2012 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 23–26, 2012, SANTANDER, SPAIN

RECURSIVE OUTLIER-ROBUST FILTERING AND SMOOTHING FOR NONLINEAR SYSTEMS USING THE MULTIVARIATE STUDENT-T DISTRIBUTION

Robert Piché

Tampere University of Technology
Department of Mathematics
Tampere, Finland

Simo Särkkä and Jouni Hartikainen

Aalto University
Department of Biomedical Engineering
and Computational Science
Espoo, Finland

ABSTRACT

Nonlinear Kalman filter and Rauch–Tung–Striebel smoother type recursive estimators for nonlinear discrete-time state space models with multivariate Student's t-distributed measurement noise are presented. The methods approximate the posterior state at each time step using the variational Bayes method. The nonlinearities in the dynamic and measurement models are handled using the nonlinear Gaussian filtering and smoothing approach, which encompasses many known nonlinear Kalman-type filters. The method is compared to alternative methods in a computer simulation.

***Index Terms*—** Robust filtering, Robust smoothing, Variational Bayes, Gaussian filter, Gaussian smoother

1. INTRODUCTION

Because large deviations typically occur in real data more frequently than is modeled by the normal (i.e. Gaussian) distribution, there is interest in developing estimation algorithms that can cope with these "outliers" while retaining the computational efficiency of Gaussian models.

In statistics, a popular approach to robustify linear multivariate regression is to use the Student's t-distribution in place of the Gaussian distribution and to compute the estimates using an Expectation-Maximization iteration [1–4]. This "outlier-accommodation" approach has many advantages: it is easy to implement, computationally light, amenable to Bayesian analysis, and doesn't require parameter tuning.

The approach has recently been extended to the sequential linear filtering and smoothing setting by Agamennoni et al. [5]. Their filter approximates the posterior state at each time step using an iterative solution of a variational Bayes formula; each iteration resembles a conventional Kalman filter update. Similar filter equations had been presented earlier in [6, 7], but the former work focused on on-line learning of the linear model's measurement noise variance parameter, while the latter work focused on outlier-detection rather than outlier-accommodation.

In this work we extend the outlier-robust filter and smoother to nonlinear state space models. For the sake of generality, we use the Gaussian filtering and smoothing framework [8–11] (also referred to as Gaussian assumed density filtering and smoothing). For Gaussian noise, this framework includes as special cases many popular nonlinear Kalman-type filters, including the extended Kalman filter (EKF) [12], the unscented Kalman filter (UKF) [13], the cubature Kalman filter (CKF) [14], the Gauss-Hermite Kalman filter (GHKF) [9], and the corresponding smoothers [11, 15, 16].

2. VARIATIONAL BAYES APPROXIMATION FOR FILTERING AND SMOOTHING

2.1. Problem Formulation

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

$$x_k \mid x_{k-1} \sim \mathrm{N}(f(x_{k-1}), Q_k) \tag{1a}$$

$$y_k \mid x_k \sim \mathrm{Student}(h(x_k), R_k, \nu), \tag{1b}$$

where $f(\cdot)$ and $h(\cdot)$ are non-linear dynamic and measurement model functions[1]. The matrix R_k is symmetric positive definite, and the Student's t-distribution's probability density is

$$p(y_k \mid x_k) \propto \left(1 + \tfrac{1}{\nu}(y_k - h(x_k))^T R_k^{-1}(y_k - h(x_k))\right)^{-\frac{\nu+d}{2}},$$

where d is the dimension of y_k and $\nu \geq 1$ is a parameter that controls the Student density's kurtosis (heavy-tailedness); for general practice a value of $\nu = 4$ is recommended [1]. The case $\nu = 1$ is the Cauchy distribution and for $\nu \to \infty$ the distribution converges to $\mathrm{N}(h(x_k), R_k)$.

In *filtering* the aim is to estimate the state distribution $p(x_k \mid y_{1:k})$ at time steps $k = 1, \ldots, T$. In contrast, in *smoothing* the aim is to estimate the state posterior given

[1]For clarity, we assume that $f(\cdot)$ and $h(\cdot)$ do not depend on k, but that dependency can be included easily to the presented algorithms.

978-1-4673-1026-0/12/$31.00

future measurements, that is, $p(x_k \mid y_{1:T})$ with $k = 1, \ldots, T$. With the model (1) these distributions are analytically intractable, and so in this article we aim to provide approximate algorithms that provide reasonably accurate solutions efficiently.

2.2. Nonlinear Kalman Filtering and Smoothing

Even in the case of Gaussian measurement noise (i.e. $\nu \to \infty$ in (1b)), the nonlinear filtering and smoothing problems are analytically intractable [8, 12]. An efficient approach for approximately solving the filtering problem with Gaussian measurement noise is the Gaussian filtering (GF) [8–10] framework. In Gaussian filter, the state posterior density $p(x_k \mid y_{1:k})$ is approximated with a multivariate Gaussian density $\mathrm{N}(x_k \mid m_k, P_k)$ at each time step $k = 1, \ldots, T$, with the mean m_k and covariance P_k computed via the following steps:

- *Prediction:*

$$
\begin{aligned}
m_k^- &= \int f(x_{k-1}) \; \mathrm{N}(x_{k-1} \mid m_{k-1}, P_{k-1}) \, \mathrm{d}x_{k-1}, \\
P_k^- &= \int (f(x_{k-1}) - m_{k-1}) \, (f(x_{k-1}) - m_{k-1})^T \\
&\qquad \times \mathrm{N}(x_{k-1} \mid m_{k-1}, P_{k-1}) \, \mathrm{d}x_{k-1} + Q_k.
\end{aligned}
$$

- *Update:*

$$
\begin{aligned}
\mu_k &= \int h(x_k) \; \mathrm{N}(x_k \mid m_k^-, P_k^-) \, \mathrm{d}x_k \\
S_k &= \int (h(x_k) - \mu_k)(h(x_k) - \mu_k)^T \\
&\qquad \times \mathrm{N}(x_k \mid m_k^-, P_k^-) \, \mathrm{d}x_k + R_k \\
C_k &= \int (x_k - m_k^-) \, (h(x_k) - \mu_k)^T \\
&\qquad \times \mathrm{N}(x_k \mid m_k^-, P_k^-) \, \mathrm{d}x_k \\
K_k &= C_k \, S_k^{-1} \\
m_k &= m_k^- + K_k \, (y_k - \mu_k) \\
P_k &= P_k^- - K_k \, S_k \, K_k^T.
\end{aligned}
$$

In [11] it was shown that approximating Gaussian smoothing densities $\mathrm{N}(x_k \mid m_k^s, P_k^s)$ for the model class (1), but with Gaussian measurement noise can obtained by solving the following Rauch–Tung–Striebel (RTS) smoother type equations in a backward order $k = T-1, \ldots, 1$:

$$
\begin{aligned}
L_k &= \int (x_k - m_k)(f(x_k) - m_{k+1}^-)^T \; \mathrm{N}(x_k \mid m_k, P_k) \, \mathrm{d}x_k \\
G_k &= L_k [P_{k+1}^-]^{-1} \\
m_k^s &= m_k + G_k (m_{k+1}^s - m_{k+1}^-) \\
P_k^s &= P_k + G_k (P_{k+1}^s - P_{k+1}^-) G_k^T.
\end{aligned}
$$

These filtering and smoothing methods, however, cannot be directly used in the case of Student's t-distributed measurement noise, but in the following we show how they can be used as part of a Variational Bayes (VB) framework to provides the required approximations efficiently.

2.3. Non-Linear Variational Bayes Filter

Introducing an auxiliary random variable λ_k, the measurement model's Student's t-distribution can be expressed as the Gaussian mixture

$$
p(y_k \mid x_k) = \int p(y_k \mid x_k, \lambda_k) p(\lambda_k) \, \mathrm{d}\lambda_k,
$$

where $y_k \mid x_k, \lambda_k \sim \mathrm{N}(h(x_k), \frac{1}{\lambda_k} R_k)$ and $\lambda_k \sim \mathrm{Gamma}(\frac{\nu}{2}, \frac{\nu}{2})$.

While assuming that $p(x_k \mid y_{1:k-1})$ is approximated with a Gaussian density $\mathrm{N}(x_k \mid m_k^-, P_k^-)$ using the prediction step of the non-linear Gaussian filter above, the state posterior can be found by marginalising λ_k out of the joint posterior

$$
\begin{aligned}
p(x_k, \lambda_k \mid y_{1:k}) &\propto \lambda_k^{\frac{d}{2}} \mathrm{e}^{-\frac{\lambda_k}{2}(y_k - h(x_k))^T R_k^{-1} (y_k - h(x_k))} \\
&\quad \cdot \mathrm{e}^{-\frac{1}{2}(x_k - m_k^-)^T (P_k^-)^{-1} (x_k - m_k^-)} \lambda_k^{\frac{\nu}{2}-1} \mathrm{e}^{-\frac{\nu \lambda_k}{2}}.
\end{aligned}
$$

In order to make this computation tractable we use a standard variational Bayes (VB) approach. We seek a separable approximation in the form of a product of probability densities, $p(x_k, \lambda_k \mid y_{1:k}) \approx q(x_k) q(\lambda_k)$. The VB approximation minimizes the Kullback-Leibler (KL) divergence between the product approximation and the true posterior,

$$
\begin{aligned}
&\mathrm{KL}(q(x_k) q(\lambda_k) \,\|\, p(x_k, \lambda_k \mid y_{1:k})) \\
&\quad = \int q(x_k) q(\lambda_k) \log \left(\frac{q(x_k) q(\lambda_k)}{p(x_k, \lambda_k \mid y_{1:k})} \right) \mathrm{d}x_k \, \mathrm{d}\lambda_k.
\end{aligned}
$$

Using the calculus of variations to minimize the KL divergence with respect to $q(\lambda_k)$ while keeping $q(x_k)$ fixed yields [17, p. 466]

$$
\begin{aligned}
\log q(\lambda_k) &= \mathrm{E}_x\big(\log p(x_k, \lambda_k, y_{1:k})\big) + \mathrm{const.} \\
&= -\tfrac{1}{2} \lambda_k \bar{\gamma}_k + \big(\tfrac{\nu + d}{2} - 1\big) \log \lambda_k - \tfrac{\nu \lambda_k}{2} + \mathrm{const.},
\end{aligned}
\tag{2}
$$

where

$$
\begin{aligned}
\bar{\gamma}_k &= \mathrm{E}_x\big((y_k - h(x_k))^T R_k^{-1} (y_k - h(x_k))\big) \\
&= \mathrm{tr}\{\mathrm{E}_x\big((y_k - h(x_k))(y_k - h(x_k))^T\big) R_k^{-1}\}.
\end{aligned}
\tag{3}
$$

This expectation can be approximated with the same methods as used in approximating the Gaussian integrals in non-linear Kalman filtering and smoothing. Thus, $q(\lambda_k)$ is the density of a $\mathrm{Gamma}(\frac{\nu + d}{2}, \frac{\bar{\gamma}_k + \nu}{2})$ distribution with mean

$$
\bar{\lambda}_k = \mathrm{E}_\lambda(\lambda_k) = \int \lambda_k \, q(\lambda_k) \, \mathrm{d}\lambda_k = \frac{\nu + d}{\nu + \bar{\gamma}_k}. \tag{4}
$$

Similarly, minimizing the KL divergence with respect to $q(x_k)$, with $q(\lambda_k)$ fixed, yields

$$
\begin{aligned}
\log q(x_k) &= \mathrm{E}_\lambda\big(\log p(x_k, \lambda_k, y_{1:k})\big) + \text{const.} \\
&= -\tfrac{1}{2}\bar{\lambda}_k (y_k - h(x_k))^T R_k^{-1} (y_k - h(x_k)) \\
&\quad - \tfrac{1}{2}(x_k - m_k^-)^T (P_k^-)^{-1} (x_k - m_k^-) + \text{const.}
\end{aligned}
$$

This can be seen as exactly the same estimation problem as in the update step of non-linear Kalman filter above, and thus we can approximate $q(x_k)$ with a Gaussian using the update step equations with measurement covariance $\frac{1}{\bar{\lambda}_k} R_k$. The variational parameters of $q(\lambda_k)$ and $q(x_k)$ are coupled, and can be solved with fixed-point iteration such that parameters of $q(\lambda_k)$ are solved while keeping $q(x_k)$ fixed and vice versa, which results in Algorithm 1.

Algorithm 1: One time step of the VB filter for the non-linear model with Student t-distributed measurement noise

---

Data: model parameters $f(), h(), Q_k, R_k, \nu$, prior state Gaussian parameters m_{k-1}, P_{k-1}, and number of VB fixed-point iterations N

Result: posterior state Gaussian parameters m_k, P_k

begin
- $m_k^- \leftarrow \int f(x_{k-1})\, \mathrm{N}(x_{k-1} \mid m_{k-1}, P_{k-1})\, \mathrm{d}x_{k-1}$
- $P_k^- \leftarrow \int (f(x_{k-1}) - m_k^-)\,(f(x_{k-1}) - m_k^-)^T \times \mathrm{N}(x_{k-1} \mid m_{k-1}, P_{k-1})\, dx_{k-1} + Q_k$
- $\bar{\lambda}_k \leftarrow 1$
- **for** n *from* 1 *to* N **do**
  - $\mu_k \leftarrow \int h(x_k)\, \mathrm{N}(x_k \mid m_k^-, P_k^-)\, \mathrm{d}x_k$
  - $S_k \leftarrow \int (h(x_k) - \mu_k)\,(h(x_k) - \mu_k)^T \times \mathrm{N}(x_k \mid m_k^-, P_k^-)\, \mathrm{d}x_k + \frac{1}{\bar{\lambda}_k} R_k$
  - $C_k \leftarrow \int (x_k - m_k^-)\,(h(x_k) - \mu_k)^T \times \mathrm{N}(x_k \mid m_k^-, P_k^-)\, \mathrm{d}x_k$
  - $K_k \leftarrow C_k\, S_k^{-1}$
  - $m_k \leftarrow m_k^- + K_k(y_k - \mu_k)$
  - $P_k \leftarrow P_k^- - K_k S_k K_k^T$
  - $D_k \leftarrow \int (y_k - h(x_k))\,(y_k - h(x_k))^T \times \mathrm{N}(x_k \mid m_k, P_k)\, \mathrm{d}x_k$
  - $\bar{\gamma}_k \leftarrow \mathrm{tr}(D_k R_k^{-1})$
  - $\bar{\lambda}_k \leftarrow \frac{\nu+d}{\nu+\bar{\gamma}_k}$
- **endfor**

end

2.4. Non-Linear Variational Bayes Smoother

The non-linear filtering algorithm can also be extended to provide approximate fixed-interval smoothing solutions, that is, factorizing solutions $q(x_k)q(\lambda_k)$ that approximate the true posterior $p(x_k, \lambda_k \mid y_{1:T})$ at each time step $k = 1, \ldots, T$. Using the standard VB approach, $\log q(\lambda_{1:T})$ can be written as

$$
\begin{aligned}
\log q(\lambda_{1:T}) &= \mathrm{E}_{x_{1:T}}(\log p(x_{1:T}, \lambda_{1:T}, y_{1:T})) + \text{const.} \\
&= \sum_{k=1}^{T} \mathrm{E}_{x_k}(\log p(y_k \mid x_k, \lambda_k)) + \log p(\lambda_k) + \text{const.}
\end{aligned}
$$

Because the terms of the sum are jointly independent, we have

$$
\begin{aligned}
\log q(\lambda_k) = &-\tfrac{1}{2}\lambda_k\, \mathrm{E}_{x_k}\big((y_k - h(x_k))^T R_k^{-1} (y_k - h(x_k))\big) \\
&+ \big(\tfrac{\nu+d}{2} - 1\big) \log \lambda_k - \tfrac{\nu \lambda_k}{2} + \text{const.},
\end{aligned}
$$

which is of the same form as (2) except that the expectation is taken over the smoothed state distribution $q(x_k)$.

Similarly, for $\log q(x_{1:T})$ we get

$$
\begin{aligned}
\log q(x_{1:T}) &= \mathrm{E}_{\lambda_{1:T}}(\log p(x_{1:T}, \lambda_{1:T}, y_{1:T})) + \text{const.} \\
&= \sum_{k=1}^{T} \log p(x_k \mid x_{k-1}) + \log p(y_k \mid x_k, \bar{\lambda}_k) + \text{const.}
\end{aligned}
$$

This can be identified as a non-linear smoothing problem with known measurement covariance, and thus be approximated with non-linear RTS smoothing algorithms [11].

As the result we have an algorithm that alternates between updating the approximations $q(x_k)$ and $q(\lambda_k)$ while keeping the other terms fixed such that $q(x_k)$ are updated using non-linear Kalman filtering and smoothing, and $q(\lambda_k)$ by solving the parameters from (5) similarly as in the non-linear VB filter above. Several iterations are needed for the parameters to converge towards a fixed value. The resulting fixed-interval smoother is shown Algorithm 2.

Because the smoothing algorithm also updates the estimates of the last time step, it would be possible to use fixed-lag smoothing to enhance the filtering results. A Gaussian fixed-lag smoother can be obtained as a simple modification to the fixed-interval smoothers as is shown in [11].

2.5. Conditionally Independent Measurements

In some cases it may be more appropriate to use a marginal t-distribution for each measurement channel. That is, in place of the multivariate t-distribution (1b), the measurement model is taken to be a product of t-marginals

$$
p(y_k \mid x_k) = \prod_{i=1}^{d} \mathrm{Student}(y_k \mid h_{k,i}(x_k), r_{i,k}, \nu_i).
$$

The VB algorithms developed above can be easily modified for this independent measurement-channel model, by using d-dimensional vectors for λ_k and $\bar{\gamma}_k$, and updating each of the components in the algorithms.

3. NUMERICAL EXAMPLE

To illustrate the use of the proposed algorithms, we consider a computer simulation of a multi-sensor bearings-only target

Algorithm 2: Variational Bayes smoother for interval $k = 1, \ldots, T$ for the non-linear state-space model with Student t-distributed measurement noise

Data: model parameters $f(), h(), Q_k, R_k, \nu$ and variational noise parameters $\bar{\lambda}_k$ for each time step $k = 1, \ldots, T$, state prior parameters m_0 and P_0 and number of VB fixed-point iterations N

Result: Smoothed state and noise parameters m_k^s, P_k^s and $\bar{\lambda}_k$ for each time step $k = 1, \ldots, T$

begin
  for *n from* 1 *to* N **do**
    for *k from* 1 *to* T **do**
      $m_k^- \leftarrow \int f(x_{k-1}) \, \mathrm{N}(x_{k-1} \,|\, m_{k-1}, P_{k-1}) \, \mathrm{d}x_{k-1}$
      $P_k^- \leftarrow \int (f(x_{k-1}) - m_k^-) \, (f(x_{k-1}) - m_k^-)^T \times \mathrm{N}(x_{k-1} \,|\, m_{k-1}, P_{k-1}) \, \mathrm{d}x_{k-1} + Q_k$
      $\mu_k \leftarrow \int h(x_k) \, \mathrm{N}(x_k \,|\, m_k^-, P_k^-) \, \mathrm{d}x_k$
      $S_k \leftarrow \int (h(x_k) - \mu_k) \, (h(x_k) - \mu_k)^T \times \mathrm{N}(x_k \,|\, m_k^-, P_k^-) \, \mathrm{d}x_k + \frac{1}{\bar{\lambda}_k} R_k$
      $C_k \leftarrow \int (x_k - m_k^-) \, (h(x_k) - \mu_k)^T \times \mathrm{N}(x_k \,|\, m_k^-, P_k^-) \, \mathrm{d}x_k$
      $K_k \leftarrow C_k S_k^{-1}$
      $m_k \leftarrow m_k^- + K_k (y_k - \mu_k)$
      $P_k \leftarrow P_k^- - K_k S_k K_k^T$
    endfor
    $m_T^s \leftarrow m_T$
    $P_T^s \leftarrow P_T$
    for *k from* $T-1$ *to* 1 **do**
      $L_k \leftarrow \int (x_k - m_k) \, (f(x_k) - m_{k+1}^-)^T \times \mathrm{N}(x_k \,|\, m_k, P_k) \, \mathrm{d}x_k$
      $G_k \leftarrow L_k [P_{k+1}^-]^{-1}$
      $m_k^s \leftarrow m_k + G_k (m_{k+1}^s - m_{k+1}^-)$
      $P_k^s \leftarrow P_k + G_k (P_{k+1}^s - P_{k+1}^-) G_k^T$
    endfor
    for *k from* 1 *to* T **do**
      $D_k \leftarrow \int (y_k - h(x_k)) \, (y_k - h(x_k))^T \times \mathrm{N}(x_k \,|\, m_k^s, P_k^s) \, \mathrm{d}x_k$
      $\bar{\gamma}_k \leftarrow \mathrm{tr}(D_k R_k^{-1})$
      $\bar{\lambda}_k \leftarrow \frac{\nu + d}{\nu + \bar{\gamma}_k}$
    endfor
  endfor
end

tracking problem, where the sensors measure the angle between the target and the sensor. Outliers in the sensor measurements are introduced using a clutter model that is not Student-t.

The dynamics of the target are modeled with coordinated turning model, where the state $x = (u, \dot{u}, v, \dot{v}, \omega)$ contains the 2d location (u, v) and the corresponding velocities $(\dot{u}, \dot{v})$ as well as the turning rate ω of the target. The dynamic model can be written as

$$x_k = \begin{pmatrix} 1 & \frac{\sin(\omega \Delta t)}{\omega} & 0 & \frac{\cos(\omega \Delta t) - 1}{\omega} & 0 \\ 0 & \cos(\omega \Delta t) & 0 & -\sin(\omega \Delta t) & 0 \\ 0 & \frac{1 - \cos(\omega \Delta t)}{\omega} & 1 & \frac{\sin(\omega \Delta t)}{\omega} & 0 \\ 0 & \sin(\omega \Delta t) & 0 & \cos(\omega \Delta t) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} x_{k-1} + q_k, \tag{5}$$

where $q_k \sim \mathrm{N}(0, Q_k)$ is the process noise. In our simulations we use the parameters

$$Q_k = \begin{pmatrix} q_1 M & 0 & 0 \\ 0 & q_1 M & 0 \\ 0 & 0 & q_2 \end{pmatrix}, \; M = \begin{pmatrix} \Delta t^3/3 & \Delta t^2/2 \\ \Delta t^2 & \Delta t \end{pmatrix},$$

where $q_1 = 0.1$, $q_2 = 1.75 \times 10^{-4}$ and $\Delta t = 1$. We generate measurements for sensors $i = 1, \ldots, d$ according to the mixture model

$$y_{i,k} \,|\, x_k, c_{i,k} \sim \begin{cases} \mathrm{U}(-\pi, \pi), & \text{if } c_{i,k} = 0 \\ \mathrm{N}\left(\arctan\left(\frac{v_k - s_v^i}{u_k - s_u^i}\right), \sigma_{i,k}^2\right), & \text{if } c_{i,k} = 1 \end{cases} \tag{6}$$

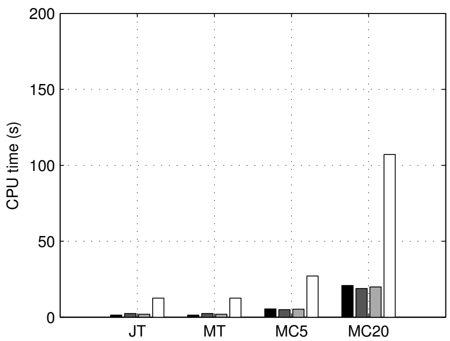
Here $c_{i,k} = 0$ indicates that the measurement from the ith sensor at the kth time step is clutter, while $c_{i,k} = 1$ if the measurement is from the actual target. Clutter is uniformly distributed over the angle measurement space $[-\pi, \pi]$, whereas the measurements from the actual target are the angle values from sensors located at (s_u^i, s_v^i) corrupted with additive Gaussian distributed noise with variances $\sigma_{i,k}^2$. In our simulation scenario we have $d = 4$ sensors placed at the corners of the square $[-10, 10] \times [-10, 10]$, and the variances are set to $\sigma_{i,k}^2 = 0.1^2$. We assume that the probability of clutter is constant over time, that is, $c_{i,k}$ is a Bernoulli random variable with $\mathrm{prob}(c_{i,k}) = p_c$. In our simulations we used the clutter probability $p_c = 0.1$. We simulated the target trajectory 100 times for 400 time steps according to dynamic model (5) and generated measurements on each time step with model (6). With the generated measurement sequences we assessed the performance of following filters and smoothers in estimating the state trajectories:

- JT: Non-linear Student's t-filter and smoother developed in this article with a joint t-distribution ($\nu = 5$) as the measurement model. In VB filter, 5 fixed point iterations were used, which is enough to achieve convergence in all cases. In smoothers we used 5 VB iterations.
- MT: Same as JT with the joint t-distribution replaced by separate t-marginals for each of the 4 sensors.
- MC5 and MC20: Rao-Blackwellized Monte Carlo data association (RBMCDA) algorithm [18] with 5 and 20 particles. This is an ideal method for this simulation

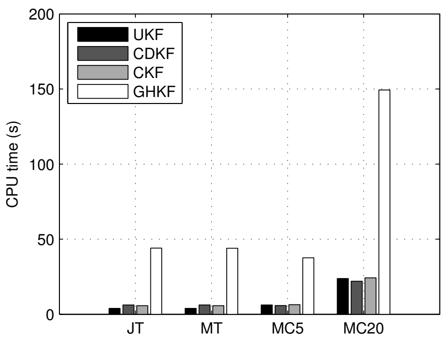
scenario since the clutter model specified above is a special case of model class assumed by RBMCDA.

- PA: Gaussian ADF with perfect associations, that is, clutter measurements are not given to the filter.

We used the numerical approximation methods of UKF, CDKF, CKF and GHKF in approximating the necessary Gaussian integrals in every tested filtering algorithm. In UKF we used the transformation parameters $(\alpha = 1, \beta = 2, \kappa = 0)$, in CDKF the step size $h_{\mathrm{cd}} = \sqrt{3}$ and in GHKF $m = 3^n = 243$ quadrature points.



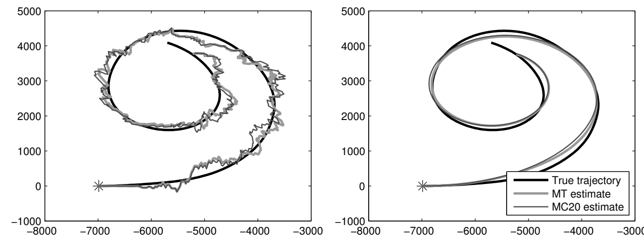
(a) CPU time (filters)



(b) CPU time (smoothers)

Fig. 1. The average CPU times taken by the methods are shown for filters (a) and for filter and smoothers combined (b).

Figure 1 shows the CPU times taken by the methods, showing that the Student's t based filters are many times faster than RBMCDA. We can also see that while GHKF gives the highest estimation accuracy it also much slower than the rest since the number of quadrature points m is so large (in general m grows exponentially with the number of state components). Among the other methods CKF provides a good trade-off in accuracy and speed. Both t-filters also provide adequate tracking accuracy, which is further illustrated in Figure 2, which show typical target trajectories and estimates of it given by MT and MC20 (with CKF/CRTS).



(a) Example filtered estimate (b) Example smoothed estimate

Fig. 2. Example trajectory with filtered and smoothed estimates.

The root mean square errors (RMSE) of estimating the target positions and turn rates are shown in panels (a)–(d) of Figure 3. We can see that the performance of RBMCDA is nearly optimal with both 5 and 20 particles. The regular Gaussian measurement model was also tested, but it diverged in all simulation cases. Of the tested integration methods GHKF gave the best performance in all cases, followed by CKF, CDKF and UKF in that order.

4. CONCLUDING REMARKS

In this article, we have presented a new outlier-robust filter and smoother for nonlinear state space models. The filter and smoother are based on using Student's t-distribution in the measurement model. The methods approximate the posterior state at each time step using the variational Bayes method and handle the non-linearities using the Gaussian filtering and smoothing approach. The method was compared to the particle filtering based method RBMCDA in computer simulation, which showed that the proposed approach provides a good trade-off between accuracy and computational efficiency.

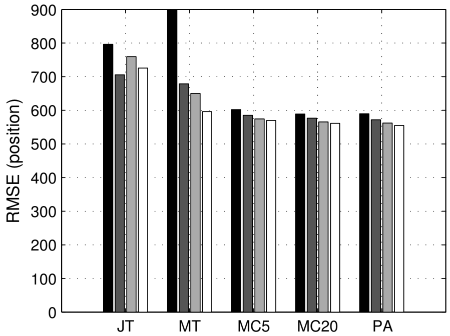
5. REFERENCES

[1] K. L. Lange, R. J. A. Little, and J. M. G. Taylor, "Robust statistical modeling using the t distribution," *Journal of the American Statistical Association*, vol. 84, no. 408, pp. 881–896, 1989.

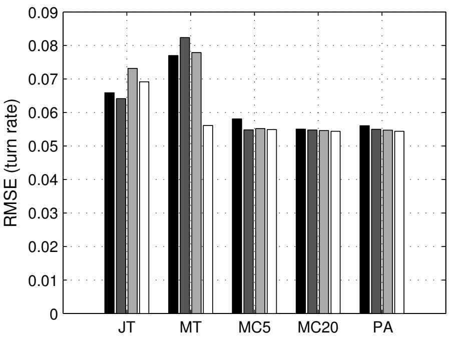
[2] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, Wiley Series in Probability and Statistics. Wiley, New York, 1987.

[3] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall/CRC, second edition, 2004.

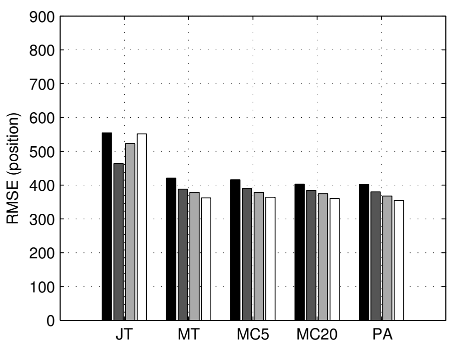
[4] R. Piché, "Robust multivariate linear regression using the Student-t distribution," Tech. Rep., Tampere University of Technology, Tampere, Finland, May 2012.



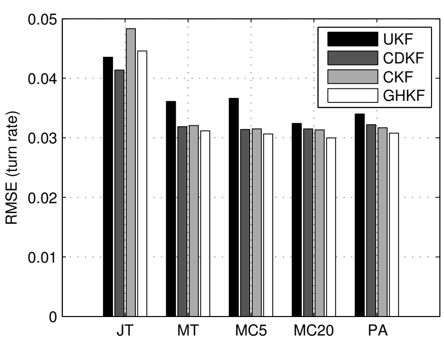
(a) Filter RMSE (position)



(b) Filter RMSE (turn rate)



(c) Smoother RMSE (position)



(d) Smoother RMSE (turn rate)

Fig. 3. RMSE values of estimating positions and turn rates with all the tested methods over 100 simulations.

[5] G. Agamennoni, J.I. Nieto, and E.M. Nebot, "An outlier-robust Kalman filter," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, May 2011, pp. 1551–1558.

[6] S. Särkkä and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 596–600, 2009.

[7] J.-A. Ting, E. Theodorou, and S. Schaal, "A Kalman filter for robust outlier detection," in *Proc. IEEE/RSJ Intl. Conf. Intelligent Robots and Systems*, Oct. 2007, pp. 1514–1519.

[8] P. Maybeck, *Stochastic Models, Estimation and Control*, vol. 2, Academic Press, 1982.

[9] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[10] Y. Wu, D. Hu, M. Wu, and X. Hu, "A numerical-integration perspective on Gaussian filters," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2910–2921, August 2006.

[11] S. Särkkä and J. Hartikainen, "On Gaussian optimal smoothing of non-linear state space models," *IEEE Transactions on Automatic Control*, vol. 55, no. 8, pp. 1938–1941, August 2010.

[12] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, 1970.

[13] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *Proceedings of the 1995 American Control, Conference, Seattle, Washington*, 1995, pp. 1628–1632.

[14] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, June 2009.

[15] S. Särkkä, "Unscented Rauch-Tung-Striebel smoother," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, pp. 845–849, April 2008.

[16] I. Arasaratnam and S. Haykin, "Cubature Kalman smoothers," *Automatica*, vol. 47, no. 10, pp. 2245–2250, 2011.

[17] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[18] S. Särkkä, A. Vehtari, and J. Lampinen, "Rao-Blackwellized particle filter for multiple target tracking," *Information Fusion*, vol. 8, no. 1, pp. 2 – 15, 2007.