Technical report from Automatic Control at Linköpings universitet

Variational Iterations for Smoothing with Unknown Process and Measurement Noise Covariances

Tohid Ardeshiri, Emre Özkan, Umut Orguner, Fredrik Gustafsson

Division of Automatic Control

E-mail: tohid@isy.liu.se, emre@isy.liu.se, umut@metu.edu.tr, fredrik@isy.liu.se

30th August 2015

Report no.: LiTH-ISY-R-3086
Submitted to

Address:
Department of Electrical Engineering
Linköpings universitet
SE-581 83 Linköping, Sweden

WWW: http://www.control.isy.liu.se

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Abstract

In this technical report, some derivations for the smoother proposed in [1] are presented. More specifically, the derivations for the cyclic iteration needed to solve the variational Bayes smoother for linear state-space models with unknown process and measurement noise covariances in [1] are presented. Further, the variational iterations are compared with iterations of the Expectation Maximization (EM) algorithm for smoothing linear state-space models with unknown noise covariances.

Keywords: Adaptive smoothing, variational Bayes, sensor calibration, Rauch-Tung-Striebel smoother, Kalman filtering, noise covariance

Variational Iterations for Smoothing with Unknown Process and Measurement Noise Covariances

Tohid Ardeshiri, Emre Özkan, Umut Orguner*
and Fredrik Gustafsson

2015-08-31

Abstract

In this technical report, some derivations for the smoother proposed in [1] are presented. More specifically, the derivations for the cyclic iteration involved in the variational Bayes smoother for linear state-space models with unknown process and measurement noise covariances in [1] are presented. Further, the variational iterations are compared with iterations of the Expectation Maximization (EM) algorithm for smoothing linear state-space models with unknown noise covariances.

1 Problem formulation

A Bayesian smoother using the variational Bayes method is given in [1]. The algorithm computes an approximation of the smoothing distribution for the state variable and the unknown noise covariances.

The dynamical model for the covariance matrix Σ_k adopted in [1] is taken from [2] where the matrix Beta-Bartlett stochastic evolution model was proposed for estimating the multivariate stochastic volatility. Let $0 \ll \lambda \leq 1$ be a covariance discount factor and $p(\Sigma_{k-1}) = \mathcal{IW}(\Sigma_{k-1}; \nu_{k-1}, \Psi_{k-1})$. 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

$$\Psi_k = \lambda \Psi_{k-1}, \tag{1a}$$
$$\nu_k = \lambda \nu_{k-1} + (1-\lambda)(2d+2). \tag{1b}$$

Furthermore, the backwards smoothing recursion is given as [2]

$$\Psi_k^{-1} \leftarrow (1-\lambda)\Psi_k^{-1} + \lambda \Psi_{k+1}^{-1}, \tag{2a}$$
$$\nu_k \leftarrow (1-\lambda)\nu_k + \lambda \nu_{k+1}. \tag{2b}$$

Here, we derive the expectations needed for the cyclic iterations of the variational Bayes smoother given in [1] which approximates the joint smoothing

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

posterior density for the states and the process and measurement noise covariance matrices. The joint smoothing posterior density

$$
\begin{aligned}
p(x_{0:K}, Q_{0:K-1}, R_{0:K}|y_{0:K}) \propto\ & p(x_{0:K}, Q_{0:K-1}, R_{0:K}, y_{0:K}) && (3)\\
=& p(x_0, Q_0, R_0) p(y_K|x_K, R_K) \prod_{l=0}^{K-2} p(Q_{l+1}|Q_l)\\
& \times \prod_{k=0}^{K-1} p(y_k|x_k, R_k) p(x_{k+1}|x_k, Q_k) p(R_{k+1}|R_k) && (4)
\end{aligned}
$$

is approximated in [1] by a factorized probability density function (PDF) in the form

$$
p(x_{0:K}, R_{0:K}, Q_{0:K-1}|y_{0:K}) \approx q_x(x_{0:K}) q_Q(Q_{0:K-1}) q_R(R_{0:K}). \tag{5}
$$

The analytical solutions for $\hat{q}_x$, $\hat{q}_Q$ and $\hat{q}_R$ can be obtained by cyclic iteration of the following form.

$$
\log \hat{q}_x(x_{0:K}) \leftarrow \mathop{\mathbb{E}}_{\hat{q}_Q \hat{q}_R} [\log p(x_{0:K}, Q_{0:K-1}, R_{0:K}, y_{0:K})] + c_x, \tag{6a}
$$

$$
\log \hat{q}_Q(Q_{0:K-1}) \leftarrow \mathop{\mathbb{E}}_{\hat{q}_x \hat{q}_R} [\log p(x_{0:K}, Q_{0:K-1}, R_{0:K}, y_{0:K})] + c_Q, \tag{6b}
$$

$$
\log \hat{q}_R(R_{0:K}) \leftarrow \mathop{\mathbb{E}}_{\hat{q}_x \hat{q}_Q} [\log p(x_{0:K}, Q_{0:K-1}, R_{0:K}, y_{0:K})] + c_R, \tag{6c}
$$

where the expected values on the right hand sides of (6) are taken with respect to the current q_x, q_Q and q_R and c_x, c_Q and c_R are constants with respect to the variables x_k, Q_k and R_k, respectively [3, Chapter 10] [4].

2 Derivations for the smoother

In subsections 2.1 to 2.3, we will derive the necessary expressions for completing one iteration of the algorithm. For brevity all constant values are denoted by c in the derivation. Starting from the last estimate of the distributions (i.e., the ith iterates), we derive the $(i+1)$th iterates which are denoted as $q_x^{(i+1)}(\cdot)$, $q_Q^{(i+1)}(\cdot)$ and $q_R^{(i+1)}(\cdot)$. The joint density $p(x_{0:K}, Q_{0:K-1}, R_{0:K}, y_{0:K})$ needed for the derivations is given as follows,

$$
\begin{aligned}
p(x_{0:K}, Q_{0:K-1}, R_{0:K}, y_{0:K}) =\ & p(x_0, Q_0, R_0) p(y_K|x_K, R_K) \prod_{l=0}^{K-2} p(Q_{l+1}|Q_l)\\
& \times \prod_{k=0}^{K-1} p(y_k|x_k, R_k) p(x_{k+1}|x_k, Q_k) p(R_{k+1}|R_k) && (7)\\
=& \mathcal{N}(x_0; m_0, P_0) \mathcal{IW}(Q_0; \nu_0, V_0) \mathcal{IW}(R_0; \mu_0, M_0) \mathcal{N}(y_K; C_K x_K, R_K)\\
& \times \prod_{k=0}^{K-1} \mathcal{N}(y_k; C_k x_k, R_k) \mathcal{N}(x_{k+1}; A_k x_k, Q_k) p(R_{k+1}|R_k)\\
& \times \prod_{l=0}^{K-2} p(Q_{l+1}|Q_l). && (8)
\end{aligned}
$$

2.1 Derivations for the approximate posterior $q_x^{(i+1)}(\cdot)$

Using (6a) and (8), we obtain

$$
\begin{aligned}
\log q_x^{(i+1)}(x_{0:K}) &= \log \mathcal{N}(x_0; m_0, P_0) \\
&- \frac{1}{2}\sum_{k=0}^{K-1} \underset{q_Q^{(i)}}{\mathbb{E}}\left[\mathrm{Tr}\left(Q_k^{-1}(x_{k+1} - A_k x_k)(x_{k+1} - A_k x_k)^{\mathrm{T}}\right)\right] \\
&- \frac{1}{2}\sum_{k=0}^{K} \underset{q_R^{(i)}}{\mathbb{E}}\left[\mathrm{Tr}\left(R_k^{-1}(y_k - C_k x_k)(y_k - C_k x_k)^{\mathrm{T}}\right)\right] + c \qquad (9)\\
&= \log \mathcal{N}(x_0; m_0, P_0) + \sum_{k=0}^{K-1} \log \mathcal{N}(x_{k+1}; A_k x_k, (\underset{q_Q^{(i)}}{\mathbb{E}}[Q_k^{-1}])^{-1}) \\
&+ \sum_{k=0}^{K} \log \mathcal{N}(y_k; C_k x_k, (\underset{q_R^{(i)}}{\mathbb{E}}[R_k^{-1}])^{-1}) + c. \qquad (10)
\end{aligned}
$$

Hence, (10) has the same form as the logarithm of the joint posterior distribution of the state trajectory in a linear-Gaussian state-space model with the process noise covariance $\widetilde{Q}_k \triangleq (\mathbb{E}_{q_Q^{(i)}}[Q_k^{-1}])^{-1}$ and with the measurement noise covariance $\widetilde{R}_k \triangleq (\mathbb{E}_{q_R^{(i)}}[R_k^{-1}])^{-1}$. The approximate posterior density $q_x^{(i+1)}(x_{0:K})$ can be computed using the well-known RTS smoother [5].

2.2 Derivations for the approximate posterior $q_Q^{(i+1)}(\cdot)$

The variational form for $q_Q(\cdot)$, using (6b) and (8) obeys

$$
\begin{aligned}
\log q_Q^{(i+1)}(Q_{0:K-1}) &= \log \mathcal{IW}(Q_0; \nu_0, V_0) + \sum_{k=0}^{K-2} \log p(Q_{k+1}|Q_k) \\
&+ \sum_{k=0}^{K-1} \underset{q_x^{(i)}}{\mathbb{E}}[\log \mathcal{N}(x_{k+1}|A_k x_k, Q_k)] + c \qquad (11)\\
&= \log \mathcal{IW}(Q_0; \nu_0, V_0) + \sum_{k=0}^{K-2} \log p(Q_{k+1}|Q_k) \\
&- \frac{1}{2}\sum_{k=0}^{K-1} \underset{q_x^{(i)}}{\mathbb{E}}\left[\mathrm{Tr}\left(Q_k^{-1}(x_{k+1} - A_k x_k)(x_{k+1} - A_k x_k)^{\mathrm{T}}\right)\right] \\
&- \frac{1}{2}\sum_{k=0}^{K-1} \log |Q_k| + c. \qquad (12)
\end{aligned}
$$

Taking the exponential of both sides, we get

$$
q_Q^{(i+1)}(Q_{0:K-1}) \propto \mathcal{IW}(Q_0; \nu_0, V_0) \prod_{k=0}^{K-2} p(Q_{k+1}|Q_k) \prod_{k=0}^{K-1} L_{Q,k}^{(i+1)}(Q_k) \qquad (13)
$$

where

$$L_{Q,k}^{(i+1)}(Q_k) \triangleq |Q_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}\left(Q_k^{-1} \underset{q_x^{(i)}}{\mathbb{E}}\left[(x_{k+1}-A_k x_k)(x_{k+1}-A_k x_k)^{\mathrm{T}}\right]\right)\right) \tag{14}$$

for $k = 0, \ldots, K-1$. Notice that the posterior density given in (13) corresponds to a smoothing problem with the following Markov model.

$$Q_0 \sim \mathcal{IW}(Q_0; \nu_0, V_0) \tag{15a}$$

$$Q_{k+1}|Q_k \sim p(Q_{k+1}|Q_k), \quad k = 0, \ldots, K-2 \tag{15b}$$

$$Z_{Q,k}^{(i+1)} \sim p(Z_{Q,k}^{(i+1)}|Q_k) \triangleq L_{Q,k}^{(i+1)}(Q_k), \quad k = 0, \ldots, K-1 \tag{15c}$$

where $Z_{Q,k}^{(i+1)}$ are some pseudo-measurements having the pseudo-likelihood $L_{Q,k}^{(i+1)}(\cdot)$. Since the problem is a standard smoothing problem for a Markov model, it can be solved using a forward-backward recursion with the following descriptions.

- **Forward Recursion:**

$$q_{Q,0|-1}^{(i+1)}(Q_0) = \mathcal{IW}(Q_0; \nu_0, V_0), \tag{16a}$$

$$q_{Q,k|k}^{(i+1)}(Q_k) \propto L_{Q,k}^{(i+1)}(Q_k) q_{Q,k|k-1}^{(i+1)}(Q_k), \tag{16b}$$

$$q_{Q,k+1|k}^{(i+1)}(Q_{k+1}) = \int p(Q_{k+1}|Q_k) q_{Q,k|k}^{(i+1)}(Q_k)\, \mathrm{d}Q_k. \tag{16c}$$

- **Backward Recursion:**

$$q_{Q,k|K}^{(i+1)}(Q_k) = \int q_{Q,k|k+1,k}^{(i+1)}(Q_k|Q_{k+1}, Z_{Q,0:k}^{(i+1)}) q_{Q,k|K}^{(i+1)}(Q_{k+1})\, \mathrm{d}Q_{k+1} \tag{17}$$

  where

$$q_{Q,k|k+1,k}^{(i+1)}(Q_k|Q_{k+1}, Z_{Q,0:k}^{(i+1)}) = \frac{p(Q_{k+1}|Q_k) q_{Q,k|k}^{(i+1)}(Q_k)}{q_{Q,k+1|k}^{(i+1)}(Q_{k+1})}. \tag{18}$$

  Note here that the given condition K in the smoothed density $q_{Q,k|K}^{(i+1)}(\cdot)$ pertains to $y_{0:K}$ in the original smoothing problem which corresponds to the pseudo-measurements $Z_{Q,0:K-1}^{(i+1)}$ in the artificial problem (15).

The forward recursion starts with the prior density $q_{Q,0|-1}^{(i+1)}(Q_0) = \mathcal{IW}(Q_0; \nu_0, V_0)$ which is an inverse Wishart density. Suppose now that the intermediate predicted density $q_{Q,k|k-1}^{(i+1)}(\cdot)$ is inverse Wishart in the following form.

$$q_{Q,k|k-1}^{(i+1)}(Q_k) = \mathcal{IW}\left(Q_k; \nu_{k|k-1}, V_{k|k-1}^{(i+1)}\right) \tag{19}$$

Thanks to the form of the pseudo-likelihood function (14), when $q_{Q,k|k-1}^{(i+1)}(\cdot)$ in (19) is substituted into the update expression (16b), the posterior $q_{Q,k|k}^{(i+1)}(\cdot)$ becomes also inverse Wishart as given below.

$$q_{Q,k|k}^{(i+1)}(Q_k) = \mathcal{IW}\left(Q_k; \nu_{k|k}, V_{k|k}^{(i+1)}\right) \tag{20}$$

where

$$V_{k|k}^{(i+1)} = V_{k|k-1}^{(i+1)} + \underset{q_x^{(i)}}{\mathbb{E}}[(x_{k+1} - A_k x_k)(x_{k+1} - A_k x_k)^{\mathrm{T}}], \tag{21a}$$

$$\nu_{k|k} = \nu_{k|k-1} + 1. \tag{21b}$$

When the posterior (20) is substituted into the prediction update expression (16c), thanks to the Beta-Bartlett transition density whose prediction updates are given by (1), the predicted density $q_{Q,k+1|k}^{(i+1)}(\cdot)$ turns out to be also inverse Wishart as given below.

$$q_{Q,k+1|k}^{(i+1)}(Q_k) = \mathcal{IW}\left(Q_{k+1}; \nu_{k+1|k}, V_{k+1|k}^{(i+1)}\right) \tag{22}$$

where

$$V_{k+1|k}^{(i+1)} = \lambda_Q V_{k|k}^{(i+1)}, \tag{23a}$$

$$\nu_{k+1|k} = \lambda_Q \nu_{k|k} + (1 - \lambda_Q)(2n_x + 2). \tag{23b}$$

As a result, via induction, all forward predicted and posterior densities are inverse Wishart.

The backward recursion starts with the final posterior density which is inverse Wishart as discussed above, i.e., we have

$$q_{Q,K-1|K}^{(i+1)}(Q_{K-1}) = \mathcal{IW}\left(Q_{K-1}; \nu_{K-1|K}, V_{K-1|K}^{(i+1)}\right). \tag{24}$$

Note here again that the given condition K in the smoothed density $q_{Q,K-1|K}^{(i+1)}(\cdot)$ and its parameters $\nu_{K-1|K}$, $V_{K-1|K}^{(i+1)}$ pertains to $y_{0:K}$ in the original smoothing problem which corresponds to the pseudo-measurements $Z_{Q,0:K-1}^{(i+1)}$ in the artificial problem (15). Suppose now that an intermediate smoothed density $q_{Q,k+1|K}^{(i+1)}(\cdot)$ is inverse Wishart as given below.

$$q_{Q,k+1|K}^{(i+1)}(Q_{k+1}) = \mathcal{IW}\left(Q_{k+1}; \nu_{k+1|K}, V_{k+1|K}^{(i+1)}\right). \tag{25}$$

When the smoothed density (25) is substituted into the backward update expression (17), thanks to the Beta-Bartlett transition density whose backward smoothing updates are given by (2), the smoothed density $q_{Q,k|K}^{(i+1)}(\cdot)$ turns out to be also inverse Wishart as given below.

$$q_{Q,k|K}^{(i+1)}(Q_k) = \mathcal{IW}\left(Q_k; \nu_{k|K}, V_{k|K}^{(i+1)}\right) \tag{26}$$

where

$$V_{k|K}^{(i+1)} \leftarrow \left((1 - \lambda_Q)\left(V_{k|k}^{(i+1)}\right)^{-1} + \lambda_Q\left(V_{k+1|K}^{(i+1)}\right)^{-1}\right)^{-1}, \tag{27a}$$

$$\nu_{k|K} \leftarrow (1 - \lambda_Q)\nu_{k|k} + \lambda_Q \nu_{k+1|K}. \tag{27b}$$

As a result, via induction, all backward smoothing densities are inverse Wishart.

2.2.1 Summary

Combining (21), (23) and (27), the marginals for the approximate joint smoothing density $q_Q^{(i+1)}(Q_{0:K-1})$ in (13) can be found as the following inverse Wishart density

$$q_Q^{(i+1)}(Q_k) = \mathcal{IW}\left(Q_k; \nu_{k|K}, V_{k|K}^{(i+1)}\right), \tag{28}$$

whose parameters can be computed using

$$V_{0|0}^{(i+1)} = V_0 + \mathop{\mathbb{E}}_{q_x^{(i)}}[(x_1 - A_0 x_0)(x_1 - A_0 x_0)^{\mathrm{T}}], \tag{29a}$$

$$\nu_{0|0} = \nu_0 + 1. \tag{29b}$$

for $k = 0$ along with the forward (filtering) recursion given by

$$V_{k|k}^{(i+1)} = \lambda_Q V_{k-1|k-1}^{(i+1)} + \mathop{\mathbb{E}}_{q_x^{(i)}}[(x_{k+1} - A_k x_k)(x_{k+1} - A_k x_k)^{\mathrm{T}}], \tag{30a}$$

$$\nu_{k|k} = \lambda_Q \nu_{k-1|k-1} + (1 - \lambda_Q)(2n_x + 2) + 1, \tag{30b}$$

for $1 \leq k \leq K-1$, followed by a backward (smoothing) recursion given by

$$V_{k|K}^{(i+1)} \leftarrow \left((1-\lambda_Q)\left(V_{k|k}^{(i+1)}\right)^{-1} + \lambda_Q\left(V_{k+1|K}^{(i+1)}\right)^{-1}\right)^{-1}, \tag{31a}$$

$$\nu_{k|K} \leftarrow (1-\lambda_Q)\nu_{k|k} + \lambda_Q \nu_{k+1|K}. \tag{31b}$$

for $0 \leq k \leq K-2$.

2.2.2 Fixed Parameter Case

When Q_k is a fixed parameter i.e., $\lambda_Q = 1$ or $Q_k = Q_{k-1}$, the forward filtering recursion (30) takes the form

$$V_{k|k}^{(i+1)} = V_{k-1|k-1}^{(i+1)} + \mathop{\mathbb{E}}_{q_x^{(i)}}[(x_{k+1} - A_k x_k)(x_{k+1} - A_k x_k)^{\mathrm{T}}], \tag{32a}$$

$$\nu_{k|k} = \nu_{k-1|k-1} + 1, \tag{32b}$$

and the backward smoothing recursion (31) becomes

$$V_{k|K}^{(i+1)} = V_{k+1|K}^{(i+1)}, \tag{33a}$$

$$\nu_{k|K} = \nu_{k+1|K}. \tag{33b}$$

Hence, the smoothing posterior for Q_k is the same for all time instances and is given by

$$q_Q^{(i+1)}(Q_k) = \mathcal{IW}\left(Q_k; \nu, V^{(i+1)}\right), \tag{34}$$

whose parameters are

$$V^{(i+1)} = V_0 + \sum_{k=0}^{K-1} \mathop{\mathbb{E}}_{q_x^{(i)}}[(x_{k+1} - A_k x_k)(x_{k+1} - A_k x_k)^{\mathrm{T}}], \tag{35a}$$

$$\nu = \nu_0 + K. \tag{35b}$$

When the variational recursions converge, the expected value of Q_k can be obtained using the stationary values of V and ν via

$$\widehat{Q_k} \triangleq \mathbb{E}[Q_k] = \frac{V}{\nu - 2n_x - 2}. \tag{36}$$

2.3 Derivations for the approximate posterior $q_R^{(i+1)}(\cdot)$

Using (6c) and (8), $q_R^{(i+1)}(\cdot)$ is given as

$$\begin{aligned}
\log q_R^{(i+1)}(R_{0:K}) =& \log \mathcal{IW}(R_0; \mu_0, M_0) \\
&+ \sum_{k=0}^{K} \mathop{\mathbb{E}}_{q_x^{(i)}} [\log \mathcal{N}(y_k; C_k x_k, R_k)] + \sum_{k=0}^{K-1} \log p(R_{k+1}|R_k) \quad (37) \\
=& \log \mathcal{IW}(R_0; \mu_0, M_0) + \sum_{k=0}^{K-1} \log p(R_{k+1}|R_k) - \frac{1}{2} \sum_{k=0}^{K} \log |R_k| \\
&- \frac{1}{2} \sum_{k=0}^{K} \mathop{\mathbb{E}}_{q_x^{(i)}} [\mathrm{Tr}\left(R_k^{-1}(y_k - C_k x_k)(y_k - C_k x_k)^{\mathrm{T}}\right)] + c. \quad (38)
\end{aligned}$$

Taking the exponential of both sides, we get

$$q_R^{(i+1)}(R_{0:K}) \propto \mathcal{IW}(R_0; \mu_0, M_0) \prod_{k=0}^{K-1} p(R_{k+1}|R_k) \prod_{k=0}^{K} L_{R,k}^{(i+1)}(R_k) \quad (39)$$

where

$$L_{R,k}^{(i+1)}(R_k) \triangleq |R_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathrm{Tr}\left(R_k^{-1} \mathop{\mathbb{E}}_{q_x^{(i)}} [(y_k - C_k x_k)(y_k - C_k x_k)^{\mathrm{T}}] \right) \right). \quad (40)$$

for $k = 0, \ldots, K$. Notice that the posterior density given in (39) corresponds to a smoothing problem with the following Markov model.

$$\begin{aligned}
R_0 &\sim \mathcal{IW}(R_0; \mu_0, M_0) && (41a) \\
R_{k+1}|R_k &\sim p(R_{k+1}|R_k), \quad k = 0, \ldots, K-1 && (41b) \\
Z_{R,k}^{(i+1)} &\sim p(Z_{R,k}^{(i+1)}|R_k) \triangleq L_{R,k}^{(i+1)}(R_k), \quad k = 0, \ldots, K && (41c)
\end{aligned}$$

where $Z_{R,k}^{(i+1)}$ are some pseudo-measurements having the pseudo-likelihood $L_{R,k}^{(i+1)}(\cdot)$. Since the problem is a standard smoothing problem for a Markov model, it can be solved using a forward-backward recursion. The rest of the derivation follows exactly the same steps as those in Section 2.2 and therefore are not repeated here. A summary of the results is given in the next subsection.

2.3.1 Summary

The marginals for the approximate joint smoothing density $q_R^{(i+1)}(R_{0:K})$ in (39) can be found as the following inverse Wishart density.

$$q_R^{(i+1)}(R_k) = \mathcal{IW}\left(R_k; \mu_{k|K}, M_{k|K}^{(i+1)}\right), \quad (42)$$

whose parameters can be computed using

$$M_{0|0}^{(i+1)} = M_0 + \mathop{\mathbb{E}}_{q_x^{(i)}} [(y_k - C_k x_k)(y_k - C_k x_k)^{\mathrm{T}}], \quad (43a)$$

$$\mu_{0|0} = \mu_0 + 1, \quad (43b)$$

for $k=0$ along with the forward (filtering) recursion given by

$$M_{k|k}^{(i+1)} = \lambda_R M_{k-1|k-1}^{(i+1)} + \mathop{\mathbb{E}}_{q_x^{(i)}}[(y_k - C_k x_k)(y_k - C_k x_k)^{\mathrm{T}}], \tag{44a}$$

$$\mu_{k|k} = \lambda_R \mu_{k-1|k-1} + (1-\lambda_R)(2n_y+2)+1, \tag{44b}$$

for $1 \le k \le K$, followed by a backwards (smoothing) recursion given by

$$M_{k|K}^{(i+1)} \leftarrow \left((1-\lambda_R)\left(M_{k|k}^{(i+1)}\right)^{-1} + \lambda_R \left(M_{k+1|K}^{(i+1)}\right)^{-1}\right)^{-1}, \tag{45a}$$

$$\mu_{k|K} \leftarrow (1-\lambda_R)\mu_{k|k} + \lambda_R \mu_{k+1|K}. \tag{45b}$$

for $0 \le k \le K-1$.

2.3.2 Fixed Parameter Case

When R_k is a fixed parameter i.e., $\lambda_R = 1$ or $R_k = R_{k-1}$, the forward filtering recursion (44) takes the form

$$M_{k|k}^{(i+1)} = M_{k-1|k-1}^{(i+1)} + \mathop{\mathbb{E}}_{q_x^{(i)}}[(y_k - C_k x_k)(y_k - C_k x_k)^{\mathrm{T}}], \tag{46a}$$

$$\mu_{k|k} = \mu_{k-1|k-1} + 1, \tag{46b}$$

and the backwards smoothing recursion (45) becomes

$$M_{k|K}^{(i+1)} = M_{k+1|K}^{(i+1)}, \tag{47a}$$

$$\mu_{k|K} = \mu_{k+1|K}. \tag{47b}$$

Hence, the smoothing posterior for R_k is the same for all time instances and is given by

$$q_R^{(i+1)}(R_k) = \mathcal{IW}\left(R_k; \mu, M^{(i+1)}\right), \tag{48}$$

whose parameters are given by

$$M^{(i+1)} = M_0 + \sum_{k=0}^{K} \mathop{\mathbb{E}}_{q_x^{(i)}}[(y_k - C_k x_k)(y_k - C_k x_k)^{\mathrm{T}}], \tag{49a}$$

$$\mu = \mu_0 + K + 1. \tag{49b}$$

When the variational recursions converge, the expected value of R_k can be obtained using the stationary values of M and μ via

$$\widehat{R_k} \triangleq \mathbb{E}[R_k] = \frac{M}{\mu - 2n_y - 2}. \tag{50}$$

2.4 Calculation of the expected values

Now we can calculate the expected values needed for the iterations in Sections 2.1 to 2.3. The approximate distribution of the random matrices Q_k and R_k

is inverse Wishart. Therefore their inverses are Wishart distributed and their expectations are given by

$$\mathop{\mathbb{E}}_{q_Q^{(i)}}[Q_k^{-1}] = (\nu_{k|K} - n_x - 1)(V_{k|K}^{(i)})^{-1}, \tag{51}$$

$$\mathop{\mathbb{E}}_{q_R^{(i)}}[R_k^{-1}] = (\mu_{k|K} - n_y - 1)(M_{k|K}^{(i)})^{-1}. \tag{52}$$

At each recursion of the algorithm, the RTS smoother provides the approximate joint posterior for $p(x_{k+1}, x_k|y_{0:K})$ denoted by $q_x^{(i)}(x_{k+1}, x_k)$ and parametrized as,

$$q_x^{(i)}(x_{k+1}, x_k) = \mathcal{N}\left(\begin{bmatrix} x_{k+1} \\ x_k \end{bmatrix}; \begin{bmatrix} m_{k+1|K}^{(i)} \\ m_{k|K}^{(i)} \end{bmatrix}, \begin{bmatrix} P_{k+1|K}^{(i)} & P_{k+1,k|K}^{(i)} \\ P_{k,k+1|K}^{(i)} & P_{k|K}^{(i)} \end{bmatrix}\right). \tag{53}$$

Using (53) we can calculate the following expected values

$$\begin{aligned} \mathop{\mathbb{E}}_{q_x^{(i)}}[(y_k - Cx_k)(y_k - Cx_k)^{\mathrm{T}}] &= CP_{k|K}^{(i)}C^{\mathrm{T}} \\ &\quad + (y_k - Cm_{k|K}^{(i)})(y_k - Cm_{k|K}^{(i)})^{\mathrm{T}}, \end{aligned} \tag{54}$$

$$\begin{aligned} \mathop{\mathbb{E}}_{q_x^{(i)}}[(x_{k+1} - A_k x_k)(x_{k+1} - A_k x_k)^{\mathrm{T}}] &= P_{k+1|K}^{(i)} \\ + A_k P_{k|K}^{(i)} A_k^{\mathrm{T}} - P_{k+1,k|K}^{(i)} A_k^{\mathrm{T}} &- A_k P_{k,k+1|K}^{(i)} \\ + (m_{k+1|K}^{(i)} - A_k m_{k|K}^{(i)})&(m_{k+1|K}^{(i)} - A_k m_{k|K}^{(i)})^{\mathrm{T}}. \end{aligned} \tag{55}$$

3 Comparison with Expectation Maximization

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

$$x_{k+1} = Ax_k + w_k, \qquad w_k \overset{iid}{\sim} \mathcal{N}(w_k; 0, Q), \tag{56a}$$

$$y_k = Cx_k + v_k, \qquad v_k \overset{iid}{\sim} \mathcal{N}(v_k; 0, R), \tag{56b}$$

where $\{x_k \in \mathbb{R}^{n_x} | 0 \le k \le K\}$ is the state trajectory, also denoted as $x_{0:K}$; $\{y_k \in \mathbb{R}^{n_y} | 0 \le k \le K\}$ is the measurement sequence, denoted in more compact form as $y_{0:K}$; $A \in \mathbb{R}^{n_x \times n_x}$ and $C \in \mathbb{R}^{n_y \times n_x}$ are known state transition and measurement matrices, respectively; $\{w_k \in \mathbb{R}^{n_x} | 0 \le k \le K-1\}$ and $\{v_k \in \mathbb{R}^{n_y} | 0 \le k \le K\}$ are mutually independent white Gaussian noise sequences. The initial state x_0 is assumed to have a Gaussian prior, i.e., $p(x_0) = \mathcal{N}(x_0; m_0, P_0)$. Q and R are the unknown (deterministic) fixed positive definite process noise and measurement noise covariance matrices.

Expectation-maximization (EM) [6] method can be used as in [7–10] to compute the maximum likelihood (ML) estimate of the noise covariance matrices. In the E (Expectation) step of the EM algorithm the conditional expectation of the joint log-likelihood is computed using the last estimates of the unknown parameters $Q^{(i)}$ and $R^{(i)}$ as in

$$\mathcal{Q} = \mathbb{E}\left[\log p(x_{0:K}, y_{0:K}) \,\middle|\, y_{0:K}, R^{(i)}, Q^{(i)}\right] \tag{57}$$

where

$$\begin{aligned}
\log p(x_{0:K}, y_{0:K}) = & \log \mathcal{N}(x_0; m_0, P_0) - \frac{K+1}{2} \log |R| \\
& - \frac{1}{2} \sum_{k=0}^{K} \mathrm{Tr}\left(R^{-1}(y_k - Cx_k)(y_k - Cx_k)^{\mathrm{T}}\right) - \frac{K}{2} \log |Q| \\
& - \frac{1}{2} \sum_{k=0}^{K-1} \mathrm{Tr}\left(Q^{-1}(x_{k+1} - Ax_k)(x_{k+1} - Ax_k)^{\mathrm{T}}\right) + c.
\end{aligned} \tag{58}$$

Therefore,

$$\begin{aligned}
\mathcal{Q} = & -\frac{1}{2} \mathbb{E}[(x_0 - m_0)P_0^{-1}(x_0 - m_0)^{\mathrm{T}} + \log |P_0|] \\
& - \frac{K+1}{2} \log |R| - \frac{1}{2} \mathrm{Tr}\left(R^{-1} \sum_{k=0}^{K} \mathbb{E}[(y_k - Cx_k)(y_k - Cx_k)^{\mathrm{T}} | y_{0:K}]\right) \\
& - \frac{K}{2} \log |Q| - \frac{1}{2} \mathrm{Tr}\left(Q^{-1} \sum_{k=0}^{K-1} \mathbb{E}[(x_{k+1} - Ax_k)(x_{k+1} - Ax_k) | y_{0:K}]\right) + c,
\end{aligned} \tag{59}$$

where the expectations can be computed using the posterior of the RTS smoother which uses $Q^{(i)}$ and $R^{(i)}$.

Now the expressions for the M (Maximization) step of the EM algorithm can be computed. Taking the partial derivatives of $\mathcal{Q}$ with respect to Q^{-1} and

R^{-1}, we get[1]

$$\frac{\partial \mathcal{Q}}{\partial R^{-1}} = \frac{K+1}{2} R - \frac{1}{2} \sum_{k=0}^{K} \mathbb{E}[(y_k - Cx_k)(y_k - Cx_k)^{\mathrm{T}} | y_{0:K}], \tag{60a}$$

$$\frac{\partial \mathcal{Q}}{\partial Q^{-1}} = \frac{K}{2} Q - \frac{1}{2} \sum_{k=0}^{K-1} \mathbb{E}[(x_{k+1} - Ax_k)(x_{k+1} - Ax_k)^{\mathrm{T}} | y_{0:K}], \tag{60b}$$

and equating the results to zero gives

$$R^{(i+1)} = \frac{1}{K+1} \sum_{k=0}^{K} \mathbb{E}[(y_k - Cx_k)(y_k - Cx_k)^{\mathrm{T}} | y_{0:K}, R^{(i)}, Q^{(i)}], \tag{61a}$$

$$Q^{(i+1)} = \frac{1}{K} \sum_{k=0}^{K-1} \mathbb{E}[(x_{k+1} - Ax_k)(x_{k+1} - Ax_k)^{\mathrm{T}} | y_{0:K}, R^{(i)}, Q^{(i)}] \tag{61b}$$

where, $R^{(i+1)}$ and $Q^{(i+1)}$ are $(i+1)$th estimates for R and Q, respectively. These estimates can be used in the E step of the EM algorithm using the RTS smoother again until convergence.

Note that the iteration of the EM algorithm for the estimation of noise covariances is not identical to the iterations of the VB algorithm but they have "close" functional forms. In particular, the RTS smoother in the EM solution uses the last estimates $R^{(i)}$ and $Q^{(i)}$ while calculating the expected values in (61) while the RTS smoother in the VB solution uses the harmonic mean of these random matrices i.e., $\widetilde{Q} \triangleq (\mathbb{E}_{q_Q}[Q^{-1}])^{-1}$ for the process noise covariance and $\widetilde{R} \triangleq (\mathbb{E}_{q_R}[R^{-1}])^{-1}$ for the measurement noise covariance for calculating the similar expected values, where

$$(\mathop{\mathbb{E}}_{q_Q}[Q^{-1}])^{-1} = V/(\nu - n_x - 1), \tag{62}$$

$$(\mathop{\mathbb{E}}_{q_R}[R^{-1}])^{-1} = M/(\mu - n_y - 1). \tag{63}$$

Note that the matrices (62) and (63) are strictly larger than (in a positive definite sense) the VB estimates given in (36) and (50), respectively, which is a result of the Jensen's inequality. Hence VB does not use the last estimates in calculating the involved expected values.

Expressions for V and ν are given in (35) and expressions for M and μ are given in (49). When V_0 and M_0 are small and

$$\nu - n_x - 1 = K \qquad \Rightarrow \qquad \nu_0 = n_x + 1, \tag{64}$$

$$\mu - n_y - 1 = K + 1 \qquad \Rightarrow \qquad \mu_0 = n_y + 1, \tag{65}$$

the noise covariances used in the RTS smoother of EM solution and VB solution will coincide. However, such quantities for the initial degree of freedom of inverse Wishart distributions are not valid since the degree of freedom should be always greater than twice the dimension of the random matrix in the inverse Wishart distribution, see [11, page 111]. Hence VB recursions to do not reduce to EM recursions for some initial (hyper-) parameter selections.

[1]Since Q and R are positive definite, taking the partial derivatives with respect to Q^{-1} and R^{-1}; equating the results to zero and solving for Q and R is equivalent to taking the partial derivatives with respect to Q and R; equating the results to zero and solving for Q and R.

References

[1] T. Ardeshiri, E. Özkan, U. Orguner, and F. Gustafsson, "Approximate Bayesian smoothing with unknown process and measurement noise covariances," *Signal Processing Letters*, submitted 2015.

[2] C. M. Carvalho and M. West, "Dynamic matrix-variate graphical models," *Bayesian Anal*, vol. 2, pp. 69–98, 2007.

[3] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2007.

[4] 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.

[5] 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.

[6] 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

[7] 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.

[8] 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

[9] Z. Ghahramani and G. E. Hinton, "Parameter estimation for linear dynamical systems," Department of Computer Science, University of Toronto, Tech. Rep., 1996.

[10] S. Särkkä, *Bayesian Filtering and Smoothing*. New York, NY, USA: Cambridge University Press, 2013.

[11] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*. Boca Raton, FL: Chapman & Hall/CRC, 2000.

Avdelning, Institution Division, Department Division of Automatic Control Department of Electrical Engineering	**Datum** Date 2015-08-30

Språk Language	**Rapporttyp** Report category	
☐ Svenska/Swedish	☐ Licentiatavhandling	**ISBN** —
☒ Engelska/English	☐ Examensarbete	**ISRN** —
	☐ C-uppsats	**Serietitel och serienummer** Title of series, numbering
	☐ D-uppsats	**ISSN** 1400-3902
☐ ______	☒ Övrig rapport	LiTH-ISY-R-3086
	☐ ______	

URL för elektronisk version

http://www.control.isy.liu.se

Titel Title: Variational Iterations for Smoothing with Unknown Process and Measurement Noise Covariances

Författare Author: Tohid Ardeshiri, Emre Özkan, Umut Orguner, Fredrik Gustafsson

Sammanfattning
Abstract

In this technical report, some derivations for the smoother proposed in [1] are presented. More specifically, the derivations for the cyclic iteration needed to solve the variational Bayes smoother for linear state-space models with unknown process and measurement noise covariances in [1] are presented. Further, the variational iterations are compared with iterations of the Expectation Maximization (EM) algorithm for smoothing linear state-space models with unknown noise covariances.

Nyckelord
Keywords: Adaptive smoothing, variational Bayes, sensor calibration, Rauch-Tung-Striebel smoother, Kalman filtering, noise covariance