Robust Inference for State-Space Models with Skewed Measurement Noise

Robust Inference for State-Space Models with Skewed Measurement Noise

Henri Nurminen, Tohid Ardeshiri, Robert Piché, *Member, IEEE*, and Fredrik Gustafsson, *Fellow, IEEE*

Abstract—Filtering and smoothing algorithms for linear discrete-time state-space models with skewed and heavy-tailed measurement noise are presented. The algorithms use a variational Bayes approximation of the posterior distribution of models that have normal prior and skew-t-distributed measurement noise. The proposed filter and smoother are compared with conventional low-complexity alternatives in a simulated pseudorange positioning scenario. In the simulations the proposed methods achieve better accuracy than the alternative methods, the computational complexity of the filter being roughly 5 to 10 times that of the Kalman filter.

Index Terms—skew t, skewness, t-distribution, robust filtering, Kalman filter, RTS smoother, variational Bayes
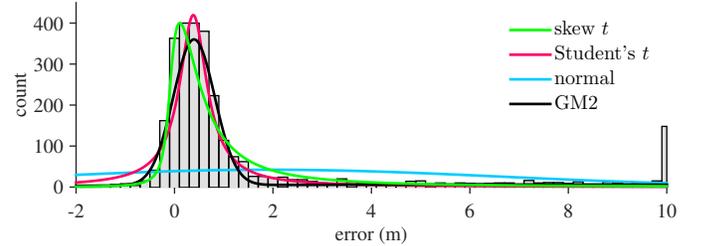
Figure 1. Skewed distributions fit better than symmetric distributions to the time-of-flight measurement errors. BIC values for 2905 data points are 9600 for skew t, 10500 for Student's t, 17200 for normal, and 10000 for GM2.

I. Introduction

The Kalman filter (KF) [1] is the linear minimum mean-square-error filter for linear state-space models, but it is optimal within the set of all filters only when the noise processes are normally distributed [2]. However, the normal distribution has small tail probabilities, and real-world data typically contain large errors ("outliers") more often than the normal distribution predicts [3]. Therefore, the KF is prone to large estimation errors when outliers occur. Hence, there is a need for filtering and smoothing algorithms that mitigate the outlier measurements' influence.

Many applications involve noise processes that have both heavy-tailed (high-kurtosis) and asymmetric (skewed) distributions. In radio signal based distance estimation [4], [5], for example, non-line-of-sight causes large positive errors [6], [7]. Fig. 1 shows the error histogram of a time-of-flight based ultra-wideband distance measurement experiment[1] and maximum likelihood fits of some probability distribution families. By the Bayesian information criterion (BIC) [8], the skewed distributions skew t [9, Ch. 4.3] and two-component Gaussian mixture (GM2) model the data better than the symmetric Student's t [10, Ch. 28] and normal. Other applications for asymmetric distributions have emerged in biostatistics [11], psychiatry [12], environmetrics [13], and economics [14].

Despite these applications, a computationally efficient estimation algorithm for time-series data with heavy-tailed and asymmetric noise has been missing. Robust algorithms that

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.

T. Ardeshiri and F. Gustafsson are with the Department of Electrical Engineering, Linköping University, 58183 Linköping, Sweden, (e-mails: tohid@isy.liu.se, fredrik@isy.liu.se). T. Ardeshiri receives funding from Swedish research council (VR), project ETT (621-2010-4301)

[1]High accuracy reference measurements are provided through the use of the Vicon real-time tracking system courtesy of the UAS Technologies Lab, Artificial Intelligence and Integrated Computer Systems Division (AIICS) at the Department of Computer and Information Science (IDA). http://www.ida.liu.se/divisions/aiics/aiicssite/index.en.shtml

model the heavy-tailed noise with a t-distribution are proposed in [15]–[17], but these do not use the skewness information. A GM2 can model skewness, but the number of mixture components in the posterior increases exponentially with the number of measurements. Furthermore, the GM2 has heavy tails only within a limited range near the component locations, and it has five parameters, while four suffices for modeling location, spread, skewness and kurtosis. Particle filters (PF) [18] can cope with a wide range of models including skewed noise processes, but their computational complexity increases rapidly as the state dimension increases.

This letter proposes approximations to the Bayesian filter and smoother that retain the computational efficiency of the KF while introducing more modeling flexibility for skewed and heavy-tailed measurement noise. The measurement noise is modelled by the skew t-distribution, and the proposed algorithms use a variational Bayes (VB) approximation of the posterior. The proposed filter and smoother are evaluated by numerical pseudorange positioning simulations, where they are compared with the state-of-the-art computationally light algorithms and a PF. To our knowledge, the only earlier work applying VB approximations to the skew t-distribution is that of Wand et al. [19]. However, Wand et al. do not consider state-space models and time-series estimation.

II. Skew t-distribution

Skewed extensions of the well-known unimodal symmetric distributions have been studied since the introduction of the skew normal distribution by Azzalini in [20]. The univariate skew t-distribution is parametrized by its location parameter $\mu \in \mathbb{R}$, spread parameter $\sigma > 0$, shape parameter $\delta \in \mathbb{R}$ and degrees of freedom $\nu > 0$, and has a probability density function (PDF) of the form

$$\mathrm{ST}(z; \mu, \sigma^2, \delta, \nu) = 2\,\mathrm{t}(z; \mu, \delta^2 + \sigma^2, \nu)\,\mathrm{T}(\widetilde{z}; 0, 1, \nu+1), \tag{1}$$

where

$$\mathrm{t}(z; \mu, \sigma^2, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{(z-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}} \tag{2}$$
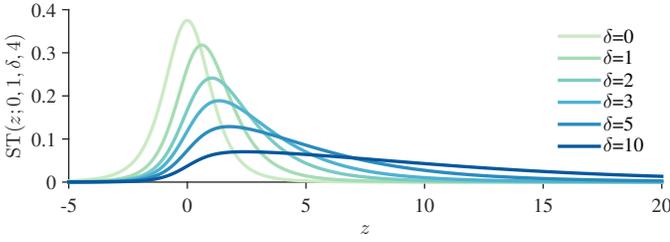
Figure 2. The PDF $\mathrm{ST}(z; 0, 1, \delta, 4)$ for different shape parameter values δ.

is the PDF of Student's t-distribution, $\Gamma(\cdot)$ is the gamma function, and $\widetilde{z} = \frac{(z-\mu)\delta}{\sigma} \left(\frac{\nu+1}{\nu(\delta^2+\sigma^2)+(z-\mu)^2} \right)^{\frac{1}{2}}$. Also, $\mathrm{T}(\cdot; 0, 1, \nu)$ denotes the cumulative distribution function (CDF) of Student's t-distribution with degrees of freedom ν. The PDF $\mathrm{ST}(z; 0, 1, \delta, 4)$ is plotted for six different values of shape parameter δ in Fig. 2. The skew t-distribution approaches normal distribution when $\nu \to \infty$ and $\delta \to 0$. Expressions for the first two moments of the univariate skew t-distribution with the parametrisation (1) can be found in [21].

Following the introduction of the multivariate skew normal distribution in [22], multivariate skew t-distributions have been proposed in [23]–[25]. In these versions, the PDF of the skew t-distribution involves only the univariate CDF of t-distribution, while the definition of skew t-distribution given in [26]–[28] involves the multivariate CDF, but a single kurtosis factor. In this letter the measurement noise distribution is a product of independent univariate skew t-distributions. This less general model is justified in applications where one-dimensional data from different sensors can be assumed to be statistically independent.

III. PROBLEM FORMULATION

Consider the linear state-space model with skew-t-distributed measurement noise

$$x_{k+1} = Ax_k + w_k, \qquad w_k \stackrel{\mathrm{iid}}{\sim} \mathcal{N}(w_k; 0, Q), \qquad (3a)$$

$$y_k = Cx_k + e_k, \qquad [e_k]_i \stackrel{\mathrm{iid}}{\sim} \mathrm{ST}([e_k]_i; 0, R_{ii}, \Delta_{ii}, \nu_i) \quad (3b)$$

where $\mathcal{N}(\cdot; \mu, \Sigma)$ denotes a (multivariate) normal PDF with mean μ and covariance Σ; $A \in \mathbb{R}^{n_x \times n_x}$ is the state transition matrix; $x_k \in \mathbb{R}^{n_x}$ indexed by $1 \leq k \leq K$ is the state to be estimated with prior distribution

$$p(x_1) = \mathcal{N}(x_1; x_{1|0}, P_{1|0}); \qquad (4)$$

where the subscript "$a|b$" is read "at time a using measurements up to time b"; $y_k \in \mathbb{R}^{n_y}$ also indexed by $1 \leq k \leq K$ are the measurements and the elements of y_k are conditionally independently skew-t-distributed; $R \in \mathbb{R}^{n_y \times n_y}$ is a diagonal matrix whose diagonal elements R_{ii} are the squares of the spread parameters of (3b); $\Delta \in \mathbb{R}^{n_y \times n_y}$ is a diagonal matrix whose diagonal elements Δ_{ii} are the shape parameters of (3b); $\nu \in \mathbb{R}^{n_y}$ is a vector whose elements ν_i are the degrees of freedom of (3b); $C \in \mathbb{R}^{n_y \times n_x}$ is the measurement matrix; $\{w_k \in \mathbb{R}^{n_x} | 1 \leq k \leq K\}$ and $\{e_k \in \mathbb{R}^{n_y} | 1 \leq k \leq K\}$ are mutually independent noise sequences; and the operator $[\cdot]_{ij}$ gives the (i, j) entry of its argument.

The aim of this letter is to derive a Bayesian filter and a Bayesian smoother using the VB method that computes an approximation of the filtering distribution $p(x_k | y_{1:k})$ and smoothing distribution $p(x_k | y_{1:K})$.

IV. VARIATIONAL SOLUTION

The likelihood function implied from (3b) has the hierarchical representation [27]

$$y_k | x_k, u_k, \Lambda_k \sim \mathcal{N}(Cx_k + \Delta u_k, \Lambda_k^{-1} R), \qquad (5a)$$

$$u_k | \Lambda_k \sim \mathcal{N}_+(0, \Lambda_k^{-1}), \qquad (5b)$$

$$[\Lambda_k]_{ii} \sim \mathcal{G}\left(\frac{\nu_i}{2}, \frac{\nu_i}{2} \right). \qquad (5c)$$

Λ_k is a diagonal matrix with independent random diagonal elements $[\Lambda_k]_{ii}$, and $\mathcal{N}_+(\mu, \Sigma)$ denotes the (multivariate) truncated normal distribution with closed positive orthant as support, location parameter μ, and squared-scale matrix Σ. Furthermore, $\mathcal{G}(\alpha, \beta)$ denotes the gamma distribution with shape parameter α and rate parameter β.

Using Bayes' theorem, the likelihood (5) and the prior (4), the joint smoothing posterior PDF can be written as

$$p(x_{1:K}, u_{1:K}, \Lambda_{1:K} | y_{1:K}) \propto p(x_1) \prod_{l=1}^{K-1} p(x_{l+1} | x_l)$$

$$\times \prod_{k=1}^{K} p(y_k | x_k, u_k, \Lambda_k) p(u_k | \Lambda_k) p(\Lambda_k) \qquad (6)$$

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

$$\times \prod_{k=1}^{K} \mathcal{N}(y_k; Cx_k + \Delta u_k, \Lambda_k^{-1} R) \mathcal{N}_+(u_k; 0, \Lambda_k^{-1})$$

$$\times \prod_{k=1}^{K} \prod_{i=1}^{n_y} \mathcal{G}\left([\Lambda_k]_{ii}; \frac{\nu_i}{2}, \frac{\nu_i}{2} \right). \qquad (7)$$

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

$$p(x_{1:K}, u_{1:K}, \Lambda_{1:K} | y_{1:K}) \approx q_x(x_{1:K}) q_u(u_{1:K}) q_\Lambda(\Lambda_{1:K}). \qquad (8)$$

In the VB approach, the Kullback-Leibler divergence (KLD) [29] of the true posterior from the factorized approximation is minimized;

$$\hat{q}_x, \hat{q}_u, \hat{q}_\Lambda = \operatorname*{argmin}_{q_x, q_u, q_\Lambda}$$

$$D_{\mathrm{KL}}(q_x(x_{1:K}) q_u(u_{1:K}) q_\Lambda(\Lambda_{1:K}) || p(x_{1:K}, u_{1:K}, \Lambda_{1:K} | y_{1:K}))$$

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

$$\log q_x(x_{1:K}) \leftarrow \mathop{\mathbb{E}}_{q_u q_\Lambda} [\log p(y_{1:K}, x_{1:K}, u_{1:K}, \Lambda_{1:K})] + c_x \quad (9a)$$

$$\log q_u(u_{1:K}) \leftarrow \mathop{\mathbb{E}}_{q_x q_\Lambda} [\log p(y_{1:K}, x_{1:K}, u_{1:K}, \Lambda_{1:K})] + c_u \quad (9b)$$

$$\log q_\Lambda(\Lambda_{1:K}) \leftarrow \mathop{\mathbb{E}}_{q_x q_u} [\log p(y_{1:K}, x_{1:K}, u_{1:K}, \Lambda_{1:K})] + c_\Lambda \quad (9c)$$

where the expected values on the right hand sides of (9) are taken with respect to the current q_x, q_u and q_Λ [30, Chapter 10] [31], [32]. Also, c_x, c_u and c_Λ are constants with respect to the variables x_k, u_k and Λ_k, respectively. This recursion is convergent to a local optimum [30, Chapter 10]. When the iterations converge, approximate densities q_u and q_Λ are integrated out from the right hand side of (8) by simply discarding them. Then, the approximate marginal smoothing density $q_x(x_k)$ is obtained, and it turns out to be a normal

Table I
SMOOTHING FOR SKEW-t MEASUREMENT NOISE

1: **Inputs:** A, C, Q, R, Δ, ν, $x_{1|0}$, $P_{1|0}$ and $y_{1:K}$
   initialization
2:   $\overline{\Lambda_k} \leftarrow I_{n_y}$ for $k = 1 \cdots K$
3:   $\overline{u_k} \leftarrow 0$ for $k = 1 \cdots K$
4: **repeat**
    update $q_x(x_{1:K})$ given $q_u(u_{1:K})$ and $q_\Lambda(\Lambda_{1:K})$
5:   **for** $k = 1$ to K **do**
6:     $K_x \leftarrow P_{k|k-1}C^{\mathrm{T}}(CP_{k|k-1}C^{\mathrm{T}} + \overline{\Lambda_k}^{-1}R)^{-1}$
7:     $x_{k|k} \leftarrow x_{k|k-1} + K_x(y_k - Cx_{k|k-1} - \Delta\overline{u_k})$
8:     $P_{k|k} \leftarrow (I - K_xC)P_{k|k-1}$
    predict $q_x(x_{k+1})$
9:     $x_{k+1|k} \leftarrow Ax_{k|k}$
10:     $P_{k+1|k} \leftarrow AP_{k|k}A^{\mathrm{T}} + Q$
11:   **end for**
12:   **for** $k = $ K-1 down to 1 **do**
13:     $G_k \leftarrow P_{k|k}A^{\mathrm{T}}P_{k+1|k}^{-1}$
14:     $x_{k|K} \leftarrow x_{k|k} + G_k(x_{k+1|K} - Ax_{k|k})$
15:     $P_{k|K} \leftarrow P_{k|k} + G_k(P_{k+1|K} - P_{k+1|k})G_k^{\mathrm{T}}$
16:   **end for**
   update $q_u(u_{1:K})$ and $q_\Lambda(\Lambda_{1:K})$ given $q_x(x_{1:K})$
17:   **for** $k = 1$ to K **do**
    update $q_u(u_k) = \mathcal{N}_+(u_k; u_{k|K}, U_{k|K})$
18:     $\widetilde{u}_k = y_k - Cx_{k|K}$
19:     $K_u \leftarrow \Delta(\Delta^2 + R)^{-1}$
20:     $u_{k|K} \leftarrow K_u\widetilde{u}_k$
21:     $U_{k|K} \leftarrow (I - K_u\Delta)\overline{\Lambda_k}^{-1}$
22:     $\overline{u_k} \leftarrow \mathbb{E}_{\mathcal{N}_+(u_{k|K}, U_{k|K})}[u_k]$ ▷ see [34] for the formula
23:     **for** $i = 1$ to n_y **do**
24:       $\Upsilon_{ii} \leftarrow \mathbb{E}_{\mathcal{N}_+(u_{k|K}, U_{k|K})}[[u_k]_i^2]$ ▷ see [34] for the formula
25:     **end for**
    update $q_\Lambda(\Lambda_k) = \prod_{i=1}^{n_y} \mathcal{G}\left([\Lambda_k]_{ii}; \frac{\nu_i}{2} + 1, \frac{\nu_i + [\Psi_k]_{ii}}{2}\right)$
26:     $\Psi_k \leftarrow R^{-1}(\widetilde{u}_k\widetilde{u}_k^{\mathrm{T}} + CP_{k|K}C^{\mathrm{T}}) + (\Delta R^{-1}\Delta + I)\Upsilon$
      $-R^{-1}\Delta\overline{u_k}\widetilde{u}_k^{\mathrm{T}} - \Delta R^{-1}\widetilde{u}_k\overline{u_k}^{\mathrm{T}}$
27:     $[\overline{\Lambda_k}]_{ii} \leftarrow \frac{\nu_i+2}{\nu_i+[\Psi_k]_{ii}}$
28:   **end for**
29: **until** converged
30: **Outputs:** $x_{k|K}$ and $P_{k|K}$ for $k = 1 \cdots K$

Table II
FILTERING FOR SKEW-t MEASUREMENT NOISE

1: **Inputs:** A, C, Q, R, Δ, ν, $x_{1|0}$, $P_{1|0}$ and $y_{1:K}$
2: **for** $k = 1$ to K **do**
   initialization
3:   $\overline{\Lambda_k} \leftarrow I_{n_y}$
4:   $\overline{u_k} \leftarrow 0$
5:   **repeat**
    update $q_x(x_k) = \mathcal{N}(x_k; x_{k|k}, P_{k|k})$ given $q_u(u_k)$ and $q_\Lambda(\Lambda_k)$
6:     $K_x \leftarrow P_{k|k-1}C^{\mathrm{T}}(CP_{k|k-1}C^{\mathrm{T}} + \overline{\Lambda_k}^{-1}R)^{-1}$
7:     $x_{k|k} \leftarrow x_{k|k-1} + K_x(y_k - Cx_{k|k-1} - \Delta\overline{u_k})$
8:     $P_{k|k} \leftarrow (I - K_xC)P_{k|k-1}$
    update $q_u(u_k) = \mathcal{N}_+(u_k; u_{k|k}, U_{k|k})$ given $q_x(x_k)$ and $q_\Lambda(\Lambda_k)$
9:     $K_u \leftarrow \Delta(\Delta^2 + R)^{-1}$
10:     $\widetilde{u}_k = y_k - Cx_{k|k}$
11:     $u_{k|k} \leftarrow K_u\widetilde{u}_k$
12:     $U_{k|k} \leftarrow (I - K_u\Delta)\overline{\Lambda_k}^{-1}$
13:     $\overline{u_k} \leftarrow \mathbb{E}_{\mathcal{N}_+(u_{k|k}, U_{k|k})}[u_k]$ ▷ see [34] for the formula
14:     **for** $i = 1$ to n_y **do**
15:       $\Upsilon_{ii} \leftarrow \mathbb{E}_{\mathcal{N}_+(u_{k|k}, U_{k|k})}[[u_k]_i^2]$ ▷ see [34] for the formula
16:     **end for**
    update $q_\Lambda(\Lambda_k) = \prod_{i=1}^{n_y} \mathcal{G}\left([\Lambda_k]_{ii}; \frac{\nu_i}{2} + 1, \frac{\nu_i + [\Psi_k]_{ii}}{2}\right)$
    given $q_u(u_k)$ and $q_x(x_k)$
17:     $\Psi_k \leftarrow R^{-1}(\widetilde{u}_k\widetilde{u}_k^{\mathrm{T}} + CP_{k|k}C^{\mathrm{T}}) + (\Delta R^{-1}\Delta + I)\Upsilon$
      $-R^{-1}\Delta\overline{u_k}\widetilde{u}_k^{\mathrm{T}} - \Delta R^{-1}\widetilde{u}_k\overline{u_k}^{\mathrm{T}}$
18:     $[\overline{\Lambda_k}]_{ii} \leftarrow \frac{\nu_i+2}{\nu_i+[\Psi_k]_{ii}}$
19:   **until** converged
   predict $q_x(x_{k+1})$
20:   $x_{k+1|k} \leftarrow Ax_{k|k}$
21:   $P_{k+1|k} \leftarrow AP_{k|k}A^{\mathrm{T}} + Q$
22: **end for**
23: **Outputs:** $x_{k|k}$ and $P_{k|k}$ for $k = 1 \cdots K$

$I_{3\times3}$, $\nu = 4 \cdot \mathbf{1}_{3\times1}$, and $\Delta = 5 \cdot I_{3\times3}$, where $\mathbf{1}$ is a vector of ones. The VB iterations of STVBF and TVBF are terminated when the change in the estimate is less than 0.01.

Some statistics of the estimation error are in Table III, and Fig. 3 shows an example of the error processes. Table III shows that the STVBF has the lowest root-mean-square error (RMSE), the TVBF and KF-G have negative bias, and the KF's error process has the highest standard deviation and positive skew. As illustrated by Fig. 3, the TVBF and KF-G react relatively slowly to positive errors, interpreting them as outliers to be discounted. The KF error's skewness is caused by excessive sensitivity to the large positive measurement errors.

distribution $q_x(x_k) = \mathcal{N}(x_k; x_{k|K}, P_{k|K})$ where the parameters $x_{k|K}$ and $P_{k|K}$ are the output of the smoothing algorithm given in Table I. The filtering algorithm and the parameters of the filtering posterior $q_x(x_k) = \mathcal{N}(x_k; x_{k|k}, P_{k|k})$ can be found in Table II. The derivations for the expectations given in (9) are relegated to [33] because of space constraints.

V. SIMULATIONS

Numerical simulations are carried out to evaluate the performance of the proposed algorithms Skew-t variational Bayes filter (STVBF) and Skew-t variational Bayes smoother (STVBS). The compared filters are t variational Bayes filter (TVBF) [16], the bootstrap Particle filter (PF), the Kalman filter (KF), and the KF with measurement validation gating (KF-G) [35, Ch. 5.7.2] that discards the individual measurement components whose normalized squared innovation is larger than the χ_1^2-distribution's 99 % quantile. The smoothers are t variational Bayes smoother (TVBS) [16], and Rauch-Tung-Striebel smoother with gating (RTSS-G) [36]. KF and RTSS use the true mean and covariance of the measurement noise distribution, and the TVBF and TVBS use the true mean and $(\nu - 2)/\nu$ times the true covariance as the shape matrix. The computations are done using MATLAB.

A. One-dimensional positioning

The simulation consists of 1000 100-step random-walks of model (3) with parameters $A = 1$, $Q = 1$, $C = \mathbf{1}_{3\times1}$, $R =$
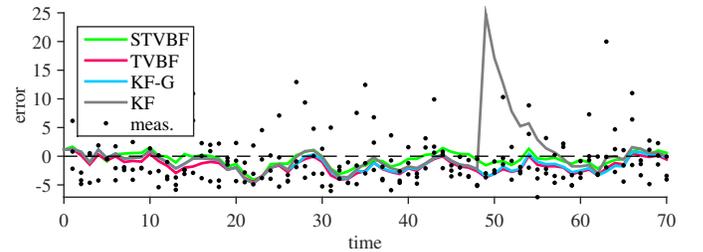

Figure 3. One-dimensional positioning example illustrates TVBF estimate's negative bias and KF's sensitivity to outliers. Measurement error of 300 at time instant 49 is not shown.

Table III
ERROR STATISTICS IN ONE-DIMENSIONAL POSITIONING

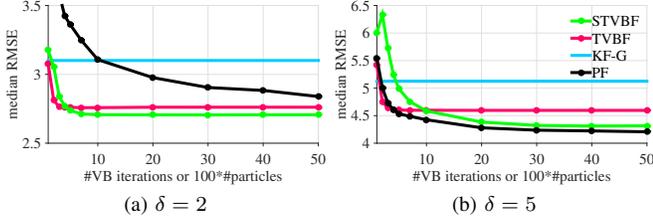| Filter	RMSE	Mean	Standard deviation	Skewness
STVBF	1.2	0.1	1.2	0.0
TVBF	1.5	-0.8	1.3	0.2
KF-G	1.5	-0.5	1.4	0.1
KF	1.6	0.0	1.6	0.5

(a) $\delta = 2$

(b) $\delta = 5$

Figure 4. Convergence of the STVBF with $q = 10$. Ten STVBF iterations is enough to outperform TVBF. One additional VB iteration gives the same accuracy gain as 100 additional PF particles.

B. Pseudorange positioning

GNSS-type (global navigation satellite system) pseudorange measurements are simulated from the model

$$[y_k]_i = \|s_i - [x_k]_{1:3}\| + [x_k]_4 + [e_k]_i, \ [e_k]_i \stackrel{\text{iid}}{\sim} \text{ST}(0, 1, \delta, 4) \tag{10}$$

where s_i is the ith satellite's position, $[x_k]_4$ is bias, e_k is noise, and δ is varied. The model is linearized, and the linearization error is negligible because the satellites are far from the receiver. The state model is a three-dimensional random walk with process noise covariance matrix $Q = \text{diag}(q^2, q^2, 0.5^2)$, where q is a parameter. The constant bias $[x_k]_4$ has prior $\mathcal{N}(0, 0.75^2)$. Satellite constellations of Global Positioning System provided by the International GNSS service [37] are used, and on average 7.6 satellites are measured. The results are based on 1000 Monte Carlo replications of a 100-step trajectory. The RMSE is computed for the components $[x_k]_{1:3}$.

1) Evaluation of the filter: Fig. 4 studies the convergence of the STVBF's VB iteration with $q = 10$. The speed of convergence depends on the parameters of the model; the larger δ, the slower convergence, and large q and a high number of sensors can also increase the required number of iterations. The RMSE reduction is fastest for the first iterations, 10 iterations is enough to outperform TVBF, and after 30 iterations the RMSE reduction is negligible. Thus, the STVBF is slower than the TVBF that requires 5 iterations. In this example, one additional VB iteration gives the same accuracy gain as 100 additional PF particles. In the remaining numerical examples, STVBF's VB iteration is terminated after 30 iterations, and TVBF's after 10 iterations.

Fig. 5 shows the distributions of the RMSE differences of the comparison methods from the STVBF's RMSE as percentages of the STVBF's RMSE. The levels of the boxes are 5 %, 25 %, 50 %, 75 %, and 95 % quantiles. With $q \geq 1$, the STVBF outperforms the comparison methods in significant majority of the replications. The problems with $q = 0.1$ are explained by the model structure: only sums of x_k and u_k are measured, so x_k and u_k are correlated *a posteriori*, which makes the VB approximation underestimate the posterior variance [30, Ch. 10.1.2]. The STVBF works well only when the process noise has enough dispersion to dominate in the prior's variance, i.e. when the signal-to-noise ratio (SNR) is not very low.

2) Real-world noise: The robustness of the STVBF is evaluated by generating the noise in Eq. (10) from the histogram distribution of the time-of-flight data set of Fig. 1 and using $q = 10$. The histogram of the RMSE differences of TVBF from the RMSE of STVBF is in Fig. 6. The proposed method has lower RMSE than the TVBF in 61 % of the 1000 Monte Carlo replications. This indicates that the proposed filter is robust to small deviations from the model that appear in real data.

(a) $q = 0.1$

(b) $q = 1$
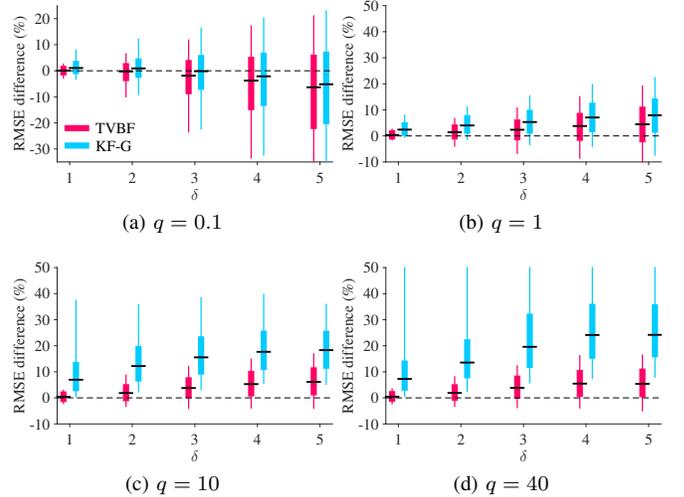
(c) $q = 10$

(d) $q = 40$

Figure 5. RMSE differences per cent of the STVBF's RMSE. The proposed STVBF outperforms the comparison methods with skewed measurements when the signal-to-noise ratio is high enough.
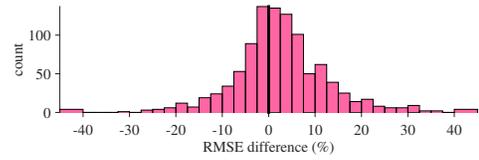

Figure 6. RMSE difference of TVBF per cent of the STVBF's RMSE with noise generated from real time-of-flight measurements' error histogram. STVBF has lower RMSE than the TVBF in 61 % of the 1000 replications.

3) Evaluation of the smoother: The smoother versions of the compared algorithms are evaluated in the same simulation of Eq. (10) with skew-t noise. The STVBS uses 30 and the TVBS 10 VB iterations, which were observed to provide convergence. Fig. 7 shows that the STVBS outperforms the TVBS also at low SNR, but the percentile differences at high SNR are smaller than those of the corresponding filters.
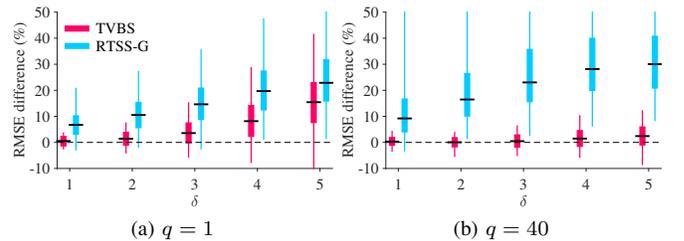

(a) $q = 1$

(b) $q = 40$

Figure 7. Smoothers' RMSE differences per cent of the STVBS's RMSE. STVBS performs well also at low SNR, but difference to TVBS is smaller than the difference between the corresponding filters.

VI. CONCLUSIONS

A filter and a smoother that take into account the skewness and heavy-tailedness of the measurement noise are proposed. The algorithms use the variational Bayes approximation. In the presented computer simulations the proposed methods outperform the conventional symmetric Kalman-type algorithms when skewness is present. The computational burden depends on the measurement dimension and model parameters. In the presented simulations the proposed filter has roughly 5 to 10 times the Kalman filter's computational cost.

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] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, ser. Prentice-Hall information and system sciences. Prentice-Hall, 1979.

[3] R. K. Pearson, "Outliers in process modeling and identification," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 1, pp. 55–63, January 2002.

[4] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 41–53, July 2005.

[5] B.-S. Chen, C.-Y. Yang, F.-K. Liao, and J.-F. Liao, "Mobile location estimator in a rough wireless environment using Extended Kalman-based IMM and data fusion," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 3, pp. 1157–1169, March 2009.

[6] K. Kaemarungsi and P. Krishnamurthy, "Analysis of WLAN's received signal strength indication for indoor location fingerprinting," *Pervasive and Mobile Computing*, vol. 8, no. 2, pp. 292–316, 2012, special Issue: Wide-Scale Vehicular Sensor Networks and Mobile Sensing.

[7] M. Kok, J. D. Hol, and T. B. Schön, "Indoor positioning using ultra-wideband and inertial measurements," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1293–1303, April 2015.

[8] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[9] A. Azzalini and A. Capitanio, *The Skew-Normal and Related Families*. Cambridge University Press, 2014.

[10] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Vol. 2*. Wiley, May 1995.

[11] S. Frühwirth-Schnatter and S. Pyne, "Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions," *Biostatistics*, vol. 11, no. 2, pp. 317–336, 2010.

[12] M. Eling, "Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models?" *Insurance: Mathematics and Economics*, vol. 51, no. 2, pp. 239–248, 2012.

[13] N. Counsell, M. Cortina-Borja, A. Lehtonen, and A. Stein, "Modelling psychiatric measures using skew-normal distributions," *European Psychiatry*, vol. 26, no. 2, pp. 112–114, 2010.

[14] Y. V. Marchenko, "Multivariate skew-t distributions in econometrics and environmetrics," Ph.D. dissertation, Texas A&M University, December 2010.

[15] G. Agamennoni, J. Nieto, and E. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5024–5037, October 2012.

[16] R. Piché, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution," in *2012 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, September 2012.

[17] M. Roth, E. Özkan, and F. Gustafsson, "A Student's t filter for heavy tailed process and measurement noise," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 5770–5774.

[18] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, July 2000.

[19] M. P. Wand, J. T. Ormerod, S. A. Padoan, and R. Frühwirth, "Mean field variational Bayes for elaborate distributions," *Bayesian Analysis*, vol. 6, no. 4, pp. 847–900, 2011.

[20] A. Azzalini, "A class of distributions which includes the normal ones," *Scandinavian Journal of Statistics*, vol. 12, no. 2, pp. 171–178, 1985.

[21] S. K. Sahu, D. K. Dey, and M. D. Branco, "Erratum: A new class of multivariate skew distributions with applications to Bayesian regression models," *Canadian Journal of Statistics*, vol. 37, no. 2, pp. 301–302, 2009.

[22] A. Azzalini and A. Dalla Valle, "The multivariate skew-normal distribution," *Biometrika*, vol. 83, no. 4, pp. 715–726, 1996.

[23] M. D. Branco and D. K. Dey, "A general class of multivariate skew-elliptical distributions," *Journal of Multivariate Analysis*, vol. 79, no. 1, pp. 99–113, October 2001.

[24] A. Azzalini and A. Capitanio, "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 65, no. 2, pp. 367–389, 2003.

[25] A. K. Gupta, "Multivariate skew t-distribution," *Statistics*, vol. 37, no. 4, pp. 359–363, 2003.

[26] S. K. Sahu, D. K. Dey, and M. D. Branco, "A new class of multivariate skew distributions with applications to Bayesian regression models," *Canadian Journal of Statistics*, vol. 31, no. 2, pp. 129–150, 2003.

[27] T.-I. Lin, "Robust mixture modeling using multivariate skew t distributions," *Statistics and Computing*, vol. 20, pp. 343–356, 2010.

[28] S. X. Lee and G. J. McLachlan, "EMMIXuskew: An R package for fitting mixtures of multivariate skew t distributions via the EM algorithm," *Journal of Statistical Software*, vol. 55, no. 12, pp. 1–22, November 2013.

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

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

[31] 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, November 2008.

[32] M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, Gatsby Computational Neuroscience Unit, University College London, 2003.

[33] T. Ardeshiri, H. Nurminen, R. Piché, and F. Gustafsson, "Variational iterations for filtering and smoothing with skew-t measurement noise," Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, Tech. Rep. LiTH-ISY-R-3076, March 2015. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-115741

[34] 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.

[35] Y. Bar-Shalom, R. X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation, Theory Algorithms and Software*. John Wiley & Sons, 2001.

[36] 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, August 1965.

[37] J. M. Dow, R. Neilan, and C. Rizos, "The international GNSS service in a changing landscape of global navigation satellite systems," *Journal of Geodesy*, vol. 83, no. 7, p. 689, February 2009.