Variational Iterations for Filtering and Smoothing with skew-t measurement noise

Tohid Ardeshiri, Henri Nurminen, Robert Piché, Fredrik Gustafsson

Division of Automatic Control

E-mail: tohid@isy.liu.se, henri.nurminen@tut.fi, robert.piche@tut.fi, fredrik@isy.liu.se

19th March 2015

Report no.: LiTH-ISY-R-3083
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 filter and smoother proposed in [1] are presented. More specifically, the derivations for the cyclic iteration needed to solve the variational Bayes filter and smoother for state space models with skew t likelihood proposed in [1] are presented.

Keywords: skew t-distribution, skewness, t-distribution, robust filtering, Kalman filter, RTS smoother, variational Bayes

Variational Iterations for Filtering and Smoothing with skew-t measurement noise

Tohid Ardeshiri, Henri Nurminen*, Robert Piché*, and Fredrik Gustafsson

2015-03-24

Abstract

In this technical report, some derivations for the filter and smoother proposed in [1] are presented. More specifically, the derivations for the cyclic iteration needed to solve the variational Bayes filter and smoother for state space models with skew t likelihood proposed in [1] are presented.

1 Problem formulation

A Bayesian filter and a Bayesian smoother using the variational Bayes method for normal prior and skew-t measurement noise are given in [1]. These algorithms compute an approximation of the filtering distribution $p(x_k|y_{1:k})$ and smoothing distribution $p(x_k|y_{1:K})$, respectively. Here, we derive the expectations needed for the cyclic iterations of the variational Bayes smoother which approximates the joint smoothing posterior density given in [1]. The joint smoothing posterior density

$$
\begin{aligned}
&p(x_{1:K}, u_{1:K}, \Lambda_{1:K}|y_{1:K}) \propto p(x_{1:K}, u_{1:K}, \Lambda_{1:K}, y_{1:K}) && (1)\\
&= p(x_1) \prod_{l=1}^{K-1} p(x_{l+1}|x_l) \prod_{k=1}^{K} p(y_k|x_k, u_k, \Lambda_k) p(u_k|\Lambda_k) p(\Lambda_k) && (2)\\
&= \mathcal{N}(x_1; x_{1|0}, P_{1|0}) \prod_{l=1}^{K-1} \mathcal{N}(x_{l+1}; Ax_l, Q)\\
&\times \prod_{k=1}^{K} \left\{ \mathcal{N}(y_k; Cx_k + \Delta u_k, \Lambda_k^{-1} R) \mathcal{N}_+(u_k; 0, \Lambda_k^{-1}) \prod_{i=1}^{n_y} \mathcal{G}\left([\Lambda_k]_{ii}; \frac{\nu_i}{2}, \frac{\nu_i}{2} \right) \right\} && (3)
\end{aligned}
$$

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

$$
p(x_{1:K}, u_{1:K}, \Lambda_{1:K}|y_{1:K}) \approx q_x(x_{1:K}) q_u(u_{1:K}) q_\Lambda(\Lambda_{1:K}). \quad (4)
$$

*H. Nurminen and R. Piché are with the Department of Automation Science and Engineering, Tampere University of Technology (TUT), PO Box 692, 33101 Tampere, Finland (e-mails: henri.nurminen@tut.fi, robert.piche@tut.fi). H. Nurminen receives funding from TUT Graduate School, Finnish Doctoral Programme in Computational Sciences (FICS), and the Foundation of Nokia Corporation.

The analytical solutions for $\hat{q}_x$, $\hat{q}_u$ and $\hat{q}_\Lambda$ can be obtained by cyclic iteration of

$$\log q_x(x_{1:K}) \leftarrow \underset{q_u q_\Lambda}{\mathbb{E}} [\log p(y_{1:K}, x_{1:K}, u_{1:K}, \Lambda_{1:K})] + c_x \tag{5a}$$

$$\log q_u(u_{1:K}) \leftarrow \underset{q_x q_\Lambda}{\mathbb{E}} [\log p(y_{1:K}, x_{1:K}, u_{1:K}, \Lambda_{1:K})] + c_u \tag{5b}$$

$$\log q_\Lambda(\Lambda_{1:K}) \leftarrow \underset{q_x q_u}{\mathbb{E}} [\log p(y_{1:K}, x_{1:K}, u_{1:K}, \Lambda_{1:K})] + c_\Lambda \tag{5c}$$

where the expected values on the right hand sides of (5) are taken with respect to the current q_x, q_u and q_Λ and c_x, c_u and c_Λ are constants with respect to the variables x_k, u_k and Λ_k, respectively [2, Chapter 10] [3].

2 Derivations for the smoother

In sections 2.1, 2.3 and 2.2 the derivations for the variational solution (5) are given. For brevity all constant values are denoted by c in the derivation. The logarithm of the joint smoothing posterior which is needed for the derivations is given as

$$\begin{aligned}\log p(x_{1:K}, u_{1:K}, \Lambda_{1:K}, y_{1:K}) = &\log \mathcal{N}(x_1; x_{1|0}, P_{1|0}) + \sum_{l=1}^{K-1} \log \mathcal{N}(x_{l+1}; Ax_l, Q) \\ &+ \sum_{k=1}^{K} \left\{ \log \mathcal{N}(y_k; Cx_k + \Delta u_k, \Lambda_k^{-1} R) + \log \mathcal{N}_+(u_k; 0, \Lambda_k^{-1}) \right\} \\ &+ \sum_{k=1}^{K} \sum_{i=1}^{n_y} \log \mathcal{G}\left([\Lambda_k]_{ii}; \frac{\nu_i}{2}, \frac{\nu_i}{2} \right)\end{aligned} \tag{6}$$

2.1 Derivations for q_x

Using equation (5a) we obtain

$$\log q_x(x_{1:K}) = \log \mathcal{N}(x_1; x_{1|0}, P_{1|0}) + \sum_{l=1}^{K-1} \log \mathcal{N}(x_{l+1}; Ax_l, Q)$$

$$+ \sum_{k=1}^{K} \underset{q_u q_\Lambda}{\mathbb{E}} \left[\log \mathcal{N}(y_k; Cx_k + \Delta u_k, \Lambda_k^{-1} R)\right] + c \tag{7}$$

$$= \log \mathcal{N}(x_1; x_{1|0}, P_{1|0}) + \sum_{l=1}^{K-1} \log \mathcal{N}(x_{l+1}; Ax_l, Q)$$

$$- \frac{1}{2} \sum_{k=1}^{K} \underset{q_u q_\Lambda}{\mathbb{E}} \left[(y_k - Cx_k - \Delta u_k)^{\mathrm{T}} R^{-1} \Lambda_k (y_k - Cx_k - \Delta u_k)\right] + c \tag{8}$$

$$= \log \mathcal{N}(x_1; x_{1|0}, P_{1|0}) + \sum_{l=1}^{K-1} \log \mathcal{N}(x_{l+1}; Ax_l, Q)$$

$$- \frac{1}{2} \sum_{k=1}^{K} (y_k - Cx_k - \Delta \overline{u_k})^{\mathrm{T}} R^{-1} \overline{\Lambda_k} (y_k - Cx_k - \Delta \overline{u_k}) + c \tag{9}$$

$$= \log \mathcal{N}(x_1; x_{1|0}, P_{1|0}) + \sum_{l=1}^{K-1} \log \mathcal{N}(x_{l+1}; Ax_l, Q)$$

$$+ \sum_{k=1}^{K} \log \mathcal{N}(y_k - \Delta \overline{u_k}; Cx_k, \overline{\Lambda_k}^{-1} R) + c \tag{10}$$

where $\overline{u_k} \triangleq \mathbb{E}_{q_u}[u_k]$ and $\overline{\Lambda_k} \triangleq \mathbb{E}_{q_\Lambda}[\Lambda_k]$ are derived in sections 2.2 and 2.3, respectively. Hence, $\log q_x(x_{1:K})$ in (10) has the same form as logarithm of the joint posterior of a linear state-space model with measurements $\widetilde{y_k} \triangleq y_k - \Delta \overline{u_k}$ and Gaussian measurement noise covariance $\widetilde{R} \triangleq \overline{\Lambda_k}^{-1} R$. Therefore, $q_x(x_{1:K})$ can be computed using the Rauch-Tung-Striebel smoother's recursion [4]. The approximate marginal distribution of x_k turns out to be

$$q_x(x_k) = \mathcal{N}(x_k; x_{k|K}, P_{k|K}), \tag{11}$$

where expressions for $x_{k|K}$, $P_{k|K}$ are given in [1, Table I].

2.2 Derivations for q_u

Using equation (5b) we obtain

$$\log q_u(u_{1:K}) = \sum_{k=1}^{K} \mathop{\mathbb{E}}_{q_\Lambda q_x} \left[\log \mathcal{N}(y_k; Cx_k + \Delta u_k, \Lambda_k^{-1} R) + \log \mathcal{N}_+(u_k; 0, \Lambda_k^{-1})\right] + c. \tag{12}$$

Therefore, $q_u(u_{1:K}) = \prod_{k=1}^{K} q_u(u_k)$ where

$$\log q_u(u_k) = \mathop{\mathbb{E}}_{q_\Lambda q_x} \left[\log \mathcal{N}(y_k; Cx_k + \Delta u_k, \Lambda_k^{-1} R) + \log \mathcal{N}_+(u_k; 0, \Lambda_k^{-1})\right] + c \tag{13}$$

$$= -\frac{1}{2} \mathop{\mathbb{E}}_{q_\Lambda q_x} [(y_k - Cx_k - \Delta u_k)^{\mathrm{T}} R^{-1} \Lambda_k (y_k - Cx_k - \Delta u_k)] - \frac{1}{2} \mathop{\mathbb{E}}_{q_\Lambda} [u_k^{\mathrm{T}} \Lambda_k u_k] + c \tag{14}$$

$$= -\frac{1}{2} (y_k - C\overline{x_k} - \Delta u_k)^{\mathrm{T}} R^{-1} \overline{\Lambda_k} (y_k - C\overline{x_k} - \Delta u_k) - \frac{1}{2} u_k^{\mathrm{T}} \overline{\Lambda_k} u_k + c, \tag{15}$$

where, $\overline{x_k} \triangleq \mathbb{E}_{q_x}[x_k] = x_{k|K}$. Therefore,

$$q_u(u_k) = \mathcal{N}_+(u_k; u_{k|K}, U_{k|K}) \tag{16}$$

where

$$u_{k|K} = K_u (y_k - Cx_{k|K}), \tag{17}$$

$$U_{k|K} = (I - K_u \Delta) \overline{\Lambda_k}^{-1}, \tag{18}$$

$$K_u = \overline{\Lambda_k}^{-1} \Delta (\Delta \overline{\Lambda_k}^{-1} \Delta + \overline{\Lambda_k}^{-1} R)^{-1} = \Delta (\Delta^2 + R)^{-1}. \tag{19}$$

The expectation $\overline{u_k}$ is needed in (10) and can be calculated using e.g., [5]. Note that the cumulative distribution function of univariate normal distribution (or some approximation of it) is required in the computation of the moments of the truncated normal distribution.

2.3 Derivations for q_Λ

Using equation (5c) we obtain

$$\log q_\Lambda(\Lambda_{1:K}) = \sum_{k=1}^{K} \left\{ \mathop{\mathbb{E}}_{q_u q_x} \left[\log \mathcal{N}(y_k; Cx_k + \Delta u_k, \Lambda_k^{-1} R) + \log \mathcal{N}_+(u_k; 0, \Lambda_k^{-1})\right] \right\} + \sum_{k=1}^{K} \sum_{i=1}^{n_y} \log \mathcal{G}\left([\Lambda_k]_{ii}; \frac{\nu_i}{2}, \frac{\nu_i}{2}\right) + c. \tag{20}$$

Therefore, $q_\Lambda(\Lambda_{1:K}) = \prod_{k=1}^{K} q_\Lambda(\Lambda_k)$ where

$$
\begin{aligned}
\log q_\Lambda(\Lambda_k) = & \mathop{\mathbb{E}}_{q_u q_x}\left[\log \mathcal{N}(y_k; Cx_k + \Delta u_k, \Lambda_k^{-1}R) + \log \mathcal{N}_+(u_k; 0, \Lambda_k^{-1})\right] \\
& + \sum_{i=1}^{n_y} \log \mathcal{G}\left([\Lambda_k]_{ii}; \frac{\nu_i}{2}, \frac{\nu_i}{2}\right) + c \qquad (21) \\
= & -\frac{1}{2}\mathop{\mathbb{E}}_{q_u q_x}[\mathrm{tr}(\Lambda_k R^{-1}(y_k - Cx_k - \Delta u_k)(y_k - Cx_k - \Delta u_k)^{\mathrm{T}})] \\
& - \frac{1}{2}\mathop{\mathbb{E}}_{q_u}[\mathrm{tr}(\Lambda_k u_k u_k^{\mathrm{T}})] + \sum_{i=1}^{n_y}\left(\frac{\nu_i}{2}\log[\Lambda_k]_{ii} - \frac{\nu_i}{2}[\Lambda_k]_{ii}\right) + c \qquad (22) \\
= & -\frac{1}{2}\mathop{\mathbb{E}}_{q_x}[\mathrm{tr}(\Lambda_k R^{-1}(y_k - Cx_k)(y_k - Cx_k)^{\mathrm{T}})] - \frac{1}{2}\mathop{\mathbb{E}}_{q_u}[\mathrm{tr}(\Lambda_k \Delta R^{-1}\Delta u_k u_k^{\mathrm{T}})] \\
& + \frac{1}{2}(y_k - C\overline{x_k})^{\mathrm{T}}\Lambda_k R^{-1}\Delta\overline{u_k} + \frac{1}{2}\overline{u_k}^{\mathrm{T}}\Delta\Lambda_k R^{-1}(y_k - C\overline{x_k}) \\
& - \frac{1}{2}\mathop{\mathbb{E}}_{q_u}[\mathrm{tr}(\Lambda_k u_k u_k^{\mathrm{T}})] + \sum_{i=1}^{n_y}\left(\frac{\nu_i}{2}\log[\Lambda_k]_{ii} - \frac{\nu_i}{2}[\Lambda_k]_{ii}\right) + c \qquad (23) \\
= & -\frac{1}{2}\mathrm{tr}(\Lambda_k R^{-1}((y_k - Cx_{k|K})(y_k - Cx_{k|K})^{\mathrm{T}} + CP_{k|K}C^{\mathrm{T}})]) \\
& - \frac{1}{2}\mathrm{tr}(\Lambda_k(\Delta R^{-1}\Delta + I)\mathop{\mathbb{E}}_{q_u}[u_k u_k^{\mathrm{T}}]) + \frac{1}{2}\mathrm{tr}(\Lambda_k R^{-1}\Delta\overline{u_k}(y_k - Cx_{k|K})^{\mathrm{T}}) \\
& + \frac{1}{2}\mathrm{tr}(\Lambda_k \Delta R^{-1}(y_k - Cx_{k|K})\overline{u_k}^{\mathrm{T}}) + \sum_{i=1}^{n_y}\left(\frac{\nu_i}{2}\log[\Lambda_k]_{ii} - \frac{\nu_i}{2}[\Lambda_k]_{ii}\right) + c \qquad (24) \\
= & \sum_{i=1}^{n_y}\left(\frac{\nu_i}{2}\log[\Lambda_k]_{ii} - \frac{\nu_i + [\Psi_k]_{ii}}{2}[\Lambda_k]_{ii}\right) + c, \qquad (25)
\end{aligned}
$$

where the commutative property of product of diagonal matrices Δ, Λ_k and R is used in several occasions and

$$
\begin{aligned}
\Psi_k = & R^{-1}((y_k - Cx_{k|K})(y_k - Cx_{k|K})^{\mathrm{T}} + CP_{k|K}C^{\mathrm{T}}) + (\Delta R^{-1}\Delta + I)\mathop{\mathbb{E}}_{q_u}[u_k u_k^{\mathrm{T}}] \\
& - R^{-1}\Delta\overline{u_k}(y_k - Cx_{k|K})^{\mathrm{T}} - \Delta R^{-1}(y_k - Cx_{k|K})\overline{u_k}^{\mathrm{T}}. \qquad (26)
\end{aligned}
$$

Therefore,

$$
q_\Lambda(\Lambda_k) = \prod_{i=1}^{n_y}\mathcal{G}\left([\Lambda_k]_{ii}; \frac{\nu_i}{2} + 1, \frac{\nu_i + [\Psi_k]_{ii}}{2}\right). \qquad (27)
$$

Note that only the diagonal elements of the matrix Ψ_k are required. As a consequence, provided that Δ and R are diagonal, only the diagonal elements of $\mathbb{E}[u_k u_k^{\mathrm{T}}]$ are required. These are second moments of univariate truncated normal distributions that can be computed using e.g., [5]. In the derivations above $\mathbb{E}_{q_\Lambda}[\Lambda_k]$ is required. The diagonal elements of $\mathbb{E}_{q_\Lambda}[\Lambda_k]$ are

$$
\mathop{\mathbb{E}}_{q_\Lambda}[[\Lambda_k]_{ii}] = \frac{\nu_i + 2}{\nu_i + [\Psi_k]_{ii}}. \qquad (28)
$$

3 Derivations for the Filter

The filtering recursions are similar to those of the smoother given in section 2. However, since the notation used in the filtering algorithm [1, Table II] is different from the notation used for smoothing algorithm, the derivation for the filter will be given separately.

Suppose that at time index k the measurement y_k is available, and the prediction PDF $p(x_k|y_{1:k-1})$ is

$$p(x_k|y_{1:k-1}) = \mathcal{N}(x_k; x_{k|k-1}, P_{k|k-1}). \tag{29}$$

Then, using Bayes' theorem the joint filtering posterior PDF can be written as

$$\begin{aligned}
p(x_k, u_k, \Lambda_k|y_{1:k}) \propto & p(y_k, x_k, u_k, \Lambda_k|y_{1:k-1}) && (30)\\
=& p(y_k|x_k, u_k, \Lambda_k)p(x_k|y_{1:k-1})p(u_k|\Lambda_k)p(\Lambda_k) && (31)\\
=& \mathcal{N}(y_k; Cx_k + \Delta u_k, \Lambda_k^{-1}R)\mathcal{N}(x_k; x_{k|k-1}, P_{k|k-1})\\
& \times \mathcal{N}_+(u_k; 0, \Lambda_k^{-1}) \prod_{i=1}^{n_y} \mathcal{G}\left([\Lambda_k]_{ii}; \frac{\nu_i}{2}, \frac{\nu_i}{2}\right). && (32)
\end{aligned}$$

This posterior is not analytically tractable. We seek an approximation in the form

$$p(x_k,u_k, \Lambda_k|y_{1:k}) \approx q_x(x_k)q_u(u_k)q_\Lambda(\Lambda_k). \tag{33}$$

In the VB approach, the Kullback-Leibler divergence (KLD) of the true posterior from the factorized approximation is minimized;

$$\hat{q}_x, \hat{q}_u, \hat{q}_\Lambda = \operatorname*{argmin}_{q_x, q_u, q_\Lambda} D_{\mathrm{KL}}(q_x(x_k)q_u(u_k)q_\Lambda(\Lambda_k)||p(x_k, u_k, \Lambda_k|y_{1:k}))$$

where $D_{\mathrm{KL}}(q(\cdot)||p(\cdot)) \triangleq \int q(x) \log \frac{q(x)}{p(x)}\, \mathrm{d}x$ is the KLD [6]. The analytical solutions for $\hat{q}_x$, $\hat{q}_u$ and $\hat{q}_\Lambda$ can be obtained by cyclic iteration of

$$\log q_x(x_k) \leftarrow \mathop{\mathbb{E}}_{q_u q_\Lambda} [\log p(y_k, x_k, u_k, \Lambda_k|y_{1:k-1})] + c_x \tag{34a}$$

$$\log q_u(u_k) \leftarrow \mathop{\mathbb{E}}_{q_x q_\Lambda} [\log p(y_k, x_k, u_k, \Lambda_k|y_{1:k-1})] + c_u \tag{34b}$$

$$\log q_\Lambda(\Lambda_k) \leftarrow \mathop{\mathbb{E}}_{q_x q_u} [\log p(y_k, x_k, u_k, \Lambda_k|y_{1:k-1})] + c_\Lambda \tag{34c}$$

where the expected values on the right hand sides of (34) are taken with respect to the current q_x, q_u and q_Λ and c_x, c_u and c_Λ are constants with respect to the variables x_k, u_k and Λ_k, respectively [2, Chapter 10] [3]. This recursion is convergent to a local optimum [2, Chapter 10]. In sections 3.1, 3.3 and 3.2 the derivations for the variational solution (34) are given. For brevity all constant values are denoted by c in the derivation. The logarithm of the joint filtering posterior which is needed for the derivations is given by

$$\begin{aligned}
\log p(y_k, x_k, u_k, \Lambda_k|y_{1:k-1}) = &-\frac{1}{2}(y_k - Cx_k - \Delta u_k)^{\mathrm{T}}\Lambda_k R^{-1}(y_k - Cx_k - \Delta u_k)\\
&-\frac{1}{2}(x_k - x_{k|k-1})^{\mathrm{T}}P_{k|k-1}^{-1}(x_k - x_{k|k-1})\\
&-\frac{1}{2}u_k^{\mathrm{T}}\Lambda_k u_k + \sum_{i=1}^{n_y}\frac{\nu_i}{2}\log[\Lambda_k]_{ii} - \frac{\nu_i}{2}[\Lambda_k]_{ii} + c.
\end{aligned} \tag{35}$$

3.1 Derivations for q_x

Using equation (34a) we obtain

$$\begin{aligned}\log q_x(x_k) =& -\frac{1}{2}\underset{q_u q_\Lambda}{\mathbb{E}}[(y_k - Cx_k - \Delta u_k)^{\mathrm{T}}\Lambda_k R^{-1}(y_k - Cx_k - \Delta u_k)] \\ & -\frac{1}{2}(x_k - x_{k|k-1})^{\mathrm{T}} P_{k|k-1}^{-1}(x_k - x_{k|k-1}) + c \quad (36) \\ =& -\frac{1}{2}(y_k - Cx_k - \Delta\overline{u_k})^{\mathrm{T}}\overline{\Lambda_k} R^{-1}(y_k - Cx_k - \Delta\overline{u_k}) \\ & -\frac{1}{2}(x_k - x_{k|k-1})^{\mathrm{T}} P_{k|k-1}^{-1}(x_k - x_{k|k-1}) + c, \quad (37)\end{aligned}$$

where $\overline{u_k} \triangleq \mathbb{E}_{q_u}[u_k]$ and $\overline{\Lambda_k} \triangleq \mathbb{E}_{q_\Lambda}[\Lambda_k]$ are derived in sections 3.2 and 3.3, respectively. Hence,

$$q_x(x_k) \propto \mathcal{N}(y_k - \Delta\overline{u_k}; Cx_k, \overline{\Lambda_k}^{-1}R)\mathcal{N}(x_k; x_{k|k-1}, P_{k|k-1}) \quad (38)$$

which can be computed using the Kalman filter's [7] measurement update. Therefore,

$$q_x(x_k) = \mathcal{N}(x_k; x_{k|k}, P_{k|k}) \quad (39)$$

where,

$$\begin{aligned} x_{k|k} &= x_{k|k-1} + K_x(y_k - Cx_{k|k-1} - \Delta\overline{u_k}), \quad (40) \\ P_{k|k} &= (I - K_x C)P_{k|k-1}, \quad (41) \\ K_x &= P_{k|k-1}C^{\mathrm{T}}(CP_{k|k-1}C^{\mathrm{T}} + \overline{\Lambda_k}^{-1}R)^{-1}. \quad (42)\end{aligned}$$

3.2 Derivations for q_u

Using equation (34b) we obtain

$$\begin{aligned}\log q_u(u_k) =& -\frac{1}{2}\underset{q_x q_\Lambda}{\mathbb{E}}[(y_k - Cx_k - \Delta u_k)^{\mathrm{T}}\Lambda_k R^{-1}(y_k - Cx_k - \Delta u_k)] \\ & -\frac{1}{2}\underset{q_\Lambda}{\mathbb{E}}[u_k^{\mathrm{T}}\Lambda_k u_k] + c \quad (43) \\ =& -\frac{1}{2}(y_k - C\overline{x_k} - \Delta u_k)^{\mathrm{T}}\overline{\Lambda_k} R^{-1}(y_k - C\overline{x_k} - \Delta u_k) \\ & -\frac{1}{2}u_k^{\mathrm{T}}\overline{\Lambda_k}u_k + c, \quad (44)\end{aligned}$$

where, $\overline{x_k} \triangleq \mathbb{E}_{q_x}[x_k] = x_{k|k}$. Therefore,

$$q_u(u_k) = \mathcal{N}_+(u_k; u_{k|k}, U_{k|k}) \quad (45)$$

where

$$\begin{aligned} u_{k|k} &= K_u(y_k - Cx_{k|k}), \quad (46) \\ U_{k|k} &= (I - K_u\Delta)\overline{\Lambda_k}^{-1}, \quad (47) \\ K_u &= \overline{\Lambda_k}^{-1}\Delta(\Delta\overline{\Lambda_k}^{-1}\Delta + \overline{\Lambda_k}^{-1}R)^{-1} = \Delta(\Delta^2 + R)^{-1}. \quad (48)\end{aligned}$$

The expectation $\overline{u_k}$ is needed in (40) and can be calculated using e.g., [5].

3.3 Derivations for q_Λ

Using equation (34c) we obtain

$$
\begin{aligned}
\log q_\Lambda(\Lambda_k) =& -\frac{1}{2}\mathop{\mathbb{E}}_{q_u q_x}[\operatorname{tr}(\Lambda_k R^{-1}(y_k - Cx_k - \Delta u_k)(y_k - Cx_k - \Delta u_k)^{\mathrm{T}})] \\
& - \frac{1}{2}\mathop{\mathbb{E}}_{q_u}[\operatorname{tr}(\Lambda_k u_k u_k^{\mathrm{T}})] + \sum_{i=1}^{n_y}\left(\frac{\nu_i}{2}\log[\Lambda_k]_{ii} - \frac{\nu_i}{2}[\Lambda_k]_{ii}\right) + c \qquad (49)\\
=& -\frac{1}{2}\mathop{\mathbb{E}}_{q_x}[\operatorname{tr}(\Lambda_k R^{-1}(y_k - Cx_k)(y_k - Cx_k)^{\mathrm{T}})] - \frac{1}{2}\mathop{\mathbb{E}}_{q_u}[\operatorname{tr}(\Lambda_k \Delta R^{-1}\Delta u_k u_k^{\mathrm{T}})] \\
& + \frac{1}{2}(y_k - C\overline{x_k})^{\mathrm{T}}\Lambda_k R^{-1}\Delta\overline{u_k} + \frac{1}{2}\overline{u_k}^{\mathrm{T}}\Delta\Lambda_k R^{-1}(y_k - C\overline{x_k}) \\
& - \frac{1}{2}\mathop{\mathbb{E}}_{q_u}[\operatorname{tr}(\Lambda_k u_k u_k^{\mathrm{T}})] + \sum_{i=1}^{n_y}\left(\frac{\nu_i}{2}\log[\Lambda_k]_{ii} - \frac{\nu_i}{2}[\Lambda_k]_{ii}\right) + c \qquad (50)\\
=& -\frac{1}{2}\operatorname{tr}(\Lambda_k R^{-1}((y_k - Cx_{k|k})(y_k - Cx_{k|k})^{\mathrm{T}} + CP_{k|k}C^{\mathrm{T}})]) \\
& - \frac{1}{2}\operatorname{tr}(\Lambda_k(\Delta R^{-1}\Delta + I)\mathop{\mathbb{E}}_{q_u}[u_k u_k^{\mathrm{T}}]) + \frac{1}{2}\operatorname{tr}(\Lambda_k R^{-1}\Delta\overline{u_k}(y_k - Cx_{k|k})^{\mathrm{T}}) \\
& + \frac{1}{2}\operatorname{tr}(\Lambda_k \Delta R^{-1}(y_k - Cx_{k|k})\overline{u_k}^{\mathrm{T}}) + \sum_{i=1}^{n_y}\left(\frac{\nu_i}{2}\log[\Lambda_k]_{ii} - \frac{\nu_i}{2}[\Lambda_k]_{ii}\right) + c \qquad (51)\\
=& \sum_{i=1}^{n_y}\left(\frac{\nu_i}{2}\log[\Lambda_k]_{ii} - \frac{\nu_i + [\Psi_k]_{ii}}{2}[\Lambda_k]_{ii}\right) + c, \qquad (52)
\end{aligned}
$$

where

$$
\begin{aligned}
\Psi_k =& R^{-1}((y_k - Cx_{k|k})(y_k - Cx_{k|k})^{\mathrm{T}} + CP_{k|k}C^{\mathrm{T}}) + (\Delta R^{-1}\Delta + I)\mathop{\mathbb{E}}_{q_u}[u_k u_k^{\mathrm{T}}] \\
& - R^{-1}\Delta\overline{u_k}(y_k - Cx_{k|k})^{\mathrm{T}} - \Delta R^{-1}(y_k - Cx_{k|k})\overline{u_k}^{\mathrm{T}}. \qquad (53)
\end{aligned}
$$

Therefore,

$$
q_\Lambda(\Lambda_k) = \prod_{i=1}^{n_y}\mathcal{G}\left([\Lambda_k]_{ii}; \frac{\nu_i}{2} + 1, \frac{\nu_i + [\Psi_k]_{ii}}{2}\right). \qquad (54)
$$

Note that only the diagonal elements of the matrix Ψ_k are required. As a consequence, provided that Δ and R are diagonal, only the diagonal elements of $\mathbb{E}[u_k u_k^{\mathrm{T}}]$ are required. These are second moments of univariate truncated normal distributions that can be computed using e.g., [5]. In the derivations above $\mathbb{E}_{q_\Lambda}[\Lambda_k]$ is required. The diagonal elements of $\mathbb{E}_{q_\Lambda}[\Lambda_k]$ are

$$
\mathop{\mathbb{E}}_{q_\Lambda}[[\Lambda_k]_{ii}] = \frac{\nu_i + 2}{\nu_i + [\Psi_k]_{ii}}. \qquad (55)
$$

References

[1] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "Robust inference for state-space models with skewed measurement noise," *Signal Processing Letters*, p. submitted, 2015. [Online]. Available: http://arxiv.org/abs/1503.06606

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

[3] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, Nov. 2008.

[4] 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.

[5] D. R. Barr and E. T. Sherrill, "Mean and variance of truncated normal distributions," *The American Statistician*, vol. 53, no. 4, pp. 357–361, 1999.

[6] T. M. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, 2006.

[7] 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.

Avdelning, Institution
Division, Department

Division of Automatic Control
Department of Electrical Engineering

Datum
Date

2015-03-19

Språk
Language

☐ Svenska/Swedish
⊠ Engelska/English
☐ ________________

Rapporttyp
Report category

☐ Licentiatavhandling
☐ Examensarbete
☐ C-uppsats
☐ D-uppsats
⊠ Övrig rapport
☐ ________________

ISBN
—

ISRN
—

Serietitel och serienummer
Title of series, numbering

ISSN
1400-3902

LiTH-ISY-R-3083

URL för elektronisk version

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

Titel
Title

Variational Iterations for Filtering and Smoothing with skew-t measurement noise

Författare
Author

Tohid Ardeshiri, Henri Nurminen, Robert Piché, Fredrik Gustafsson

Sammanfattning
Abstract

In this technical report, some derivations for the filter and smoother proposed in [1] are presented. More specifically, the derivations for the cyclic iteration needed to solve the variational Bayes filter and smoother for state space models with skew t likelihood proposed in [1] are presented.

Nyckelord
Keywords

skew t-distribution, skewness, t-distribution, robust filtering, Kalman filter, RTS smoother, variational Bayes