Recursive Noise Adaptive Kalman Filtering by Variational Bayesian Approximations

Simo Särkkä, *Member, IEEE* and Aapo Nummenmaa

***Abstract*— This article considers the application of variational Bayesian methods to joint recursive estimation of the dynamic state and the time-varying measurement noise parameters in linear state space models. The proposed adaptive Kalman filtering method is based on forming a separable variational approximation to the joint posterior distribution of states and noise parameters on each time step separately. The result is a recursive algorithm, where on each step the state is estimated with Kalman filter and the sufficient statistics of the noise variances are estimated with a fixed-point iteration. The performance of the algorithm is demonstrated with simulated data.**

***Index Terms*— Kalman filtering, Bayesian filtering, Variational Bayesian methods, adaptive filtering, noise adaptive filtering, adaptive Kalman filtering**

I. Introduction

The *Kalman filter* (KF) [1] considers estimation of the dynamic state from noisy measurements in the class of estimation problems where the dynamic and measurement processes can be approximated by linear Gaussian state space models. The *extended Kalman filter* (EKF) and the *unscented Kalman filter* (UKF) extend this method to non-linear dynamic and measurement models by forming a Gaussian approximation to the posterior state distribution [2]–[5]. A serious limitation in these filters is that they assume complete *a priori* knowledge of the measurement and dynamic model parameters, including the noise statistics. The exact knowledge of the parameters, and especially, the noise statistics is not in many practical situations a plausible assumption. Examples of such applications are low cost integrated GPS/INS positioning systems and fault tolerant systems (see, *e.g.*, [6], [7] and references therein).

The classical way (see, *e.g.*, [8], [9]) of solving the problem of uncertain parameters is to use *adaptive filters* where the model parameters or the noise statistics are estimated together with the dynamic state. The classical noise adaptive filtering approaches can be divided into Bayesian, maximum likelihood, correlation and covariance matching methods [8]. *The Bayesian approach* is the most general of these and the other approaches can often be interpreted as approximations to the Bayesian approach. Examples of Bayesian approaches to noise adaptive filtering are state augmentation based methods [10], [11], multiple model methods such as the *interacting multiple models* (IMM) algorithm [3], [9] and particle methods [12]–[15].

Variational Bayesian (VB) methods have been developed for a wide range of models to perform approximate posterior inference at low computational cost in comparison with the sampling methods (for a review of VB methods, see, *e.g.*, [16]–[18]). These methods assume a simpler, analytically tractable form for the posterior. Two main approaches are either to derive a factored free form distribution (for models in the conjugate-exponential class), or to assume a fixed-form posterior distribution (*e.g.*, a multivariate Gaussian, with possibly a suitable parametrization of the model).

Smidl and Quinn [19] present a theoretical framework for VB-approximations, with special emphasis on recursive Bayesian inference and signal processing applications. Related to the present work, VB-approximations of AR-models with unknown variances are also considered, but general linear state space models with unknown time-varying variances are not. A general variational version of the Kalman smoother has been developed in [20] utilizing the conjugate-exponential framework, but with stationary noise. Time varying variance models have been addressed in [21] by using the fixed form approach.

Our proposed method approximates the joint posterior distribution of the state and the noise variances by a factorized free form distribution. This approximation is formed on each time step separately, and the result is a recursive algorithm, where on each step the sufficient statistics of the state and the noise variances are estimated with a fixed-point iteration of a Kalman filter. We also propose a heuristic dynamical model for the variances, which can be used for modeling the time behavior of the noise, still retaining the functional form of the approximate posterior distribution.

II. Main Results

A. Problem formulation

The discrete-time linear state space model considered here is

$$\begin{aligned} x_k &= A_k\, x_{k-1} + q_k \\ y_k &= H_k\, x_k + r_k, \end{aligned} \tag{1}$$

where $q_k \sim \mathrm{N}(0, Q_k)$ is the Gaussian process noise, $r_k \sim \mathrm{N}(0, \Sigma_k)$ is the measurement noise with diagonal covariance Σ_k, and the initial state has a Gaussian prior distribution $x_0 \sim \mathrm{N}(m_0, P_0)$. The measurement y_k is a d-dimensional vector and the state x_k is an n-dimensional vector. Time is indexed by k and the matrices A_k, H_k, Q_k, as well as the parameters of the initial state (prior) distribution m_0, P_0 are assumed to be known.

Now, departing from the case of standard Kalman filter, we assume that the observation noise variance parameters $\sigma^2_{k,i}, i = 1, \ldots, d$, are stochastic with independent dynamic models. We denote the diagonal covariance matrix comprising of these variances by $\Sigma_k = \mathrm{diag}(\sigma^2_{k,1}, ..., \sigma^2_{k,d})$. The construction of a suitable dynamical model of the observation noise variances will be discussed in Section (II-C), and at this stage we denote it generically by $p(\Sigma_k \,|\, \Sigma_{k-1})$. We assume that the dynamic models of the states and the variance parameters are independent:

$$p(x_k, \Sigma_k \,|\, x_{k-1}, \Sigma_{k-1}) = p(x_k \,|\, x_{k-1})\, p(\Sigma_k \,|\, \Sigma_{k-1}). \tag{2}$$

Manuscript received September 13, 2008; revised September 13, 2008.

Simo Särkkä* is with Helsinki University of Technology, P.O. Box 9203. FIN-02015 TKK, Finland. E-mail: simo.sarkka@tkk.fi, Tel: +358 40 757 0730, Fax: +358 9 451 4830.

Aapo Nummenmaa is with Helsinki University of Technology, P.O. Box 9203. FIN-02015 TKK, Finland. E-mail: aapo.nummenmaa@tkk.fi, Tel: +358 9 451 4834, Fax: +358 9 451 4830.

The goal of Bayesian optimal filtering of the above model is to compute the posterior distribution $p(x_k, \Sigma_k \,|\, y_{1:k})$. Formally, the well-known recursive solution to this filtering problem consists of the following steps [2]:

- *Initialization*: The recursion starts from the prior distribution $p(x_0, \Sigma_0)$.
- *Prediction*: The predictive distribution of the state x_k and measurement noise covariance Σ_k is given by the Chapman-Kolmogorov equation:

$$\begin{aligned} p(x_k, \Sigma_k \,|\, y_{1:k-1}) = \int & p(x_k \,|\, x_{k-1})\, p(\Sigma_k \,|\, \Sigma_{k-1}) \\ & \times p(x_{k-1}, \Sigma_{k-1} \,|\, y_{1:k-1})\, \mathrm{d}x_{k-1}\, \mathrm{d}\Sigma_{k-1}. \end{aligned} \tag{3}$$

- *Update*: Given the next measurement y_k, the predictive distribution above is updated to a posterior distribution by the Bayes' rule:

$$p(x_k, \Sigma_k \,|\, y_{1:k}) \propto p(y_k \,|\, x_k, \Sigma_k)\, p(x_k, \Sigma_k \,|\, y_{1:k-1}). \tag{4}$$

The integrations involved in the general solution are usually not analytically tractable. In the following, we solve the above recursion effectively by using a VB approximation for the posterior update, accompanied by suitable (heuristic) dynamics for the observation noise variances.

B. Variational approximation

Assume that we can approximate the conditional distribution for x_{k-1} and Σ_{k-1} given the measurements $y_1, \ldots, y_{k-1}$ as a product of Gaussian and independent Inverse-Gamma distributions[1] as follows:

$$\begin{aligned} p(x_{k-1}, \Sigma_{k-1} \,|\, y_{1:k-1}) &= \mathrm{N}(x_{k-1} \,|\, m_{k-1}, P_{k-1}) \\ &\times \prod_{i=1}^{d} \text{Inv-Gamma}(\sigma_{k-1,i}^2 \,|\, \alpha_{k-1,i}, \beta_{k-1,i}). \end{aligned}$$

We have chosen this approximation, because the Inverse-Gamma distribution is the conjugate prior distribution for the variance of a Gaussian distribution. For this reason, using an Inverse-Gamma model for variances of Gaussian models is common in Bayesian analysis [22]. And for the same reason, the Inverse-Gamma distribution turns out to be the natural approximating distribution in the VB-approach considered here.

Because the dynamics of the state and observation noise variances are assumed to be independent, the factored form will remain in the prediction step, as can be easily seen from the Equation (3). The linear dynamics will result in a Gaussian distribution for predicted x_k, with parameters m_k^-, P_k^- given by the standard KF prediction equation.

Let us assume that the dynamical models for the variance parameters $p(\sigma_{k,i}^2 \,|\, \sigma_{k-1,i}^2)$ are of such form that the prediction step will always result in an Inverse-Gamma predicted distribution for each $\sigma_{k,i}^2$, with parameters $\alpha_{k,i}^-, \beta_{k,i}^-$ (we will discuss possibilities for such dynamics later on). The joint prediction distribution is thus

$$\begin{aligned} p(x_k, \Sigma_k \,|\, y_{1:k-1}) &= p(x_k \,|\, y_{1:k-1})\, p(\Sigma_k \,|\, y_{1:k-1}) \\ &= \mathrm{N}(x_k \,|\, m_k^-, P_k^-) \prod_{i=1}^{d} \text{Inv-Gamma}(\sigma_{k,i}^2 \,|\, \alpha_{k,i}^-, \beta_{k,i}^-). \end{aligned}$$

Now, in the posterior update step, the state and observation noise variance parameters will be coupled through the likelihood $p(y_k \,|\, x_k, \Sigma_k)$, and the exact posterior will not be of tractable form.

In order to make the computations tractable we will now form a variational approximation to the posterior distribution. We follow the standard VB-approach (see, *e.g.*, [16]–[19]) and search for a free form factored approximate distribution for $p(x_k, \Sigma_k \,|\, y_{1:k})$ as follows:

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

The VB-approximation can now be formed by minimizing the the Kullback-Leibler (KL) divergence [23] between the separable approximation and the true posterior:

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

Using the methods from calculus of variations (see, *e.g.*, [24]) for minimizing the KL-divergence with respect to the probability densities $Q_x(x_k)$ and $Q_\Sigma(\Sigma_k)$ in turn, while keeping the other fixed yields:

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

This cannot be solved directly, as the above equations are coupled. However, computing the expectation in the first equation gives the following:

$$\begin{aligned} &\int \log p(y_k, x_k, \Sigma_k \,|\, y_{1:k-1})\, Q_\Sigma(\Sigma_k)\, \mathrm{d}\Sigma_k \\ &\quad = -\frac{1}{2} (y_k - H_k\, x_k)^T \langle \Sigma_k^{-1} \rangle_\Sigma (y_k - H_k\, x_k) \\ &\qquad - \frac{1}{2} (x_k - m_k^-)^T \left(P_k^- \right)^{-1} (x_k - m_k^-) + C_1, \end{aligned} \tag{5}$$

where $\langle \cdot \rangle_\Sigma = \int (\cdot)\, Q_\Sigma(\Sigma_k)\, \mathrm{d}\Sigma_k$ denotes the expected value with respect to the approximating distribution $Q_\Sigma(\Sigma_k)$ and C_1 denotes terms independent of x_k. As a function of x_k, this is a quadratic implying that $Q_x(x_k)$ is a Gaussian, mean and covariance of which can be found with standard matrix manipulations.

[1] In this article, we use the same parametrization of Gamma and Inverse-Gamma distributions as in [22].

Similarly, the second expectation can be computed as follows:

$$\begin{aligned}&\int \log p(y_k, x_k, \Sigma_k \,|\, y_{1:k-1})\, Q_x(x_k)\, \mathrm{d}x_k \\ &\quad = -\sum_{i=1}^{d} \left(\frac{3}{2} + \alpha_{k,i}\right) \ln(\sigma_{k,i}^2) - \sum_{i=1}^{d} \frac{\beta_{k,i}}{\sigma_{k,i}^2} \\ &\qquad - \frac{1}{2} \sum \frac{1}{\sigma_{k,i}^2} \langle (y_k - H_k\, x_k)_i^2 \rangle_x + C_2,\end{aligned} \tag{6}$$

where $\langle \cdot \rangle_x = \int (\cdot)\, Q_x(x_k)\, \mathrm{d}x_k$. From the equation it can be seen that $Q_\Sigma(\Sigma_k)$ is a product of Inverse-Gamma distributions. By evaluating the expectations in Equations (5) and (6), and matching the parameters of the distributions on left and right hand sides gives the following densities:

$$\begin{aligned}Q_x(x_k) &= \mathrm{N}(x_k \,|\, m_k, P_k) \\ Q_\Sigma(\Sigma_k) &= \prod_{i=1}^{d} \text{Inv-Gamma}(\sigma_{k,i}^2 \,|\, \alpha_{k,i}, \beta_{k,i}),\end{aligned}$$

where the parameters $m_k, P_k, \alpha_{k,i}, \beta_{k,i}$ are the solution to the following coupled set of equations:

$$\begin{aligned}m_k &= m_k^- + P_k^- H_k^T (H_k P_k^- H_k^T + \hat{\Sigma}_k)^{-1} (y_k - H_k m_k^-) \\ P_k &= P_k^- - P_k^- H_k^T (H_k P_k^- H_k^T + \hat{\Sigma}_k)^{-1} H_k P_k^- \\ \alpha_{k,i} &= 1/2 + \alpha_{k-1,i} \\ \beta_{k,i} &= \beta_{k-1,i} + \frac{1}{2} \left[(y_k - H_k\, m_k)_i^2 + (H_k\, P_k\, H_k^T)_{ii} \right],\end{aligned} \tag{7}$$

where $i = 1, \ldots, d$ and

$$\hat{\Sigma}_k = \text{diag}(\beta_{k,1}/\alpha_{k,1}, \ldots, \beta_{k,d}/\alpha_{k,d}).$$

In this manner, we have completed one prediction-update cycle ending up with an approximate posterior distribution of the same functional form as the one we began with. This same procedure can now be applied to the next measurement and the measurement after that, which results in a recursive type of filtering algorithm. However, in order to complete the algorithm description we will construct a suitable dynamic model for the noise variances in the next section.

C. Dynamics of observation noise variances

It is not straightforward to choose a dynamical model $p(\Sigma_k \,|\, \Sigma_{k-1})$ such that the Chapman-Kolmogorov equation (3) for the observation noise variances would yield Inverse-Gamma distributions as their predicted distributions. Usually, the dynamical model of the noise variances is not known in detail, and hence we propose a heuristic dynamics for the variances, which simply "spreads" their previous approximate posteriors. We choose to keep the expected measurement noise *precisions* (inverse variances) constant, and increase their variances by a factor of ρ^{-1}, $\rho \in (0, 1]$. This is obtained by the following, for $i = 1, \ldots, d$:

$$\begin{aligned}\alpha_{k,i}^- &= \rho_i\, \alpha_{k-1,i} \\ \beta_{k,i}^- &= \rho_i\, \beta_{k-1,i}.\end{aligned} \tag{8}$$

Note that the value $\rho = 1$ corresponds to stationary variances and lower values increase their assumed time-fluctuations. A strong practical motivation for this specific heuristic is the stability of the VB scheme: it ensures that the algorithm starts from the same posterior (parameter values) where it was terminated at the previous time point and thus if the parameters already have the correct values, they remain unaffected (see the fixed point iteration in Algorithm 1).

D. The Adaptive Filtering Algorithm

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

$$\begin{aligned}m_k^- &= A_k\, m_{k-1} \\ P_k^- &= A_k\, P_{k-1}\, A_k^T + Q_k \\ \alpha_{k,i}^- &= \rho_i\, \alpha_{k-1,i}, && i = 1, \ldots, d \\ \beta_{k,i}^- &= \rho_i\, \beta_{k-1,i}, && i = 1, \ldots, d\end{aligned}$$

- *Update*: First set $m_k^{(0)} = m_k^-$, $P_k^{(0)} = P_k^-$, $\alpha_{k,i} = 1/2 + \alpha_{k,i}^-$, and $\beta_{k,i}^{(0)} = \beta_{k,i}^-$, for $i = 1, \ldots, d$. Then iterate the following a few, say N, steps:

$$\begin{aligned}\hat{\Sigma}_k^{(n)} &= \text{diag}(\beta_{k,1}^{(n)}/\alpha_{k,1}^{(n)}, \ldots, \beta_{k,d}^{(n)}/\alpha_{k,d}^{(n)}) \\ m_k^{(n+1)} &= m_k^- \\ &\quad + P_k^- H_k^T (H_k P_k^- H_k^T + \hat{\Sigma}_k^{(n)})^{-1} (y_k - H_k m_k^-) \\ P_k^{(n+1)} &= P_k^- \\ &\quad - P_k^- H_k^T (H_k P_k^- H_k^T + \hat{\Sigma}_k^{(n)})^{-1} H_k P_k^- \\ \beta_{k,i}^{(n+1)} &= \beta_{k,i}^- + \frac{1}{2} \left(y_k - H_k\, m_k^{(n+1)} \right)_i^2 \\ &\quad + \frac{1}{2} \left(H_k\, P_k^{(n+1)}\, H_k^T \right)_{ii}, \qquad i = 1, \ldots, d,\end{aligned}$$

  and set $\beta_{k,i} = \beta_{k,i}^{(N)}$, $m_k = m_k^{(N)}$, $P_k = P_k^{(N)}$.

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

The adaptive filtering algorithm is presented as Algorithm 1. The algorithm should be started from a prior of the form

$$p(x_0, \Sigma_0) = \mathrm{N}(x_0 \,|\, m_0, P_0) \prod_{i=1}^{d} \text{Inv-Gamma}(\sigma_{0,i}^2 \,|\, \alpha_{0,i}, \beta_{0,i}),$$

and the approximation formed by the algorithm on step k is then

$$\begin{aligned}&p(x_k, \Sigma_k \,|\, y_{1:k}) \\ &\qquad \approx \mathrm{N}(x_k \,|\, m_k, P_k) \prod_{i=1}^{d} \text{Inv-Gamma}(\sigma_{k,i}^2 \,|\, \alpha_{k,i}, \beta_{k,i}).\end{aligned}$$

The prediction step of the algorithm consists of the standard KF prediction step and the heuristic prediction presented in Equations (8). The update step is simply a fixed point iteration algorithm for the Equations (7). Note that the degrees of freedom parameters of the Inverse-Gamma distributions do not change in the fixed point iteration. Also, the update of

$Q_x(x_k)$ is the standard KF update step with the known noise covariance $\hat{\Sigma}_k$. This is used to find a distribution for the noise variances, which would give the best separable match to the (intractable) joint posterior of state and observation noise variances. The KF update is then performed again with the obtained new expected noise covariance. The fixed-point algorithm is in fact a type of natural gradient method, which can be shown to have asymptotic optimal convergence properties [25].

In experiments we have noticed that the fixed point iteration tends to converge very fast and most of the time only few, say 2 or 3, iterations are enough in practice.

III. Illustrative Example

A. Stochastic Resonator Model

In this simulation the phenomenon is a randomly drifting stochastic resonator, which is measured at time intervals of $\Delta t = 1/10$ seconds. The dynamic model for the resonator with angular velocity ω can be written in form

$$\begin{pmatrix} x_{1,k+1} \\ x_{2,k+1} \\ x_{3,k+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\omega\,\Delta t) & \frac{\sin(\omega\,\Delta t)}{\omega} \\ 0 & -\omega\,\sin(\omega\,\Delta t) & \cos(\omega\,\Delta t) \end{pmatrix} \begin{pmatrix} x_{1,k} \\ x_{2,k} \\ x_{3,k} \end{pmatrix} + \begin{pmatrix} w_{1,k} \\ w_{2,k} \\ w_{3,k} \end{pmatrix}, \tag{9}$$

where $w_k = (w_{k,1}\ w_{k,2}\ w_{k,3})^T$ is a white Gaussian noise sequence with a known covariance Q. The measurement model at time step k is

$$y_k = x_{k,1} + x_{k,2} + v_k,$$

where v_k is a Gaussian random variable with unknown time varying variance σ_k^2. In the simulation, angular velocity $\omega = 0.05$ was used. The initial state was assumed to be known with variance 1 in each component, that is, $x_{0,i} \sim \mathrm{N}(0,1)$ for $k = 1,2,3$. Because no prior information was available on the measurement noise variance, a rather diffuse prior $\sigma_0^2 \sim$ Inv-Gamma$(1,1)$ was used.

The simulated data is shown in the Figure 1. In the simulation the measurement noise first had the variance 0.2 (see Figure 2). Around time $t = 100$ the variance quickly increased to 1 and then around $t = 200$ it again quickly decreased back to the value 0.2.

The following methods were tested:

- *Variational Bayesian adaptive Kalman filter (VB-AKF)* with the variance decreasing factor $\rho = 1 - \exp(-4)$. The number of fixed point iterations per time step was $N = 2$.
- *Interacting multiple models (IMM)* [3], [9], where a separate mode was used for each variance level $0.1 \ldots 1.2$ in steps of 0.01 (total of 111 models). The mode transition probabilities were chosen such that the probability of switching from mode i to j was proportional to $p_{ij} = \exp(-\lambda\,\Delta t|i-j|)$ with $\lambda = 5$.
- *Kalman filter (KF)* [1] for each noise level $0.1 \ldots 1.2$ separately, in steps of 0.01.

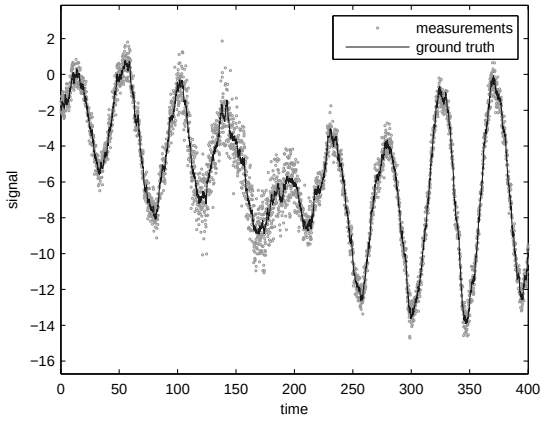

Fig. 1. Simulated noisy resonator model. The *ground truth* is the actual simulated position and the *measurements* are the noisy measurements obtained from it once per 1/10 seconds.

Note that none of the dynamic models is exactly matched to the true simulated variance dynamics. Instead, the models are selected to represent sufficiently smooth changes, which approximate the true dynamics of the variance.

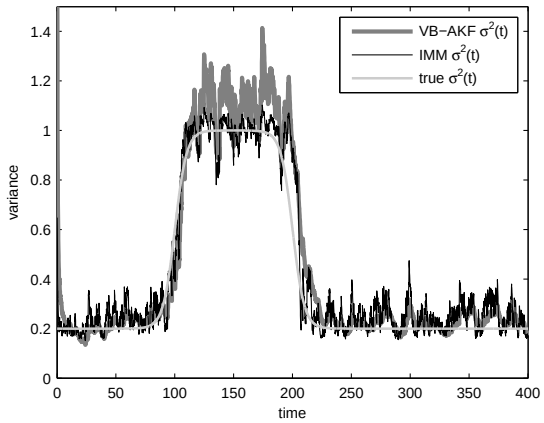

Fig. 2. Results of tracking the variance parameter of noisy resonator by VB-AKF and IMM.

The result of variance estimation, when VB-AKF and IMM methods were applied to the simulated data is shown in the Figure 2. In the figure, the actual variance and the estimated variance (posterior means) of VB-AKF and IMM estimates are shown. It can be seen that both the estimates follow the true variance trajectory quite well. IMM does seem to follow the trajectory a bit better than VB-AKF. But the performance of IMM comes with penalty of CPU time, because the CPU time needed for IMM was at least 100 times the CPU time of VB-AKF.

The root mean squared errors (RMSE) of VB-AKF and IMM are both lower than of any standard Kalman filters without the noise adaptivity. This can be seen from the Figure 3 where the RMSE results of Kalman filters with noise variance $0.1, \ldots, 1.2$ are shown together with the errors of VB-AKF and IMM. None of the Kalman filters has as low RMSE error as the VB-AKF or IMM. The CPU time of single

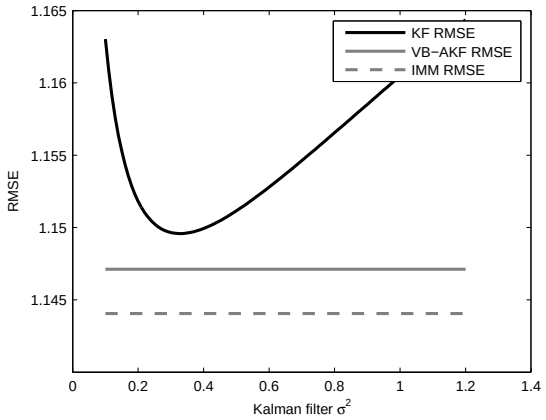

Fig. 3. RMSE errors of Kalman filters with different noise variances, and the RMSE errors of VB-AKF and IMM. It can be seen that both VB-AKF and IMM give lower RMSE errors than any of the standard Kalman filters.

KF was approximately half of the CPU time of VB-AKF.

As can be seem from the RMSE in Figure 3 and variance trajectory in Figure 2, the IMM algorithm performs a bit better than VB-AKF. This is natural, because IMM forms a grid approximation to the variance distribution and forms Gaussian state distributions conditionally to each of these grid values. In the limit of infinite number of grid points the estimate becomes exact and thus it is likely to be good approximation with a sufficiently large number of grid points – but with cost of high CPU time. The CPU time needed by IMM also grows exponentially in the number of unknown variances and thus IMM can practically be applied only to models with very low dimensional measurements.

IV. Discussion

Although, the VB-AKF method presented in this article is a Bayesian method for noise identification, the algorithm much resembles the covariance matching methods presented in, for example, [8], [10]. This is because the algorithm essentially estimates noise parameters such that the parameters become consistent with the observed residual, which is also the basis of covariance matching methods.

Because the algorithm presented in this article contains explicit Kalman filter prediction and update steps as parts of the algorithm, the extension to non-linear filtering problems can be done by replacing them with, for example, the corresponding extended Kalman filter (EKF) or unscented Kalman filter (UKF) equations [2]–[5]. The update equations of the parameters $\beta_{k,i}$ can also be easily approximated with EKF and UKF type of approximations.

V. Conclusion

In this article, we have presented a new adaptive Kalman filtering algorithm, which is based on recursively forming separable approximations to the joint distribution of state and noise parameters by variational Bayesian methods. The performance of the algorithm has been demonstrated in a simulated application.

Acknowledgment

The authors would like to thank Jouko Lampinen for helpful insights on the subject.

References

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME, Journal of Basic Engineering*, vol. 82, pp. 35–45, March 1960.
[2] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
[3] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. Wiley Interscience, 2001.
[4] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *Proceedings of the 1995 American Control, Conference, Seattle, Washington*, 1995, pp. 1628–1632.
[5] E. A. Wan and R. van der Merwe, "The unscented Kalman filter," in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. Wiley, 2001, ch. 7.
[6] C. Hide, T. Moore, and M. Smith, "Adaptive Kalman filtering algorithms for integrating GPS and low cost INS," in *Position Location and Navigation Symposium*, April 2004, pp. 227 – 233.
[7] J. Gertler, "Survey of model-based failure detection and isolation in complex plants," *IEEE Control Systems Magazine*, vol. 8(6), pp. 3–11, 1988.
[8] R. Mehra, "Approaches to adaptive filtering," *IEEE Transactions on Automatic Control*, vol. 17(5), pp. 693–698, Oct. 1972.
[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] P. Maybeck, *Stochastic Models, Estimation and Control, Volume 2*. Academic Press, 1982.
[11] E. A. Wan and A. T. Nelson, "Dual EKF methods," in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. Wiley, 2001, ch. 6.
[12] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Springer, 2001.
[13] 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.
[14] 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.
[15] S. Särkkä, "On sequential Monte Carlo sampling of discretely observed stochastic differential equations," in *Proceedings of the Nonlinear Statistical Signal Processing Workshop*, September 2006.
[16] T. S. Jaakkola, "Tutorial on variational approximation methods," in *Advanced Mean Field Methods – Theory and Practice*, M. Opper and D. Saad, Eds. MIT Press, 2001, pp. 129–159.
[17] M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, Gatsby Computational Neuroscience Unit, University College London, 2003.
[18] H. Lappalainen and J. W. Miskin, "Ensemble learning," in *Advances in Independent Component Analysis*, M. Girolami, Ed. Springer-Verlag, 2000, pp. 75–92.
[19] V. Smidl and A. Quinn, *The Variational Bayes Method in Signal Processing*. Springer, 2006.
[20] M. J. Beal and Z. Ghahramani, "The variational Kalman smoother," Gatsby Unit, Tech. Rep. TR01-003, 2001.
[21] H. Valpola, M. Harva, and J. Karhunen, "Hierarchical models of variance sources," *Signal Processing*, vol. 84(2), pp. 267–282, 2004.
[22] A. Gelman, J. B. Carlin, H. S. Stern, and D. R. Rubin, *Bayesian Data Analysis*. Chapman & Hall, 1995.
[23] D. J. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.
[24] R. Weinstock, *Calculus of Variations, with Applications to Physics and Engineering*. New York: Dover, 1974.
[25] M.-A. Sato, "Online model selection based on the variational Bayes," *Neural Computation*, no. 13, pp. 1649–1681, 2001.