

# Approximate Bayesian Smoothing with Unknown Process and Measurement Noise Covariances

Tohid Ardeshiri, Emre Özkan, Umut Orguner, Fredrik Gustafsson

**Abstract**—We present an adaptive smoother for linear state-space models with unknown process and measurement noise covariances. The proposed method utilizes the variational Bayes technique to perform approximate inference. The resulting smoother is computationally efficient, easy to implement, and can be applied to high dimensional linear systems. The performance of the algorithm is illustrated on a target tracking example.

**Index Terms**—Adaptive smoothing, variational Bayes, sensor calibration, Rauch-Tung-Striebel smoother, Kalman filtering, noise covariance, time-varying noise covariances.

## I. INTRODUCTION

Model uncertainties directly affect the performance in filtering and smoothing problems which demand an accurate knowledge of true model parameters. In most practical cases, one’s knowledge about the model may not represent the true system. Kalman filters [1], which have been widely used in many applications, also require full knowledge of model parameters for reliable estimation. The same requirement is inherited in smoothing methods which use Kalman filters as their building blocks such as Rauch-Tung-Striebel (RTS) smoother [2]. The sensitivity of the Kalman filter to model parameters has been studied in [3]–[5] and extensive research is dedicated to the identification of the parameters [6]–[8]. Noise statistics parameters are of particular interest since they determine the reliability of the information assumed to be hidden in the measurements and the system dynamics.

In the Bayesian approach, one can define priors on the unknown noise parameters and try to compute their posteriors. Here, we use variational approximation for computation of such posteriors where an analytical solution does not exist. Variational inference based techniques have been used for filtering and smoothing in a number of recent studies. For example, [9] has proposed a procedure for variational Bayesian learning of nonparametric nonlinear state-space models based on sparse Gaussian processes. In the proposed procedure the noise covariances are treated as hyper-parameters and are found via a gradient descent optimization. Variational Bayesian (VB) expectation maximization is used in [10] to identify the parameters of linear state-space models, where the process noise covariance matrix is set to the identity matrix and the remaining parameters are identified up to an unknown transformation. In [11], the measurement noise covariance is modeled as a diagonal matrix whose entries are assumed to be distributed as inverse Gamma. This result is extended and used in interactive multiple model framework for jump Markov linear systems in [12]. In [13], the conjugacy of the inverse

Wishart distributed prior with Gaussian likelihood is exploited to model and estimate the measurement noise covariance in the VB framework. It is also shown in [13] that the mean square error of state estimates can be reduced by using the proposed VB measurement update for robust filtering and smoothing. In [14], the robust filtering and smoothing for nonlinear state-space models with t-distributed measurement noise are given. In [15] the parameters of a state-space model and the noise parameters are identified using VB. Although, identification of non-diagonal noise covariances using inverse Wishart distributions is mentioned in [15], neither the analysis nor the expressions for the approximate posterior of the inverse Wishart distributed noise covariances are given. The smoothing under parameter uncertainty can also be cast into an optimization problem; Examples of recent algorithms for robust smoothing for nonlinear state-space models are presented in [16]–[20]. Such optimization based approaches can be used to compute both maximum a posteriori (MAP) and maximum likelihood (ML) estimates of the states and parameters. When the ML estimate is desired, Expectation-Maximization (EM) [21] method can be used as in [8], [20], [22], [23] to compute the ML point estimate of the noise covariance matrices. In comparison to EM, the VB method, approximates the posterior distribution of the unknown noise parameters and the state variables instead of providing only a point estimate. Further information can be extracted from the posterior as well as the point estimates with respect to different criterion. In econometrics literature concerning multivariate stochastic volatility such as [24], the estimation of covariance matrices is discussed.

In this letter, we present a novel smoothing algorithm for joint estimation of the state, measurement noise and process noise covariances using the VB technique [25, Ch. 10], [26]. Such estimation problems arise when the parameters of a state-space model are found via physical modeling of a system but the noise covariances are unknown. Our contribution is closely related to [15]. However, we consider a more general case where both of the noise covariance matrices can be non-diagonal and time-varying.

## II. PROBLEM DEFINITION

Consider the following linear time-varying state-space representation,

$$x_{k+1} = A_k x_k + w_k, \quad w_k \stackrel{iid}{\sim} \mathcal{N}(w_k; 0, Q_k), \quad (1a)$$

$$y_k = C_k x_k + v_k, \quad v_k \stackrel{iid}{\sim} \mathcal{N}(v_k; 0, R_k), \quad (1b)$$

where  $\{x_k \in \mathbb{R}^{n_x} | 0 \leq k \leq K\}$  is the state trajectory, also denoted as  $x_{0:K}$ ;  $\{y_k \in \mathbb{R}^{n_y} | 0 \leq k \leq K\}$  is the measurement sequence, denoted in more compact form as  $y_{0:K}$ ;  $A_k \in \mathbb{R}^{n_x \times n_x}$  and  $C_k \in \mathbb{R}^{n_y \times n_x}$  are known state transition and measurement matrices, respectively;  $\{w_k \in \mathbb{R}^{n_x} | 0 \leq k \leq K - 1\}$  and  $\{v_k \in \mathbb{R}^{n_y} | 0 \leq k \leq K\}$  are mutually independent and white Gaussian noise sequences.

T. Ardeshiri, E. Özkan and F. Gustafsson are with the Department of Electrical Engineering, Linköping University, 58183 Linköping, Sweden, (e-mail: tohid@isy.liu.se; emre@isy.liu.se, fredrik@isy.liu.se). This work is supported by Swedish research council (VR), project scalable Kalman filters.

U. Orguner is with Department of Electrical and Electronics Engineering, Middle East Technical University, 06531 Ankara, Turkey, (email: umut@metu.edu.tr).

The initial state  $x_0$  is assumed to have a Gaussian prior, i.e.,  $p(x_0) = \mathcal{N}(x_0; m_0, P_0)$ , where  $\mathcal{N}(\cdot; \mu, \Sigma)$  denotes the Gaussian probability density function (PDF) with mean  $\mu$  and covariance  $\Sigma$ .  $Q_k$  and  $R_k$  are the unknown positive definite process noise and measurement noise covariance matrices assumed to have initial inverse Wishart priors

$$p(Q_0) = \mathcal{IW}(Q_0; \nu_0, V_0), \quad (2a)$$

$$p(R_0) = \mathcal{IW}(R_0; \mu_0, M_0). \quad (2b)$$

The inverse Wishart PDF we use in this work is given in the following form

$$\mathcal{IW}(\Sigma; \nu, \Psi) \triangleq \frac{|\Psi|^{\frac{1}{2}(\nu-d-1)} \exp \text{Tr} \left( -\frac{1}{2}\Psi\Sigma^{-1} \right)}{2^{\frac{1}{2}(\nu-d-1)d} \Gamma_d \left[ \frac{1}{2}(\nu-d-1) \right] |\Sigma|^{\frac{\nu}{2}}}, \quad (3)$$

where  $\Sigma$  is a symmetric positive definite random matrix of dimension  $d \times d$ ,  $\nu > 2d$  is the scalar degrees of freedom and  $\Psi$  is a symmetric positive definite matrix of dimension  $d \times d$  and is called the scale matrix. This form of the inverse Wishart distribution is used in [27]. When  $\Sigma \sim \mathcal{IW}(\Sigma; \nu, \Psi)$ , then  $\Sigma^{-1} \sim \mathcal{W}(\Sigma; \nu-d-1, \Psi^{-1})$  and  $\mathbb{E}[\Sigma] = \frac{\Psi}{\nu-2d-2}$  when  $\nu-2d-2 > 0$  and  $\mathbb{E}[\Sigma^{-1}] = \Psi^{-1}(\nu-d-1)$ . Further, any diagonal element of an inverse Wishart matrix is distributed as inverse gamma [27, Corollary 3.4.2.1]. Therefore, the proposed inverse Wishart model is more general than models with diagonal covariance assumption where the diagonal entries are inverse gamma distributed, see e.g. [11].

It is common in Kalman filtering and smoothing literature to assume that the noise covariances are fixed parameters [28]. However, the noise covariances can be unknown and time-varying. In such cases, the noise parameters can be treated as state variables with dynamics. Dynamical models for covariance matrices is adopted here from [29] where the matrix Beta-Bartlett stochastic evolution model was proposed for estimating the multivariate stochastic volatility. The dynamical models for the covariance matrices  $R_k$  and  $Q_k$  are parametrized by the covariance discount factors  $0 \ll \lambda_R \leq 1$  and  $0 \ll \lambda_Q \leq 1$ , respectively. The matrix Beta-Bartlett stochastic evolution model for a generic random matrix  $\Sigma_k$  having a covariance discount factor  $0 \ll \lambda \leq 1$  is described in the following.

Let  $p(\Sigma_{k-1}) = \mathcal{IW}(\Sigma_{k-1}; \nu_{k-1}, \Psi_{k-1})$ . The forward predictive model  $p(\Sigma_k | \Sigma_{k-1})$  is such that, the forward prediction marginal density becomes the inverse Wishart density parametrized by  $p(\Sigma_k) = \mathcal{IW}(\Sigma_k; \nu_k, \Psi_k)$  where

$$\Psi_k = \lambda \Psi_{k-1}, \quad (4a)$$

$$\nu_k = \lambda \nu_{k-1} + (1-\lambda)(2d+2). \quad (4b)$$

Similar to the Kalman filter's prediction update, the forward prediction keeps the marginal expected value of  $\Sigma_k$  unchanged while the spread is increased. Furthermore, the backwards smoothing recursion is given by [29]

$$\Psi_k^{-1} \leftarrow (1-\lambda)\Psi_k^{-1} + \lambda\Psi_{k+1}^{-1}, \quad (5a)$$

$$\nu_k \leftarrow (1-\lambda)\nu_k + \lambda\nu_{k+1}. \quad (5b)$$

Note that in the prediction and smoothing iterations above, setting  $\lambda = 1$  corresponds to the fixed parameter case. The aforementioned dynamical model for noise covariances is adopted in [30] for adaptive Kalman filtering framework and in [13] for filtering and smoothing<sup>1</sup> with heavy-tailed

measurement noise covariance.

Our aim is to obtain an analytical approximation of the posterior density for the state trajectory  $x_{0:K}$  and noise covariances  $R_{0:K}$  and  $Q_{0:K-1}$ . We will derive an approximate smoother which will propagate the sufficient statistics of the approximate distributions through fixed point iterations with guaranteed convergence.

### III. VARIATIONAL SOLUTION

The *a priori* for the joint density  $p(x_0, Q_0, R_0)$  is given as follows,

$$p(x_0, Q_0, R_0) = \mathcal{N}(x_0; m_0, P_0) \mathcal{IW}(Q_0; \nu_0, V_0) \times \mathcal{IW}(R_0; \mu_0, M_0). \quad (6)$$

Then, the posterior for the state trajectory and the unknown parameters denoted by  $\mathcal{Z} \triangleq \{x_{0:K}, Q_{0:K-1}, R_{0:K}\}$  is given by Bayes' theorem as

$$p(\mathcal{Z} | y_{0:K}) \propto p(x_0, Q_0, R_0) p(y_K | x_K, R_K) \prod_{l=0}^{K-2} p(Q_{l+1} | Q_l) \times \prod_{k=0}^{K-1} p(y_k | x_k, R_k) p(x_{k+1} | x_k, Q_k) p(R_{k+1} | R_k). \quad (7)$$

There is no analytical solution for this posterior. We are going to look for an approximate analytical solution using the following variational approximation.

$$p(\mathcal{Z} | y_{0:K}) \approx q(\mathcal{Z}) \triangleq q_x(x_{0:K}) q_Q(Q_{0:K-1}) q_R(R_{0:K}), \quad (8)$$

where the densities  $q_x(\cdot)$ ,  $q_Q(\cdot)$  and  $q_R(\cdot)$  are the approximate posterior densities for  $x_{0:K}$ ,  $Q_{0:K-1}$  and  $R_{0:K}$ , respectively. The well-known technique of VB [25, Ch. 10], [26] chooses the estimates  $\hat{q}_x(\cdot)$ ,  $\hat{q}_Q(\cdot)$  and  $\hat{q}_R(\cdot)$  for the factors in (8) using the following optimization problem

$$\hat{q}_x, \hat{q}_Q, \hat{q}_R = \underset{q_x, q_Q, q_R}{\text{argmin}} D_{KL}(q(\mathcal{Z}) || p(\mathcal{Z} | y_{0:K})) \quad (9)$$

where  $D_{KL}(q(x) || p(x)) \triangleq \int q(x) \log \frac{q(x)}{p(x)} dx$  is the Kullback-Leibler divergence [31]. The optimal solution satisfies the following set of equations.

$$\log \hat{q}_x(x_{0:K}) = \mathbb{E}_{\hat{q}_x \hat{q}_R} [\log p(\mathcal{Z}, y_{0:K})] + c_x, \quad (10a)$$

$$\log \hat{q}_Q(Q_{0:K-1}) = \mathbb{E}_{\hat{q}_x \hat{q}_R} [\log p(\mathcal{Z}, y_{0:K})] + c_Q, \quad (10b)$$

$$\log \hat{q}_R(R_{0:K}) = \mathbb{E}_{\hat{q}_x \hat{q}_Q} [\log p(\mathcal{Z}, y_{0:K})] + c_R, \quad (10c)$$

where  $c_x$ ,  $c_Q$  and  $c_R$  are constants with respect to the variables  $x_{0:K}$ ,  $Q_{0:K-1}$  and  $R_{0:K}$ , respectively. The solution to (10) can be obtained via fixed-point iterations where only one factor in (8) is updated and all the other factors are fixed to their last estimated values [25, Ch. 10]. The iterations converge to a local optima of (9) [25, Ch. 10], [32, Ch. 3]. The complete (standard but tedious) derivations for the variational iterations are given in [33].

The implementation pseudo-code for the proposed algorithm is given in Table I. When the recursions of the proposed algorithm converge, the expected values or the modes of the posteriors for  $x_k$ ,  $R_k$  and  $Q_k$  can be used as the point estimates for the random variables. When an estimate of uncertainty for the point estimate is required the posterior variances can be used. Nevertheless, it is well-known that the VB method underestimates the covariance when the posterior is multimodal [25, Ch. 10].

<sup>1</sup>The expressions given in lines 21 and 22 of Algorithm 3 in [13] are inaccurate. The correct version is given in [29].

## IV. SIMULATIONS

## A. Unknown time-varying noise covariances

We illustrate the performance of the proposed smoother in an object tracking scenario. In the simulation scenario, a point object moves according to the continuous white noise acceleration model in two dimensional Cartesian coordinates [34, p. 269]. The sampling time is  $\tau = 1s$  and the simulation length  $K$  is chosen to be 4000. The state vector is defined as the position and the velocity of the object. A sensor collects noisy measurements of the object's position according to (1b). The true parameters of the linear state-space model are given as

$$\begin{aligned} A_k &= \text{Diag}(a, a), & Q_0 &= \text{Diag}(q, q), \\ a &= \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix}, & q &= \sigma_\nu^2 \begin{bmatrix} \tau^3/3 & \tau^2/2 \\ \tau^2/2 & \tau \end{bmatrix}, \\ R_k^{\text{True}} &= \left(2 - \cos\left(\frac{4\pi k}{K}\right)\right) R_0, & R_0 &= \sigma_e^2 \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}, \\ Q_k^{\text{True}} &= \left(\frac{2}{3} + \frac{1}{3} \cos\left(\frac{4\pi k}{K}\right)\right) Q_0, & C_k &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \end{aligned}$$

The noise related parameters are  $\sigma_e^2 = 2m^2$  and  $\sigma_\nu^2 = 3m^2/s^3$ . The initial values of the parameters at time index  $k = 0$  are used in the RTS smoother which are  $R_0$  and  $Q_0$ , respectively. Using the simulated measurement data, we compare four smoothers; RTS smoother using the fixed noise covariances  $R_0$  and  $Q_0$  (denoted by RTS), VB smoother for estimating only  $R_k$  as given in [13, Algorithm 3] (denoted by VBS-R), the proposed VB algorithm for estimating  $R_k$  and  $Q_k$  simultaneously (denoted by VBS-RQ), and the oracle RTS smoother which knows the true noise covariances (denoted by Oracle-RTS).

We perform  $N_{MC} = 5000$  Monte Carlo (MC) simulations. In each MC run, a new trajectory with an initial state and the corresponding measurements are generated. The prior for the initial state is assumed to be Gaussian, i.e.,  $x_0^j \stackrel{iid}{\sim} p(x_0) = \mathcal{N}(x_0; m_0, P_0)$  for the  $j^{\text{th}}$  MC simulation where,

$$m_0 = [0m, 5m/s, 0m, 5m/s]^T, \quad (11a)$$

$$P_0 = \text{Diag}([30^2, 30^2, 30^2, 30^2]). \quad (11b)$$

The initial parameters of the inverse Wishart prior densities in (2) for the smoother VBS-RQ are chosen as

$$\nu_0 = 2n_x + 3, \quad V_0 = (\nu_0 - 2n_x - 2)Q_0, \quad (12a)$$

$$\mu_0 = 2n_y + 3, \quad M_0 = (\mu_0 - 2n_y - 2)R_0. \quad (12b)$$

This choice of initial parameters yields the expected value of the initial prior densities of  $R_k$  and  $Q_k$  to coincide with the nominal values of  $R_0$  and  $Q_0$ . Similarly, the initial parameters of the inverse Wishart prior density for VBS-R are given in (12b). In the VBS-RQ and VBS-R, covariance discount factors are set to  $\lambda = 0.98$  and, the number of iterations in the variational update is set to 50.

We compare the four smoothers in terms of the root mean square error (RMSE) of the position estimates

$$\text{RMSE}(j) \triangleq \left( \frac{1}{K+1} \sum_{k=0}^K \left\| C(m_{k|K}^j - x_k^j) \right\|_2^2 \right)^{\frac{1}{2}}. \quad (13)$$

and its average RMSE over all MC simulations denoted by ARMSE. In (13),  $m_{k|K}^j$  and  $x_k^j$  denote the estimated mean of state  $x_k$  and its true value in the  $j^{\text{th}}$  MC run, respectively. For

Table I  
SMOOTHING WITH UNKNOWN NOISE COVARIANCES

---

1: **Inputs:**  $A_k, C_k, \nu_0, V_0, \mu_0, M_0, m_0, P_0, \lambda_Q, \lambda_R$  and  $y_{0:K}$ .  
*initialization*  
2:  $V_{k|K} \leftarrow V_0, \nu_{k|K} \leftarrow \nu_0, M_{k|K} \leftarrow M_0, \mu_{k|K} \leftarrow \mu_0$  for  $0 \leq k \leq K$   
3: **repeat**  
   *update*  $q_x(x_{0:K})$  given  $q_Q(Q_{0:K-1})$  and  $q_R(R_{0:K})$   
4:  $\tilde{Q}_k \leftarrow V_{k|K}/(\nu_{k|K} - n_x - 1)$  for  $0 \leq k \leq K$   
5:  $\tilde{R}_k \leftarrow M_{k|K}/(\mu_{k|K} - n_y - 1)$  for  $0 \leq k \leq K$   
6:  $m_{0|-1} \leftarrow m_0, P_{0|-1} \leftarrow P_0$   
7: **for**  $k = 0$  to  $K$  **do**  
8:  $K_k \leftarrow P_{k|k-1} C_k^T (C_k P_{k|k-1} C_k^T + \tilde{R}_k)^{-1}$   
9:  $m_{k|k} \leftarrow m_{k|k-1} + K_k (y_k - C_k m_{k|k-1})$   
10:  $P_{k|k} \leftarrow (I - K_k C_k) P_{k|k-1}$   
11:  $m_{k+1|k} \leftarrow A_k m_{k|k}$   
12:  $P_{k+1|k} \leftarrow A_k P_{k|k} A_k^T + \tilde{Q}_k$   
13: **end for**  
14: **for**  $k = K-1$  down to  $0$  **do**  
15:  $G_k \leftarrow P_{k|k} A_k^T P_{k+1|k}^{-1}$   
16:  $m_{k|K} \leftarrow m_{k|k} + G_k (m_{k+1|K} - A_k m_{k|k})$   
17:  $P_{k|K} \leftarrow P_{k|k} + G_k (P_{k+1|K} - P_{k+1|k}) G_k^T$   
18: **end for**  
   *update*  $q_Q(Q_{0:K-1})$  and  $q_R(R_{0:K})$  given  $q_x(x_{0:K})$   
19:  $\nu_{0|-1} \leftarrow \nu_0, V_{0|-1} \leftarrow V_0, \mu_{0|-1} \leftarrow \mu_0, M_{0|-1} \leftarrow M_0$   
20: **for**  $k = 0$  to  $K$  **do**  
21:  $\mu_{k|k} \leftarrow \mu_{k|k-1} + 1$   
22:  $M_{k|k} \leftarrow M_{k|k-1} + C_k P_{k|K} C_k^T$   
    $+ (y_k - C_k m_{k|K})(y_k - C_k m_{k|K})^T$   
23:  $\mu_{k+1|k} \leftarrow \mu_{k|k} + (1 - \lambda_R)(2n_y + 2), M_{k+1|k} \leftarrow \lambda_R M_{k|k}$   
24: **end for**  
25: **for**  $k = 0$  to  $K-1$  **do**  
26:  $\nu_{k|k} \leftarrow \nu_{k|k-1} + 1$   
27:  $V_{k|k} \leftarrow V_{k|k-1} + P_{k+1|K} + A_k P_{k|K} A_k^T - P_{k+1|K} A_k^T$   
    $- A_k P_{k+1|K} + (m_{k+1|K} - A_k m_{k|K})(m_{k+1|K} - A_k m_{k|K})^T$   
28:  $\nu_{k+1|k} \leftarrow \nu_{k|k} + (1 - \lambda_Q)(2n_x + 2), V_{k+1|k} \leftarrow \lambda_Q V_{k|k}$   
29: **end for**  
30: **for**  $k = K-1$  down to  $0$  **do**  
31:  $\mu_{k|K} \leftarrow (1 - \lambda_R)\mu_{k|k} + \lambda_R \mu_{k+1|K}$   
32:  $M_{k|K} \leftarrow \left( (1 - \lambda_R)M_{k|k}^{-1} + \lambda_R M_{k+1|K}^{-1} \right)^{-1}$   
33: **end for**  
34: **for**  $k = K-2$  down to  $0$  **do**  
35:  $\nu_{k|K} \leftarrow (1 - \lambda_Q)\nu_{k|k} + \lambda_Q \nu_{k+1|K}$   
36:  $V_{k|K} \leftarrow \left( (1 - \lambda_Q)V_{k|k}^{-1} + \lambda_Q V_{k+1|K}^{-1} \right)^{-1}$   
37: **end for**  
38: **until converged**  
39: **Outputs:**  $m_{k|K}, P_{k|K}, M_{k|K}, \mu_{k|K}$  and,  
    $\hat{R}_k \triangleq \mathbb{E}_{q_R}[R_k | y_{0:K}] = M_{k|K}/(\mu_{k|K} - 2n_y - 2)$  for  $k = 0 \dots K$ .  
    $V_{k|K}, \nu_{k|K}$  and  $\hat{Q}_k \triangleq \mathbb{E}_{q_Q}[Q_k | y_{0:K}] = V_{k|K}/(\nu_{k|K} - 2n_x - 2)$  for  $k = 0 \dots K - 1$ .

---

the matrices, we use the square root of the average Frobenius norm square normalized by the number of elements as the error measure

$$E_R(j) \triangleq \left( \frac{1}{n_y^2(K+1)} \sum_{k=0}^K \text{Tr} \left( (\hat{R}_k^j - R_k^{\text{True}})^2 \right) \right)^{\frac{1}{4}} \quad (14a)$$

$$E_Q(j) \triangleq \left( \frac{1}{n_x^2 K} \sum_{k=0}^{K-1} \text{Tr} \left( (\hat{Q}_k^j - Q_k^{\text{True}})^2 \right) \right)^{\frac{1}{4}}. \quad (14b)$$

In (14),  $\hat{R}_k^j, \hat{Q}_k^j, R_k^{\text{True}}$  and  $Q_k^{\text{True}}$  denote the estimated mean of measurement noise covariance, process noise covariance and their true values in the  $j^{\text{th}}$  MC run, respectively.

The estimates of some elements of the noise covariances versus time for some random samples of MC runs along with their true value are given in Fig. 1.

Error corresponding to the value for  $R_0$  and  $Q_0$  used in RTS computed using (14) are  $E_R = 2.972$  and  $E_Q = 2.224$ , respectively. The error values for the smoothers are given in

Table II  
TIME-VARYING NOISE COVARIANCES: COMPARISON OF FOUR SMOOTHERS  
IN TERMS OF ESTIMATION ERRORS.

Smoothers	Errors (Mean $\pm$ Standard deviation)		
	RMSE	$E_R$	$E_Q$
Oracle-RTS	3.608 $\pm$ 0.045	–	–
RTS	3.879 $\pm$ 0.047	–	–
VBS-R	3.712 $\pm$ 0.047	1.687 $\pm$ 0.079	–
VBS-RQ	3.653 $\pm$ 0.047	1.485 $\pm$ 0.070	1.572 $\pm$ 0.063

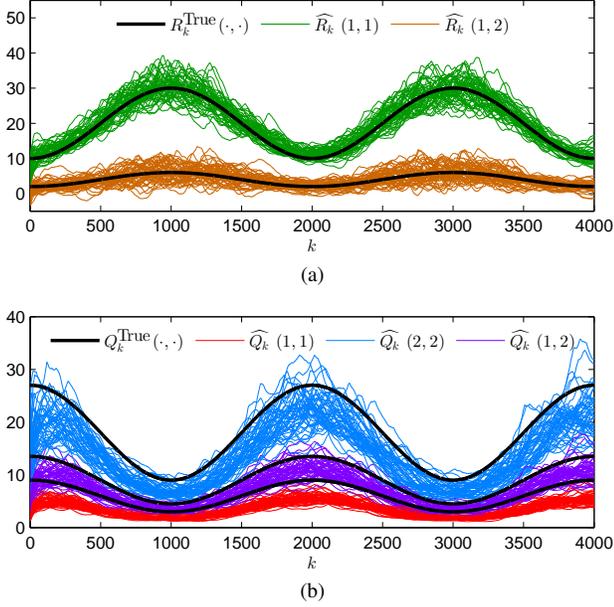


Figure 1. (a) First diagonal and the off-diagonal elements of the estimated measurement noise covariance  $\hat{R}_k$  using VBS-RQ are plotted versus time index  $k$  along with their true value (in black). (b) First two diagonal and an off-diagonal element of the estimated process noise covariance  $\hat{Q}_k$  using VBS-RQ are plotted versus time index  $k$  along with their true value (in black).

Table II. For those algorithms which do not estimate  $R_k$  and  $Q_k$  the corresponding error terms are not given in Table II.

### B. Unknown time-invariant noise covariances

When the noise covariances are time-invariant unknown parameters, the EM algorithm [20, page 182] offers an alternative to the proposed VB algorithm. An EM algorithm for smoothing with unknown noise covariances (denoted by EMS-RQ) is given in [33] and is compared to VBS-RQ. Here, we will repeat the simulation scenario in Section IV-A for  $K = 1000$  and compare six smoothers; RTS, VBS-R, VBS-RQ, Oracle-RTS, EMS-RQ and the VB smoother with diagonal covariance matrices with inverse gamma distributed entries proposed in [15] (denoted by VBS-RQ-D). Since the noise covariances are now fixed parameters, we set  $\lambda = 1$  in VBS-RQ. Furthermore, we drop the time index  $k$  and denote the noise covariances by  $R$  and  $Q$  in the rest of this section.

The nominal values of the parameters used in the RTS smoother are  $R_0$  and  $Q_0$ , respectively while their true values are  $R^{\text{True}} = 2R_0$  and  $Q^{\text{True}} = 0.2Q_0$ . Error corresponding to the nominal value for  $R$  and  $Q$  computed using (14) are  $E_R = 2.685$  and  $E_Q = 2.842$ , respectively. The error values for the smoothers are given in Table III. For those algorithms which do not estimate  $R$  and  $Q$  the corresponding error terms are not given in Table III. The convergence of some elements of the  $\hat{R}$  and  $\hat{Q}$  versus the number of iterations is illustrated in Fig. 2.

Table III  
TIME-INVARIANT NOISE COVARIANCES: COMPARISON OF SIX SMOOTHERS  
IN TERMS OF ESTIMATION ERRORS.

Smoothers	Errors (Mean $\pm$ Standard deviation)		
	RMSE	$E_R$	$E_Q$
Oracle-RTS	3.399 $\pm$ 0.088	–	–
RTS	3.786 $\pm$ 0.090	–	–
VBS-R	3.595 $\pm$ 0.090	1.326 $\pm$ 0.195	–
VBS-RQ	3.402 $\pm$ 0.088	0.929 $\pm$ 0.211	0.668 $\pm$ 0.129
EMS-RQ	3.407 $\pm$ 0.088	0.975 $\pm$ 0.212	0.851 $\pm$ 0.075
VBS-RQ-D	3.433 $\pm$ 0.089	1.764 $\pm$ 0.056	1.515 $\pm$ 0.009

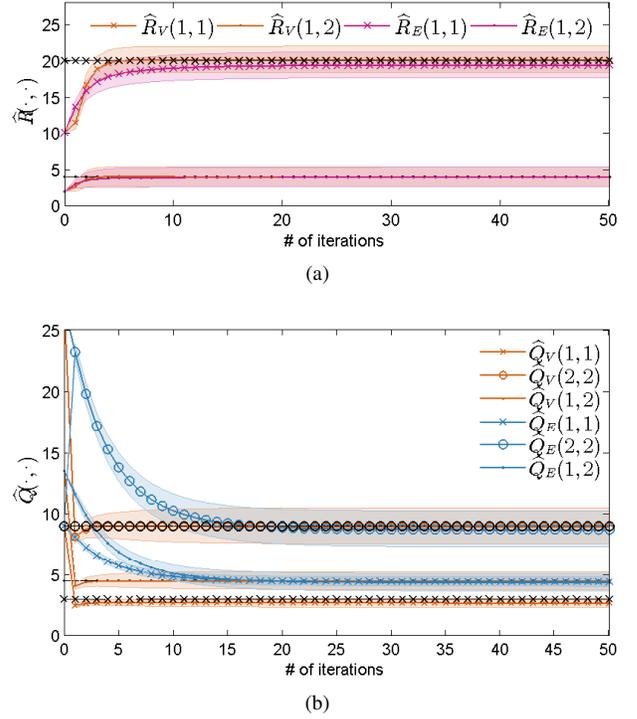


Figure 2. Estimated elements of noise covariance matrices using VBS-RQ (denoted by subscript  $V$ ) and EMS-RQ (denoted by subscript  $E$ ) are plotted versus number of iterations of the algorithm along with their true corresponding value (in black). The median values are plotted as solid line along with shaded areas in the same color as the median curve illustrating the interval between 5 and 95 percentiles. (a) First diagonal element and the off-diagonal element of estimated measurement noise covariance matrices. (b) First two diagonal elements and an off-diagonal element of estimated process noise covariance matrices.

## V. DISCUSSION AND CONCLUSION

We have proposed a smoothing technique based on a variational Bayes approximation. We have shown a successful numerical simulation using variational Bayes for approximate inference for a linear state-space model with unknown time-varying measurement noise and process noise covariances. In our simulations we obtain lower ARMSE for the state estimate compared to RTS smoother in presence of modeling mismatch. Furthermore, we obtain lower ARMSE for the state estimate compared to other state-of-the-art smoothers which identify the noise covariances using EM which is a consequence of the fact that the algorithm iteratively finds a better estimate of the process noise and measurement noise covariances.

The proposed algorithm for general time-varying noise covariance estimation can be restricted to a fixed noise parameter estimation algorithm by choosing a unity covariance discount factor. Furthermore, when the sparsity pattern in a covariance matrix is known a priori, the elements which are zero can be set to zero to obtain a tailored algorithm.

## REFERENCES

- [1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME—Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.
- [2] H. E. Rauch, C. T. Striebel, and F. Tung, "Maximum Likelihood Estimates of Linear Dynamic Systems," *Journal of the American Institute of Aeronautics and Astronautics*, vol. 3, no. 8, pp. 1445–1450, Aug 1965.
- [3] H. Heffes, "The effect of erroneous models on the kalman filter response," *Automatic Control, IEEE Transactions on*, vol. 11, no. 3, pp. 541–543, Jul 1966.
- [4] S. Sangsuk-Iam and T. Bullock, "Analysis of discrete-time kalman filtering under incorrect noise covariances," *Automatic Control, IEEE Transactions on*, vol. 35, no. 12, pp. 1304–1309, Dec 1990.
- [5] B. Anderson and J. Moore, *Optimal Filtering*, ser. Dover Books on Electrical Engineering. Dover Publications, 2012. [Online]. Available: <http://books.google.se/books?id=iYMqLQp49UMC>
- [6] R. Mehra, "On the identification of variances and adaptive Kalman filtering," *Automatic Control, IEEE Transactions on*, vol. 15, no. 2, pp. 175 – 184, apr 1970.
- [7] —, "Approaches to adaptive filtering," *Automatic Control, IEEE Transactions on*, vol. 17, no. 5, pp. 693 – 698, oct 1972.
- [8] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the em algorithm," *Journal of time series analysis*, vol. 3, no. 4, pp. 253–264, 1982.
- [9] R. Frigola, Y. Chen, and C. Rasmussen, "Variational Gaussian process state-space models," in *Advances in Neural Information Processing Systems 27*, 2014, pp. 3680–3688. [Online]. Available: <http://papers.nips.cc/paper/5375-variational-gaussian-process-state-space-models.pdf>
- [10] M. J. Beal, "Variational algorithms for approximate bayesian inference," Ph.D. dissertation, Gatsby Computational Neuroscience Unit, University College London, 2003.
- [11] S. Särkkä and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *Automatic Control, IEEE Transactions on*, vol. 54, no. 3, pp. 596 –600, march 2009.
- [12] W. Li and Y. Jia, "State estimation for jump Markov linear systems by variational Bayesian approximation," *Control Theory Applications, IET*, vol. 6, no. 2, pp. 319 –326, 19 2012.
- [13] G. Agamennoni, J. Nieto, and E. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *Signal Processing, IEEE Transactions on*, vol. 60, no. 10, pp. 5024 –5037, oct. 2012.
- [14] R. Pichè, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate student-t distribution," in *Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on*, sept. 2012, pp. 1 –6.
- [15] D. Barber and S. Chiappa, "Unified inference for variational Bayesian linear Gaussian state-space models," in *Advances in Neural Information Processing Systems 19*. MIT Press, 2007, pp. 81–88. [Online]. Available: <http://papers.nips.cc/paper/3023-unified-inference-for-variational-bayesian-linear-gaussian-state-space-models.pdf>
- [16] G. Agamennoni and E. Nebot, "Robust estimation in non-linear state-space models with state-dependent noise," *Signal Processing, IEEE Transactions on*, vol. 62, no. 8, pp. 2165–2175, April 2014.
- [17] A. Y. Aravkin, J. V. Burke, and G. Pillonetto, "Optimization viewpoint on Kalman smoothing, with applications to robust and sparse estimation," *ArXiv e-prints*, Mar. 2013.
- [18] —, "Robust and Trend Following Student's t Kalman Smoothers," *ArXiv e-prints*, Mar. 2013.
- [19] A. Aravkin, B. Bell, J. Burke, and G. Pillonetto, "An  $\ell_1$  -Laplace robust Kalman smoother," *Automatic Control, IEEE Transactions on*, vol. 56, no. 12, pp. 2898–2911, Dec 2011.
- [20] S. Särkkä, *Bayesian Filtering and Smoothing*. New York, NY, USA: Cambridge University Press, 2013.
- [21] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. pp. 1–38, 1977. [Online]. Available: <http://www.jstor.org/stable/2984875>
- [22] S. Gibson and B. Ninness, "Robust maximum-likelihood estimation of multivariable dynamic systems," *Automatica*, vol. 41, no. 10, pp. 1667 – 1682, 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0005109805001810>
- [23] Z. Ghahramani and G. E. Hinton, "Parameter estimation for linear dynamical systems," Department of Computer Science, University of Toronto, Tech. Rep., 1996.
- [24] J. Fan, Y. Fan, and J. Lv, "High dimensional covariance matrix estimation using a factor model," *Journal of Econometrics*, vol. 147, no. 1, pp. 186–197, November 2008. [Online]. Available: <http://ideas.repec.org/a/eee/econom/v147y2008i1p186-197.html>
- [25] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2007.
- [26] D. Tzikas, A. Likas, and N. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.
- [27] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*. Boca Raton, FL: Chapman & Hall/CRC, 2000.
- [28] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*, ser. Prentice-Hall information and system sciences series. Prentice Hall, 2000. [Online]. Available: <https://books.google.se/books?id=zNjFAQAIAAJ>
- [29] C. M. Carvalho and M. West, "Dynamic matrix-variate graphical models," *Bayesian Anal.*, vol. 2, pp. 69–98, 2007.
- [30] S. Särkkä and J. Hartikainen, "Non-linear noise adaptive Kalman filtering via variational Bayes," in *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*, Sept 2013, pp. 1–6.
- [31] T. M. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, 2006.
- [32] M. Wainwright and M. Jordan, *Graphical Models, Exponential Families, and Variational Inference*, ser. Foundations and trends in machine learning. Now Publishers, 2008.
- [33] T. Ardashiri, E. Özkan, U. Orguner, and F. Gustafsson, "Variational iterations for smoothing with unknown process and measurement noise covariances," Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, Tech. Rep. LiTH-ISY-R-3086, August 2015. [Online]. Available: <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-120700>
- [34] Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. Wiley, 2004.