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 variational Bayes approximation of the posterior distribution of normal prior and skew-t-distributed measurement noise. The proposed filter and smoother are compared with conventional low-complexity alternatives in simulated pseudorange positioning. In general the proposed algorithms show improvement in inference with skewed and heavy-tailed measurement noise. The computational burden depends on the measurement dimension and model parameters, and it can be tens of Kalman filters.**

***Index Terms*—skew t, skewness, t-distribution, robust filtering, Kalman filter, RTS smoother, variational Bayes**

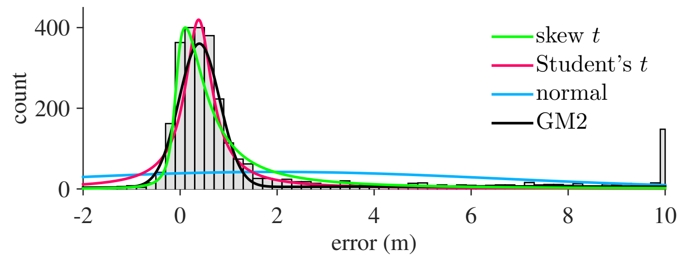


Figure 1. Skew distributions fit better than symmetric distributions to the time-of-flight measurement errors. The average log-likelihoods are -1.64 for skew t, -1.80 for Student's t, -2.96 for normal, and -1.71 for GM2.

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 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]. The skewed distributions skew t [8, Ch. 4.3] and two-component Gaussian mixture (GM2) attain better maximum likelihood fits than the symmetric Student's t [9, Ch. 28] and normal. Other applications for asymmetric distributions have emerged at least in biostatistics [10], psychiatry [11], environmetrics [12], and economics [13].

In spite of these applications, a computationally efficient estimation algorithm for time-series data with heavy-tailed and asymmetric noise has been missing. Robust filters and smoothers that model the heavy-tailed noise with a t-distribution are proposed in [14]–[16], but these do not use the skewness information. A GM2 can model skewness, but the number of components in the posterior GM increases exponentially with the number of measurements. Furthermore, the GM2 has heavy tails only within limited range from the component locations, and it has six parameters, while four is enough for modelling location, spread, skewness and kurtosis. Particle filters [17] 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 the variational Bayes (VB) approximation of the posterior. The proposed filter and smoother are evaluated by numerical simulations, and they are shown to outperform the state-of-the-art computationally light algorithms in a pseudorange positioning problem. To our knowledge, the only earlier work applying the VB approximation to the skew t-distribution is that of Wand et al. [18]. 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 [19]. 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

$$\mathrm{t}(z;\mu,\sigma^2,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{(z-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}} \tag{2}$$

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, $\mathrm{T}(\cdot;0,1,\nu)$

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

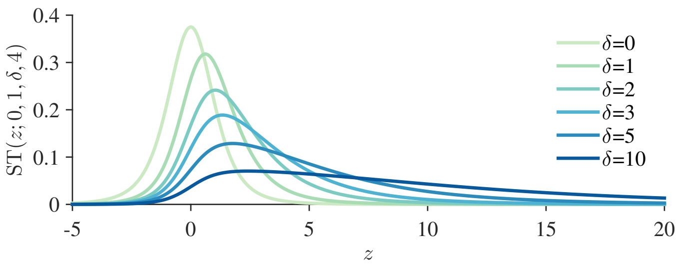


Figure 2. The PDF $\mathrm{ST}(z;0,1,\delta,4)$ for different shape parameter values δ.

denotes the Student's t-distribution's cumulative distribution function (CDF) with degrees of freedom ν and scale 1. The PDF of $\mathrm{ST}(z;0,1,\delta,4)$ 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 parametrization (1) can be found in [20].

Following the introduction of the multivariate skew normal distribution in [21], multivariate skew t-distribution has been proposed in [22]–[24]. 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 [25]–[27] 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

$$x_{k+1} = Ax_k + w_k, \quad w_k \overset{\text{iid}}{\sim} \mathcal{N}(w_k;0,Q), \tag{3a}$$

$$y_k = Cx_k + e_k, \qquad e_k \overset{\text{iid}}{\sim} \prod_{i=1}^{n_y} \mathrm{ST}([e_k]_i;0,R_{ii},\Delta_{ii},\nu_i) \tag{3b}$$

where $\mathcal{N}(\cdot;\mu,\Sigma)$ denotes a (multivariate) Gaussian 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 $1\le k\le K$ is the state to be estimated with prior distribution

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

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 $1\le k\le K$ 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 multiplicative factors 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 multiplicative factors of (3b); $\nu\in\mathbb{R}^{n_y}$ is a vector whose elements ν_i are the degrees of freedom of multiplicative factors of (3b); $C\in\mathbb{R}^{n_y\times n_x}$ is the measurement matrix; $\{w_k\in\mathbb{R}^{n_x}|1\le k\le K\}$ and $\{e_k\in\mathbb{R}^{n_y}|1\le k\le K\}$ are mutually independent noise sequences; 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 variational Bayes method that compute an approximation of the filtering distribution $p(x_k|y_{1:k})$ and smoothing distribution $p(x_k|y_{1:K})$, respectively.

IV. Variational solution

The likelihood function implied from (3b) has the hierarchical representation [26]

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

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

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

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

$$\begin{aligned} p(x_{1:K},u_{1:K},\Lambda_{1:K}|y_{1:K}) &\propto p(x_1)\prod_{l=1}^{K-1}p(x_{l+1}|x_l) \\ &\times\prod_{k=1}^{K}p(y_k|x_k,u_k,\Lambda_k)p(u_k|\Lambda_k)p(\Lambda_k) \end{aligned} \tag{6}$$

$$\begin{aligned} =&\mathcal{N}(x_1;x_{1|0},P_{1|0})\prod_{l=1}^{K-1}\mathcal{N}(x_{l+1};Ax_l,Q) \\ &\times\prod_{k=1}^{K}\mathcal{N}(y_k;Cx_k+\Delta u_k,\Lambda_k^{-1}R)\mathcal{N}_+(u_k;0,\Lambda_k^{-1}) \\ &\times\prod_{k=1}^{K}\prod_{i=1}^{n_y}\mathcal{G}\left([\Lambda_k]_{ii};\frac{\nu_i}{2},\frac{\nu_i}{2}\right). \end{aligned} \tag{7}$$

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

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

In the VB approach, the Kullback-Leibler divergence (KLD) [28] of the true posterior from the factorized approximation is minimized;

$$\begin{aligned} &\hat{q}_x,\hat{q}_u,\hat{q}_\Lambda = \underset{q_x,q_u,q_\Lambda}{\operatorname{argmin}} \\ &D_{\mathrm{KL}}(q_x(x_{1:K})q_u(u_{1:K})q_\Lambda(\Lambda_{1:K})||p(x_{1:K},u_{1:K},\Lambda_{1:K}|y_{1:K})) \end{aligned}$$

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

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

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

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

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 [29, Chapter 10] [30], [31]. 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 [29, 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

Table I
SMOOTHING FOR SKEW-t MEASUREMENT NOISE

Inputs: A, C, Q, R, Δ, ν, $x_{1|0}$, $P_{1|0}$ and $y_{1:K}$
initialization
$\overline{\Lambda_k} \leftarrow I_{n_y}$ for $k = 1 \cdots K$
$\overline{u_k} \leftarrow 0$ for $k = 1 \cdots K$
repeat
  update $q_x(x_{1:K})$ *given* $q_u(u_{1:K})$ *and* $q_\Lambda(\Lambda_{1:K})$
  for k = 1 to K **do**
    $K_x \leftarrow P_{k|k-1} C^{\mathrm{T}} (C P_{k|k-1} C^{\mathrm{T}} + \overline{\Lambda_k}^{-1} R)^{-1}$
    $x_{k|k} \leftarrow x_{k|k-1} + K_x (y_k - C x_{k|k-1} - \Delta \overline{u_k})$
    $P_{k|k} \leftarrow (I - K_x C) P_{k|k-1}$
   predict $q_x(x_{k+1})$
    $x_{k+1|k} \leftarrow A x_{k|k}$
    $P_{k+1|k} \leftarrow A P_{k|k} A^{\mathrm{T}} + Q$
  end for
  for k = K-1 down to 1 **do**
    $G_k \leftarrow P_{k|k} A^{\mathrm{T}} P_{k+1|k}^{-1}$
    $x_{k|K} \leftarrow x_{k|k} + G_k (x_{k+1|K} - A x_{k|k})$
    $P_{k|K} \leftarrow P_{k|k} + G_k (P_{k+1|K} - P_{k+1|k}) G_k^{\mathrm{T}}$
  end for
  update $q_u(u_{1:K})$ *and* $q_\Lambda(\Lambda_{1:K})$ *given* $q_x(x_{1:K})$
  for k = 1 to K **do**
   update $q_u(u_k) = \mathcal{N}_+(u_k; u_{k|K}, U_{k|K})$
    $\widetilde{u}_k = y_k - C x_{k|K}$
    $K_u \leftarrow \Delta (\Delta^2 + R)^{-1}$
    $u_{k|K} \leftarrow K_u \widetilde{u}_k$
    $U_{k|K} \leftarrow (I - K_u \Delta) \overline{\Lambda_k}^{-1}$
    $\overline{u_k} \leftarrow \mathbb{E}_{\mathcal{N}_+(u_{k|K}, U_{k|K})}[u_k]$ ▷ see [33] for the formula
    for i = 1 to n_y **do**
      $\Upsilon_{ii} \leftarrow \mathbb{E}_{\mathcal{N}_+(u_{k|K}, U_{k|K})}[[u_k]_i^2]$ ▷ see [33] for the formula
    end for
   update $q_\Lambda(\Lambda_k) = \prod_{i=1}^{n_y} \mathcal{G}\left([\Lambda_k]_{ii}; \frac{\nu_i}{2} + 1, \frac{\nu_i + [\Psi_k]_{ii}}{2}\right)$
    $\Psi_k \leftarrow R^{-1} (\widetilde{u}_k \widetilde{u}_k^{\mathrm{T}} + C P_{k|K} C^{\mathrm{T}}) + (\Delta R^{-1} \Delta + I) \Upsilon$
      $- R^{-1} \Delta \overline{u_k} \widetilde{u}_k^{\mathrm{T}} - \Delta R^{-1} \widetilde{u}_k \overline{u_k}^{\mathrm{T}}$
    $[\overline{\Lambda_k}]_{ii} \leftarrow \frac{\nu_i + 2}{\nu_i + [\Psi_k]_{ii}}$
  end for
until converged
Outputs: $x_{k|K}$ **and** $P_{k|K}$ **for** $k = 1 \cdots K$

Table II
FILTERING FOR SKEW-t MEASUREMENT NOISE

Inputs: A, C, Q, R, Δ, ν, $x_{1|0}$, $P_{1|0}$ and $y_{1:K}$
for k = 1 to K **do**
  initialization
  $\overline{\Lambda_k} \leftarrow I_{n_y}$
  $\overline{u_k} \leftarrow 0$
  repeat
   update $q_x(x_k) = \mathcal{N}(x_k; x_{k|k}, P_{k|k})$ *given* $q_u(u_k)$ *and* $q_\Lambda(\Lambda_k)$
    $K_x \leftarrow P_{k|k-1} C^{\mathrm{T}} (C P_{k|k-1} C^{\mathrm{T}} + \overline{\Lambda_k}^{-1} R)^{-1}$
    $x_{k|k} \leftarrow x_{k|k-1} + K_x (y_k - C x_{k|k-1} - \Delta \overline{u_k})$
    $P_{k|k} \leftarrow (I - K_x C) P_{k|k-1}$
   update $q_u(u_k) = \mathcal{N}_+(u_k; u_{k|k}, U_{k|k})$ *given* $q_x(x_k)$ *and* $q_\Lambda(\Lambda_k)$
    $K_u \leftarrow \Delta (\Delta^2 + R)^{-1}$
    $\widetilde{u}_k = y_k - C x_{k|k}$
    $u_{k|k} \leftarrow K_u \widetilde{u}_k$
    $U_{k|k} \leftarrow (I - K_u \Delta) \overline{\Lambda_k}^{-1}$
    $\overline{u_k} \leftarrow \mathbb{E}_{\mathcal{N}_+(u_{k|k}, U_{k|k})}[u_k]$ ▷ see [33] for the formula
    for i = 1 to n_y **do**
      $\Upsilon_{ii} \leftarrow \mathbb{E}_{\mathcal{N}_+(u_{k|k}, U_{k|k})}[[u_k]_i^2]$ ▷ see [33] for the formula
    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* $q_x(x_k)$
    $\Psi_k \leftarrow R^{-1} (\widetilde{u}_k \widetilde{u}_k^{\mathrm{T}} + C P_{k|k} C^{\mathrm{T}}) + (\Delta R^{-1} \Delta + I) \Upsilon$
      $- R^{-1} \Delta \overline{u_k} \widetilde{u}_k^{\mathrm{T}} - \Delta R^{-1} \widetilde{u}_k \overline{u_k}^{\mathrm{T}}$
    $[\overline{\Lambda_k}]_{ii} \leftarrow \frac{\nu_i + 2}{\nu_i + [\Psi_k]_{ii}}$
  until converged
  predict $q_x(x_{k+1})$
  $x_{k+1|k} \leftarrow A x_{k|k}$
  $P_{k+1|k} \leftarrow A P_{k|k} A^{\mathrm{T}} + Q$
end for
Outputs: $x_{k|k}$ **and** $P_{k|k}$ **for** $k = 1 \cdots K$

density $q_x(x_k)$ is obtained which turns out to be a normal distribution with parametrization $q_x(x_k) = \mathcal{N}(x_k; x_{k|K}, P_{k|K})$ where the parameters $x_{k|K}$ and $P_{k|K}$ are the output of the smoothing algorithm given in Table I. When the filtering density is required, the filtering algorithm and the parameters of the posterior $q_x(x_k) = \mathcal{N}(x_k; x_{k|k}, P_{k|k})$ can be found in Table II. The derivations for computing the expectations given in (9) are provided in [32] for space consideration.

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 the STVBF, t variational Bayes filter (TVBF) [15], and Kalman filter (KF). The smoother versions are the STVBS, t variational Bayes smoother (TVBS) [15], and Rauch-Tung-Striebel smoother (RTSS) [34]. The KF and RTSS are fed the true mean and covariance matrix of the measurement noise distribution, and the TVBF and TVBS are fed the true mean and $(\nu - 2)/\nu$ times the true covariance matrix as the shape matrix. The computations are done using MATLAB®.

A. One-dimensional positioning

The simulation consists of 1000 100-step trajectories of model (3) with parameters $A = 1$, $Q = 1$, $C = \mathbf{1}_{3\times1}$, $R = I_{3\times3}$, $\nu = 4 \cdot \mathbf{1}_{3\times1}$, 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 has negative bias, and the KF's error process has the highest standard deviation and positive skew. As illustrated by Fig. 3, the TVBF reacts relatively slowly to occasional large 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.

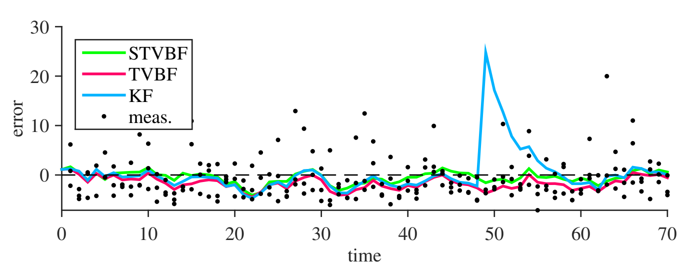
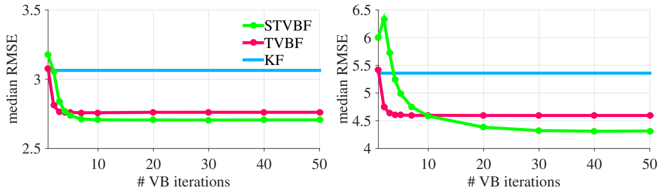


Figure 3. One-dimensional positioning example illustrates TVBF estimate's negative bias and KF's sensitivity to outliers. Measurement error of 300 at time instant 49 is not shown. A location parameter makes the measurement noise zero-mean.

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	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 beat the TVBF, but convergence can require tens of iterations.

B. Pseudorange positioning

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

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

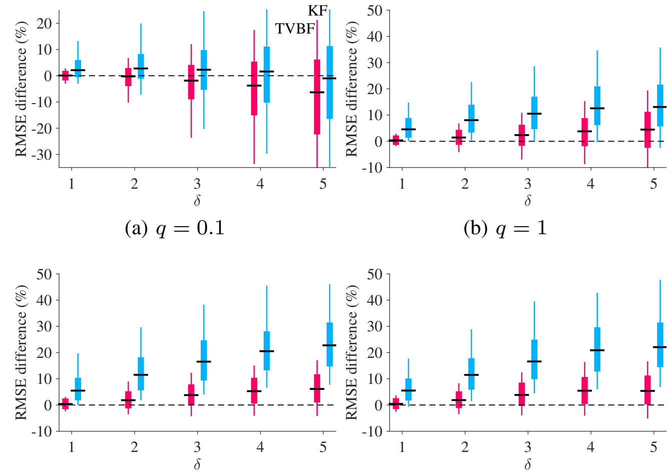
where s_i is the ith satellite's position, $[x_k]_4$ is bias, and e_k noise. The degree of freedom parameter is $\nu = 4$, and δ is varied. The model is linearised, and the linearisation error is negligible because the satellites are far from the receiver. The state model is a three-dimensional random walk with process noise covariance matrix $Q = \mathrm{diag}(q^2, q^2, 0.5^2)$, 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 [35] are used, and 7.6 satellites are measured on average. The results are based on 1000 Monte Carlo replications of a 100-step trajectory. The RMSEs are computed for the state 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, and after 30 iterations the reduction is usually negligible. Thus, the STVBF is slower than the TVBF that usually converges within five iterations. In the rest of the numerical examples, the STVBF's VB iteration is terminated after 30 iterations, and the 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 highly correlated *a posteriori*, which makes the VB approximation underestimate the posterior variance [29, 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 with $q = 10$. 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.

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) $q = 0.1$ (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.

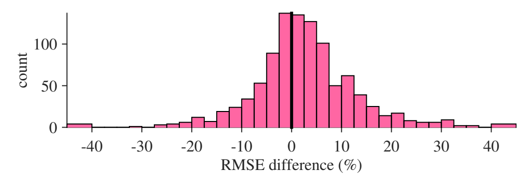
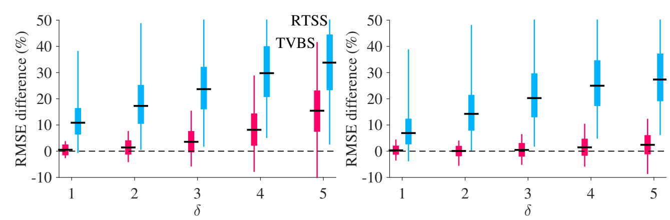


Figure 6. RMSE difference of TVBF per cent of the STVBF's RMSE with noise generated from real time-of-flight measurements' error histogram. STVBF has lower RMSE than the TVBF in 61 % of the 1000 replications.



(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, and it can be tens of Kalman filters for the filter. The algorithms are prone to underestimation of posterior variance, which can cause performance problems when signal-to-noise ratio is low.

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*, p. to appear, 2015, (special section on indoor localization, tracking, and mapping with heterogeneous technologies).

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

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

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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*, September 2012.

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.

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

[22] 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.

[23] 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.

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

[25] 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.

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

[27] 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, 11 2013.

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

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

[30] 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.

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

[32] 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

[33] 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.

[34] 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.

[35] 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.