Signal Processing 140 (2017) 53–59

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Short communication

Robust stochastic integration filtering for nonlinear systems under multivariate *t*-distributed uncertainties

Syed Safwan Khalid[a], Naveed Ur Rehman[a], Shafayat Abrar[b,*]

[a] COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
[b] School of Science and Engineering, Habib University, Karachi 75290, Pakistan

ARTICLE INFO

Article history:
Received 8 November 2016
Revised 1 May 2017
Accepted 5 May 2017
Available online 6 May 2017

Keywords:
Nonlinear filtering
Robust filtering
Stochastic integration
Heavy-tailed noise
Student *t*-distribution

ABSTRACT

Bayesian filtering solutions that are developed under the assumption of heavy-tailed uncertainties are more robust to outliers than the standard Gaussian ones. In this work, we consider robust nonlinear Bayesian filtering in the presence of multivariate *t*-distributed process and measurement noises. We develop a robust stochastic integration filter (RSIF) based on stochastic spherical-radial integration rule that achieves asymptotically exact evaluations of multivariate *t*-weighted integrals of nonlinear functions that arise in nonlinear Bayesian filtering framework. The superiority of the proposed scheme is demonstrated by comparing its performance against the cubature Kalman filter (CKF), a robust CKF, and the standard SIF in a representative example concerning bearings-only target tracking.

1. Introduction

Bayesian filtering provides a theoretical framework for recursive estimation of unknown dynamic state vectors in linear/nonlinear filtering applications. In Bayesian paradigm, the posterior probability of the state vector given the noisy observations is recursively updated, at each instant, using a process and a measurement model. However, in general, the evaluation of the posterior probability is analytically intractable, and hence only approximate solutions are available [1]. A widely used approximation utilizes the assumption that the required posterior distribution is Gaussian and the corresponding filters are termed as the Gaussian assumption (GA) filters [2]. In Table 1, we list a number of important GA filters available in the literature. The Gaussian assumption, however, is frequently violated in practice. For instance, the occurrence of outliers is a type of non-Gaussian phenomenon which is found in many applications of practical interest [3]. Consequently, filters based on the Gaussian assumption perform poorly in the presence of outliers.

A filter can be made robust to outliers by incorporating heavy-tailed uncertainties in the process and measurement models. A popular choice in literature is the use of multivariate generalization of Student *t*-distribution [2,4–6]; hereafter, referred to as *t*-distribution. In Table 2, we list a number of robust filters that rely on the assumption of *t*-distributed uncertainties. An early attempt to incorporate *t*-distribution in the Kalman filtering framework was discussed in [7]; however, the approach is purely one dimensional and cannot be extended to multivariate case. In [3,5,8], *t*-distributed uncertainties have been incorporated in the Bayesian filtering framework using the variational Bayes (VB) approach. In VB method, a solution is developed by converting the intractable posterior probability density function into a tractable factored form. The resulting factors are, however, coupled and the procedure requires a number of fixed-point iterations to obtain an admissible solution. A more direct approach has appeared in [4], where an approximate closed-form solution has been developed for systems under *t*-distributed process and measurement noise; however, the solution is valid for linear systems only.

In [2,6], the Student *t*-filter of Roth et al. [4] is extended to nonlinear systems using sigma-point methods that employed deterministic cubature rules to evaluate multivariate *t*-weighted nonlinear integrals. Such deterministic cubature rules, though computationally efficient, lead to approximate integral evaluations. Consequently, they perform inadequately in problems involving nonlinearities with large uncertainties [9,10]. On the other hand, numerical integration based on stochastic methods can achieve asymptotically exact evaluations [10–12]. The motivation behind our work is to employ the stochastic integration methods in the development of a robust nonlinear Bayesian filter assuming *t*-distributed process and measurement noises. We first describe stochastic in-

* Corresponding author.
E-mail addresses: safwan_khalid@comsats.edu.pk (S.S. Khalid), naveed.rehman@comsats.edu.pk (N.U. Rehman), shafayat.abrar@sse.habib.edu.pk (S. Abrar).

http://dx.doi.org/10.1016/j.sigpro.2017.05.007

Table 1
Traditional filters based on Gaussian distribution.

Filters based on deterministic rules	Remarks
Cubature Kalman Filter (CKF) [15]	Based on third-degree cubature rule.
Smooth Variable Structure Filters [16,17]	Variants of Kalman filters robust to modelling uncertainties.
Unscented Kalman Filter (UKF) [18]	Based on unscented transformation.
Gauss-Hermite Quadrature (GHQ) Filter [13]	More accurate than CKF at the cost of computations.
Sparse-Grid Quadrature Filter [19]	A Low-cost variant of GHQ.
Embedded CKF [20]	Based on embedded cubature rule.
Based on stochastic rules	
Monte-Carlo Kalman Filter (MKF) [21]	Based on Mote-Carlo integration.
Stochastic Integration Filter (SIF) [10]	Asymptotically exact integration; faster convergence than Monte-Carlo rule.

Table 2
Robust filters based on t-distribution.

Filters based on deterministic rules	Remarks
Poly-t robust Kalman filter [7]	For single dimension, not extendable to multivariate case.
VB based robust filter [3,5]	Posterior is approximated using variational Bayes (VB).
VB based smoother [8]	VB based fixed-interval smoother.
t-distributed Kalman filter [4]	Approximate recursive closed-form solution for linear systems.
Robust CKF [2,6]	Extension of [4] to nonlinear systems using deterministic cubature rules.
Based on stochastic rules	
Our proposed work in this paper.	Degree-three stochastic integration rule.

tegration rules of an arbitrary degree to evaluate multivariate t-weighted nonlinear integrals, and then proceed to develop a third-degree robust stochastic integration filter (RSIF).

This paper is organized as follows: Section 2 describes briefly the existing statistical linear regression based Bayesian filtering under t-distributed uncertainties [2,5,13,14]. Section 3 describes generic stochastic evaluation rules for t-weighted spherical-radial integrals, and obtains a third-degree t-weighted stochastic integration rule. Section 4 obtains a robust Bayesian filter based on the third-degree integration rule as developed in Section 3. Section 5 presents simulation results, and Section 6 draws conclusions.

1.1. Notations

Scalars are represented by small letters, and matrices are represented by capital letters. Bold-faced small letters are used for column vectors that represent states and measurements; whereas, bold-faced capital letters are used to represent a set of column vectors. We use superscript T to represent transpose of a matrix. Notations $\boldsymbol{x} \sim \mathrm{St}(\boldsymbol{\mu}, \Sigma, \eta)$ denote that $\boldsymbol{x}$ is distributed according to the generalized multivariate Student t-distribution, i.e.,

$$p(\boldsymbol{x}) = \frac{\Gamma((\eta+d)/2)}{\Gamma(\eta/2)} \frac{1}{(\eta\pi)^{d/2}\sqrt{\Sigma}} \left(1 + \frac{\delta^2(\boldsymbol{x})}{\eta}\right)^{-(\eta+d)/2},$$

where $\delta^2(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$, $d = \dim(\boldsymbol{x})$, $\boldsymbol{\mu}$ represents the mean, η is the degree-of-freedom parameter, and Σ is the covariance matrix parameter of the t-distribution.

Remark 1. Note that η is also referred to as the shape parameter of $p(\boldsymbol{x})$ [22] and determines its tail-behavior, i.e., increasingly heavier tails are obtained as η decreases towards one. Conversely, as η tends to infinity, $p(\boldsymbol{x})$ approaches the standard Gaussian distribution. Also note that for $\eta < 2$, the resulting distribution has infinite variance. Since our work relies on propagation of first and second order moments, therefore, we consider the cases for which $\eta > 2$. The covariance matrix of $\boldsymbol{x} \sim \mathrm{St}(\boldsymbol{\mu}, \Sigma, \eta)$ is equal to $\frac{\eta}{\eta-2}\Sigma$ for $\eta > 2$. Note that multivariate t-distributions appear in various different forms in literature [22,23]. The definition, we have considered here, constrains the multivariate t-distribution to be an elliptic distribution, and this is the most common definition available in literature.

2. Bayesian filtering under t-distributed uncertainties

Consider a representative nonlinear system:

$$\boldsymbol{x}_k = f(\boldsymbol{x}_{k-1}) + \boldsymbol{w}_{k-1}, \tag{1a}$$

$$\boldsymbol{y}_k = h(\boldsymbol{x}_k) + \boldsymbol{v}_k, \tag{1b}$$

where $\boldsymbol{x}_k \in \mathbb{R}^n$ and $\boldsymbol{y}_k \in \mathbb{R}^m$ are the state and the observation vectors, respectively. The system model $f(\cdot)$ and the observation model $h(\cdot)$ are nonlinear functions. The noise processes $\boldsymbol{w}_{k-1}$ and $\boldsymbol{v}_k$ represent the uncertainties. To incorporate the effect of possible outliers $\boldsymbol{w}_{k-1}$ and $\boldsymbol{v}_k$ are modeled as independent t-distributed processes, i.e., $\boldsymbol{w}_{k-1} \sim \mathrm{St}(\boldsymbol{0}, Q_{k-1}, \eta)$ and $\boldsymbol{v}_k \sim \mathrm{St}(\boldsymbol{0}, R_k, \eta)$. Also, the noise sequences, $\{\boldsymbol{w}_k\}$ and $\{\boldsymbol{v}_k\}$, are assumed to be independent and identically distributed. Let $\boldsymbol{Y}_k = \{\boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_k\}$ be the set of all available observations up until the kth instant. The aim of the filtering process is to provide an estimate of the state vector given $\boldsymbol{Y}_k$. We know that the optimal Bayesian estimate, in terms of minimum mean square error, is given by $\widehat{\boldsymbol{x}}_{k|k} = \mathrm{E}[\boldsymbol{x}_k|\boldsymbol{Y}_k]$, i.e., $\widehat{\boldsymbol{x}}_{k|k} = \int \boldsymbol{x}_k p(\boldsymbol{x}_k|\boldsymbol{Y}_k)\mathrm{d}\boldsymbol{x}_k$. The required probability $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$ is evaluated using a *time-update* and a *measurement-update*. Before we develop the expressions for these updates, we list two important properties of t-distribution that are utilized in the subsequent derivations; refer to [24] for proof.

P1: Let $\boldsymbol{x} \sim \mathrm{St}(\widehat{\boldsymbol{x}}, \Sigma, \eta)$ and $\boldsymbol{z}$ be an affine transformation, i.e., $\boldsymbol{z} = A\boldsymbol{x} + \boldsymbol{b}$, where A and $\boldsymbol{b}$ are deterministic matrix and vector of appropriate dimensions. Then, $\boldsymbol{z}$ is t-distributed with density $\mathrm{St}(A\widehat{\boldsymbol{x}} + \boldsymbol{b}, A\Sigma A^T, \eta)$.

P2: Let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$ be jointly t-distributed, i.e.,

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2) = \mathrm{St}\left(\begin{bmatrix}\widehat{\boldsymbol{x}}_1 \\ \widehat{\boldsymbol{x}}_2\end{bmatrix}, \begin{bmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22}\end{bmatrix}, \eta\right)$$

Then, the conditional distribution $p(\boldsymbol{x}_1|\boldsymbol{x}_2)$ is given as $p(\boldsymbol{x}_1|\boldsymbol{x}_2) = \mathrm{St}(\widehat{\boldsymbol{x}}_{1|2}, \Sigma_{1|2}, \widetilde{\eta})$, where $\widetilde{\eta} = \eta + n$, $\widehat{\boldsymbol{x}}_{1|2} = \widehat{\boldsymbol{x}}_1 + \Sigma_{1|2}\Sigma_{22}^{-1}(\boldsymbol{x}_2 - \widehat{\boldsymbol{x}}_2)$ and

$$\Sigma_{1|2} = \frac{\eta + (\boldsymbol{x}_2 - \widehat{\boldsymbol{x}}_2)^T \Sigma_{22}^{-1}(\boldsymbol{x}_2 - \widehat{\boldsymbol{x}}_2)}{\eta + n}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T).$$

In the following, we re-visit the Bayesian formulation of t-distributed estimation problem as advocated in [4,6].

2.1. Time-update

We assume that $p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1})$ is t-distributed, i.e., $p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1}) = \mathrm{St}(\widehat{\boldsymbol{x}}_{k-1|k-1}, \Sigma_{k-1|k-1}, \eta)$. To evaluate $p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1})$, we linearize (1a) using the *statistical linear regression* (SLR) approximation [6,13,14,25]:

$$\boldsymbol{x}_k \approx F_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{b}_{k-1} + \boldsymbol{e}_{k-1}^f + \boldsymbol{w}_{k-1}, \tag{2}$$

where $F_{k-1} \in \mathbb{R}^{n\times n}$, $\boldsymbol{b}_{k-1} \in \mathbb{R}^n$ and $\boldsymbol{e}_{k-1}^f \in \mathbb{R}^n$ are to be determined. Taking the conditional mean of both sides of $f(\boldsymbol{x}_{k-1}) \approx F_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{b}_{k-1} + \boldsymbol{e}_{k-1}^f$, we obtain the value of $\boldsymbol{b}_{k-1}$ as follows:

$$\boldsymbol{b}_{k-1} = \mathrm{E}[(f(\boldsymbol{x}_{k-1}) - F_{k-1}\boldsymbol{x}_{k-1})|\boldsymbol{Y}_{k-1}] = \widetilde{\boldsymbol{x}}_{k|k-1} - F_{k-1}\widehat{\boldsymbol{x}}_{k-1|k-1} \tag{3}$$

where

$$\tilde{\boldsymbol{x}}_{k|k-1} := \int f(\boldsymbol{x}_{k-1}) p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1})\, d\boldsymbol{x}_{k-1} \tag{4}$$

Note that, for a linear system, $\boldsymbol{b}_{k-1} = 0$. The value of F_{k-1} is evaluated by minimizing the mean square linearization error, i.e.,

$$\begin{aligned} F^{\dagger}_{k-1} = \operatorname*{argmin}_{F} \mathrm{E}[(f(\boldsymbol{x}_{k-1}) - F_{k-1}\boldsymbol{x}_{k-1} - \boldsymbol{b}_{k-1})^T \\ \times (f(\boldsymbol{x}_{k-1}) - F_{k-1}\boldsymbol{x}_{k-1} - \boldsymbol{b}_{k-1})|\boldsymbol{Y}_{k-1}]. \end{aligned} \tag{5}$$

The closed-form solution is: $F_{k-1} = (P^{xf}_{k-1})^T P^{-1}_{k-1|k-1}$, where $P_{k-1|k-1} = \frac{\eta}{\eta-2}\Sigma_{k-1|k-1}$, $\Sigma_{k-1|k-1}$ is as specified in $p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1})$, and

$$P^{xf}_{k-1} := \mathrm{E}[(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1|k-1})(f(\boldsymbol{x}_{k-1}) - \tilde{\boldsymbol{x}}_{k|k-1})^T|\boldsymbol{Y}_{k-1}]. \tag{6}$$

Note that the linearization error $\boldsymbol{e}^f_{k-1}$, in the SLR framework, is generally assumed to be Gaussian distributed. In our case, however, we require it to be t-distributed to develop an approximate closed-form solution, i.e., $\boldsymbol{e}^f_{k-1}$ is assumed to be distributed in $\mathrm{St}(\boldsymbol{0}, \frac{\eta-2}{\eta}\Omega^f_{k-1}, \eta)$. Using the expression for F_{k-1}, the covariance matrix of $\boldsymbol{e}^f_{k-1} \approx f(\boldsymbol{x}_{k-1}) - F_{k-1}\boldsymbol{x}_{k-1} - \boldsymbol{b}_{k-1}$ is evaluated as

$$\Omega^f_{k-1} := \mathrm{E}[\boldsymbol{e}^f_{k-1}(\boldsymbol{e}^f_{k-1})^T] = P^{ff}_{k-1} - F_{k-1}P_{k-1|k-1}F^T_{k-1} \tag{7}$$

where

$$P^{ff}_{k-1} := \int (f(\boldsymbol{x}_{k-1}) - \tilde{\boldsymbol{x}}_{k|k-1})\,(f(\boldsymbol{x}_{k-1}) - \tilde{\boldsymbol{x}}_{k|k-1})^T p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1})\, d\boldsymbol{x}_{k-1}. \tag{8}$$

We note that $\boldsymbol{x}_{k-1}$, $\boldsymbol{e}^f_{k-1}$ and $\boldsymbol{w}_{k-1}$ are uncorrelated t-distributed variables with same degrees of freedom. Therefore, as suggested in [4], we assume that the joint-distribution $p(\boldsymbol{x}_{k-1}, \boldsymbol{e}^f_{k-1}, \boldsymbol{w}_{k-1})$ is also t-distributed with density $\mathrm{St}([\hat{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{0}, \boldsymbol{0}]^T, \mathrm{diag}\{[\Sigma_{k-1|k-1}, \frac{\eta-2}{\eta}\Omega^f_{k-1}, Q_{k-1}]\}, \eta)$.[1] From (2), we note that, after the linearization, we may write

$$\boldsymbol{x}_k = [F_{k-1} \quad I \quad I]\begin{bmatrix} \boldsymbol{x}_{k-1} \\ \boldsymbol{e}^f_{k-1} \\ \boldsymbol{w}_{k-1} \end{bmatrix} + \boldsymbol{b}_{k-1}, \tag{9}$$

and using **P1**, we can write $p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}) = \mathrm{St}(\hat{\boldsymbol{x}}_{k|k-1}, \Sigma_{k|k-1}, \eta)$, where

$$\begin{aligned} \hat{\boldsymbol{x}}_{k|k-1} &= \mathrm{E}[\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}] = F_{k-1}\hat{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{b}_k \\ &= F_{k-1}\hat{\boldsymbol{x}}_{k-1|k-1} + \tilde{\boldsymbol{x}}_{k|k-1} - F_{k-1}\hat{\boldsymbol{x}}_{k-1|k-1} = \tilde{\boldsymbol{x}}_{k|k-1}. \end{aligned} \tag{10}$$

and

$$\begin{aligned} \Sigma_{k|k-1} &= \frac{\eta-2}{\eta} \cdot \mathrm{cov}[\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}] \\ &= \frac{\eta-2}{\eta} \cdot \left(F_{k-1}P_{k-1|k-1}F^T_{k-1} + \Omega^f_{k-1} + \frac{\eta}{\eta-2}Q_{k-1}\right) \\ &= \frac{\eta-2}{\eta} \cdot P^{ff}_{k-1} + Q_{k-1} \end{aligned} \tag{11}$$

Expressions (10) and (11) dictate that the evaluation of expression F_{k-1} (or P^{xf}_{k-1}) is not required.

[1] Note that uncorrelated random variables with joint t-elliptic distribution are in general 'not independent' [4]; however, despite this limitation, they have been successfully employed in the development of robust filtering solutions in a number of works in literature [2,6,24].

2.2. Measurement-update

Following the steps provided in Section 2.1, we linearize measurement Eq. (1b) as follows:

$$\boldsymbol{y}_k \approx H_k\boldsymbol{x}_k + \boldsymbol{c}_k + \boldsymbol{e}^h_k + \boldsymbol{v}_k, \tag{12}$$

where, $H_k = (P^{xh}_k)^T P^{-1}_{k|k-1}$, $\boldsymbol{c}_k = \hat{\boldsymbol{y}}_{k|k-1} - H_k\hat{\boldsymbol{x}}_{k|k-1}$, $\boldsymbol{e}^h_k \sim \mathrm{St}(\boldsymbol{0}, \frac{\eta-2}{\eta}\Omega^h_k, \eta)$, where $P_{k|k-1} = \frac{\eta}{\eta-2}\Sigma_{k|k-1}$ and the covariance matrix of $\boldsymbol{e}^h_k$ is expressed as:

$$\Omega^h_k := \mathrm{E}[\boldsymbol{e}^h_k(\boldsymbol{e}^h_k)^T] = P^{hh}_k - H_kP_{k|k-1}H^T_k. \tag{13}$$

Again, the linearization error $\boldsymbol{e}^h_k$ has been assumed to be t-distributed to facilitate obtaining a closed-form solution for $\hat{\boldsymbol{x}}_{k|k}$ as described in the sequel. Further, the expressions for $\hat{\boldsymbol{y}}_{k|k-1}$, P^{xh}_k and P^{hh}_k are given as follows:

$$\hat{\boldsymbol{y}}_{k|k-1} = \int h(\boldsymbol{x}_k) p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1})\, d\boldsymbol{x}_k, \tag{14}$$

$$P^{xh}_k = \int (\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1})(h(\boldsymbol{x}_k) - \hat{\boldsymbol{y}}_{k|k-1})^T p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1})\, d\boldsymbol{x}_k, \tag{15}$$

$$P^{hh}_k = \int (h(\boldsymbol{x}_k) - \hat{\boldsymbol{y}}_{k|k-1})(h(\boldsymbol{x}_k) - \hat{\boldsymbol{y}}_{k|k-1})^T p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1})\, d\boldsymbol{x}_k. \tag{16}$$

Again we assume that $\boldsymbol{e}^h_k$ and $\boldsymbol{v}_k$ are distributed such that $[\boldsymbol{x}_k, \boldsymbol{e}^h_k, \boldsymbol{v}_k]^T$ is jointly t-distributed, i.e., $p(\boldsymbol{x}_k, \boldsymbol{e}^h_k, \boldsymbol{v}_k|\boldsymbol{Y}_{k-1}) = \mathrm{St}([\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{0}, \boldsymbol{0}]^T, \mathrm{diag}\{[\Sigma_{k|k-1}, \frac{\eta-2}{\eta}\Omega^h_k, R_k]\}, \eta)$. Using (12), we can write

$$\begin{bmatrix} \boldsymbol{x}_k \\ \boldsymbol{y}_k \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ H_k & I & I \end{bmatrix}\begin{bmatrix} \boldsymbol{x}_k \\ \boldsymbol{e}^h_k \\ \boldsymbol{v}_k \end{bmatrix}. \tag{17}$$

Exploiting **P1**, the joint probability of $\boldsymbol{x}_k$ and $\boldsymbol{y}_k$ is given as

$$\begin{aligned} &p(\boldsymbol{x}_k, \boldsymbol{y}_k|\boldsymbol{Y}_{k-1}) \\ &= \mathrm{St}\left(\begin{bmatrix} \hat{\boldsymbol{x}}_{k|k-1} \\ H_k\hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{c}_k \end{bmatrix}, \begin{bmatrix} \Sigma_{k|k-1} & \Sigma_{k|k-1}H^T_k \\ H_k\Sigma_{k|k-1} & S_k \end{bmatrix}, \eta\right), \end{aligned} \tag{18}$$

where $S_k = H_k\Sigma_{k|k-1}H^T_k + R_k + \frac{\eta-2}{\eta}\Omega^h_k = (\eta-2)/\eta \cdot P^{hh}_k + R_k$. Furthermore, we note that $H_k\hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{c}_k = \hat{\boldsymbol{y}}_{k|k-1}$ and $\Sigma_{k|k-1}H^T_k = (\eta-2)/\eta \cdot P^{xh}_k$. Finally, using **P2**, the conditional probability of $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$ can be written as $p(\boldsymbol{x}_k|\boldsymbol{Y}_k) = \mathrm{St}(\hat{\boldsymbol{x}}_{k|k}, \hat{\Sigma}_{k|k}, \tilde{\eta})$, where

$$\tilde{\eta} = \eta + m, \tag{19a}$$

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + K_k(\boldsymbol{y}_k - \hat{\boldsymbol{y}}_{k|k-1}), \tag{19b}$$

$$\hat{\Sigma}_{k|k} = \frac{\eta + \delta^2_k(\boldsymbol{y}_k)}{\eta + m}(I - K_kH_k)\Sigma_{k|k-1}, \tag{19c}$$

where $K_k = \Sigma_{k|k-1}H^T_kS^{-1}_k$ and $\delta^2_k(\boldsymbol{y}_k) = (\boldsymbol{y}_k - \hat{\boldsymbol{y}}_{k|k-1})^TS^{-1}_k(\boldsymbol{y}_k - \hat{\boldsymbol{y}}_{k|k-1})$. Note that, after the measurement-update, we have $\tilde{\eta} > \eta$; however, we require $\tilde{\eta}$ to be equal to η to facilitate the time-update step in the next iteration. To resolve this, owing to Dunik et al. [24], we adopt $p(\boldsymbol{x}_k|\boldsymbol{Y}_k) \approx \mathrm{St}(\hat{\boldsymbol{x}}_{k|k}, \Sigma_{k|k}, \eta)$, where $\Sigma_{k|k}$ is scaled appropriately by matching moments as follows:

$$\Sigma_{k|k} = \left(\frac{\tilde{\eta}}{\tilde{\eta}-2}\right)\left(\frac{\eta-2}{\eta}\right)\hat{\Sigma}_{k|k}. \tag{20}$$

3. Stochastic integration method

Note that the Bayesian filtering process described above requires the evaluation of Student t-weighted integrals of the form $I(s) = \int s(\boldsymbol{x})\,\mathrm{St}(\hat{\boldsymbol{x}}, \Sigma, \eta)d\boldsymbol{x}$ in (4), (8) and (14)–(16). These integrals, in general, do not admit closed-form solutions. In contrast to conventional sigma-point methods, here, in this work, we suggest to apply stochastic integration methods [11] for an approximate yet accurate evaluation of $I(s)$.

First, we transform $\boldsymbol{x} = \widehat{\boldsymbol{x}} + \Sigma^{1/2}\boldsymbol{c}$, where $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^T$ and $\Sigma^{1/2}$ is a lower triangular matrix stemming from the Cholesky decomposition of Σ. Accordingly, the Student t-weighted integral is written as $\int s(\widehat{\boldsymbol{x}} + \Sigma^{1/2}\boldsymbol{c})\,\mathrm{St}(\boldsymbol{0}, I, \eta)\mathrm{d}\boldsymbol{c} = \int g(\boldsymbol{c})\,\mathrm{St}(\boldsymbol{0}, I, \eta)\mathrm{d}\boldsymbol{c} =: I(g)$, where $g(\boldsymbol{c}) := s(\widehat{\boldsymbol{x}} + \Sigma^{1/2}\boldsymbol{c})$. Secondly, we introduce a change of variable to convert the multi-variate integral into a radial-spherical form [11], i.e., we let $\boldsymbol{c} = r\boldsymbol{z}$, with $\boldsymbol{z}\boldsymbol{z}^T = 1$, $r^2 = \boldsymbol{c}^T\boldsymbol{c}$, and

$$w(||\boldsymbol{c}||) := \frac{\Gamma(\frac{\eta+n}{2})}{\Gamma(\frac{\eta}{2})} \frac{1}{(\eta\pi)^{n/2}} \left(1 + \frac{\boldsymbol{c}^T\boldsymbol{c}}{\eta}\right)^{-(\eta+n)/2}$$

which facilitates us to obtain

$$\begin{aligned} I(g) &= \int_{\|\boldsymbol{z}\|=1} \int_0^\infty w(r) r^{n-1} g(r\boldsymbol{z}) \mathrm{d}r \mathrm{d}\boldsymbol{z} \\ &= \frac{1}{2} \int_{\|\boldsymbol{z}\|=1} \int_{-\infty}^\infty w(r)|r|^{n-1} g(r\boldsymbol{z}) \mathrm{d}r \mathrm{d}\boldsymbol{z}, \end{aligned} \tag{21}$$

We approximate the radial integral stochastically as follows:

$$I_r(g) = \int_{-\infty}^\infty w(r)|r|^{(n-1)} g(r)\mathrm{d}r \tag{22a}$$

$$\approx \sum_{i=0}^{N_r} \varpi_{r,i} \left[\frac{g(\rho_i) + g(-\rho_i)}{2}\right], \tag{22b}$$

where the weights $\{\varpi_{r,\,i}\}$ with a set of random points $\{\rho_i\}$ are selected such that (22b) becomes a dth-degree integration rule for (22a). Similarly, we have a spherical rule

$$I_{\boldsymbol{z}}(g) = \int_{\|\boldsymbol{z}\|=1} g(\boldsymbol{z})\mathrm{d}\boldsymbol{z} \approx \sum_{j=0}^{N_s} \varpi_{s,j} g(\mathcal{Q}\boldsymbol{z}_j). \tag{23}$$

Combining (22b) and (23), a product stochastic spherical-radial rule is obtained to approximate $I(g)$ as follows:

$$I(g) \approx \frac{1}{2} \sum_{j=0}^{N_s} \varpi_{s,j} \sum_{i=0}^{N_r} \varpi_{r,i} \left[\frac{g(\rho_i\mathcal{Q}\boldsymbol{z}_j) + g(-\rho_i\mathcal{Q}\boldsymbol{z}_j)}{2}\right]. \tag{24}$$

where $\{\varpi_{s,\,j}\}$ are random weights, and $\mathcal{Q}$ is an orthogonal matrix.

Remark 2. The spherical-radial rule in (24) is a dth-degree rule if **a)** it is exact for a $g(\boldsymbol{x})$ that can be described by a linear combination of monomials up to degree d. **b)** It is not exact for at least one monomial of degree $d+1$. Moreover, if the radial rule in (22b) and the spherical rule in (23) are both dth-degree, then the resulting spherical-radial rule in (24) is dth-degree as well [9].

3.1. Stochastic radial rule

To realize the radial rule (22b), we have a proposition:

Proposition 1. *[11]: If weights $\varpi_{r,\,i}$ in (22b) are defined by*

$$\varpi_{r,i} = I_r\left(\prod_{k=0,k\neq i}^{N_r} \frac{r^2 - \rho_k^2}{\rho_i^2 - \rho_k^2}\right), \tag{25}$$

where $\rho_0 = 0$ and $\{\rho_i\}$ are distinct non-negative real numbers chosen from a distribution proportional to $p(\rho_1, \rho_2, \cdots, \rho_{N_r}) = \prod_{i=1}^{N_r} \rho_i^{n+1} w(\rho_i) \prod_{k=1}^{i-1} (\rho_i - \rho_k)^2 (\rho_i + \rho_k)$, then (22b) is an unbiased degree $2N_r + 1$ integration rule for $I_r(g)$.

3.2. Stochastic spherical rule

A stochastic spherical rule can be developed by modifying a given deterministic rule; we have the following proposition:

Proposition 2. *[11]: Let $S(g) = \sum_{j=0}^{N_s} \varpi_{s,j} g(\boldsymbol{z}_j)$ be an integration rule of degree d for the integral $I_{\boldsymbol{z}}(g)$. If $\mathcal{Q}$ is a uniformly chosen $n \times n$ orthogonal matrix, then $S_{\mathcal{Q}}(g) = \sum_{j=0}^{N_s} \varpi_{s,j} g(\mathcal{Q}\boldsymbol{z}_j)$ is also an unbiased integration rule of degree d for $I_{\boldsymbol{z}}(g)$.*

Remark 3. We can develop a stochastic spherical rule of an arbitrary degree using Proposition 2 and any of the various deterministic rules available in the literature [9,26]. The standard method for generating $\mathcal{Q}$ is to set it equal to the Q matrix of the QR-factorization of an $n \times n$ random matrix X, where each entry of X is independent and distributed in $\mathcal{N}(0, 1)$. More efficient methods can be found in [27].

3.3. Third-degree stochastic spherical radial rule

To develop a third-degree stochastic radial rule ($N_r = 1$), we note from Proposition 1 that the corresponding weights $\varpi_{r,\,0}$ and $\varpi_{r,\,1}$ are evaluated as follows:

$$\varpi_{r,0} = I_r\left(\frac{\rho^2 - r^2}{\rho^2}\right) = \left(1 - \frac{n\eta}{(\eta-2)\rho^2}\right)T \tag{26a}$$

$$\varpi_{r,1} = I_r\left(\frac{r^2}{\rho^2}\right) = \frac{n\eta}{(\eta-2)\rho^2}T, \tag{26b}$$

where $T = \pi^{-n/2}\Gamma(n/2)$. To generate ρ, we note from Proposition 1 that we need to sample from $p(\rho) \propto \rho^{n+1}(1 + \rho^2/\eta)^{-(n+\eta)/2}$. While $p(\rho)$ is not a standard distribution; however, it can be sampled using a Beta distribution with an appropriate transformation as dictated by the following proposition:

Proposition 3. *Let $y \sim \mathrm{Beta}(\eta/2 - 1, n/2 + 1)$ and let ζ be a non-negative real number, $\zeta = \sqrt{\eta(1-y)/y}$, then $p(\zeta) \propto p(\rho)$.*

Proof. Considering $p(\zeta) = p(y)|\frac{dy}{d\zeta}||_{y=\widehat{y}}$, where $\widehat{y} = (1 + \zeta^2/\eta)^{-1}$; $p(\zeta) = \frac{1}{K}\zeta^{n+1}\left(1 + \zeta^2/\eta\right)^{-(n+\eta)/2}$ is found to be proportional to $p(\rho)$ if ζ is replaced with ρ. The positive constant K is evaluated as $K = \frac{1}{2}\frac{\Gamma(\eta/2-1)\Gamma(n/2+1)}{\Gamma(\eta/2+n/2)}\eta^{(n+2)/2}$.

Note that the mentioning of the Beta distribution appears lightly in [11, p. 9] without any discussion of the required transformation. Accordingly, to generate ρ, we choose y from $\mathrm{Beta}(\eta/2 - 1, n/2 + 1)$ and then apply the transformation $\rho = \sqrt{\eta(1-y)/y}$.

For the third-degree stochastic spherical rule, we first employ the simple deterministic rule given as

$$I_{\boldsymbol{z}}(g) \approx \frac{2\varpi_s}{T} \sum_{j=1}^{n} \left[\frac{g(\boldsymbol{e}_j) + g(-\boldsymbol{e}_j)}{2}\right], \tag{27}$$

where $2/T$ is the surface area of unit sphere, $\boldsymbol{e}_j$ is a unit vector in the jth coordinate and $\varpi_s = \frac{1}{n}$. We then make use of Proposition 2 to convert it into a stochastic rule. Finally, using (24)–(27), the integral $I(g)$ in (21) can be approximated using the stochastic spherical-radial rule as follows:

$$\begin{aligned} I(g) &\approx \frac{\varpi_s}{T} \sum_{j=1}^{n} [\varpi_{r,0}\, g(\boldsymbol{0}) + \varpi_{r,1}\, \widetilde{g}(\rho\mathcal{Q}\boldsymbol{e}_j)] \\ &= \varpi_0\, g(\boldsymbol{0}) + \frac{\eta}{(\eta-2)\rho^2} \sum_{j=1}^{n} \widetilde{g}(\rho\mathcal{Q}\boldsymbol{e}_j), \end{aligned} \tag{28}$$

where $\varpi_0 = 1 - n\eta/((\eta-2)\rho^2)$, and $\widetilde{g}(\boldsymbol{x}) = \frac{1}{2}(g(\boldsymbol{x}) + g(-\boldsymbol{x}))$. □

Remark 4. To achieve global convergence, the stochastic integration is evaluated N_m times and averaged. For each evaluation, independent realizations of random entities ρ and $\mathcal{Q}$ are considered. From (28), we note that each iteration operates for $2n+1$ points. Hence, the total number of function evaluations required is $(2n+1)N_m$. Note that the CKF [15] requires $2n$ function evaluations; consequently, the complexity of the proposed scheme is approximately N_m times greater than that of CKF filter.

4. Stochastic integration filtering

Here, we describe the procedure to recursively estimate $\widehat{\boldsymbol{x}}_{k|k}$ using the stochastic rule described in Section 3.3. The filter is initialized with $\widehat{\boldsymbol{x}}_{0|0} = \mathrm{E}[\boldsymbol{x}_0]$ and $\Sigma_{0|0} = (\eta-2)/\eta E[(\boldsymbol{x}_0 - \widehat{\boldsymbol{x}}_{0|0})(\boldsymbol{x}_0 - \widehat{\boldsymbol{x}}_{0|0})^T]$. The filtering procedure is carried out by repeating the following steps for each instance k.

For the time-update, we set $\boldsymbol{\mu} = \widehat{\boldsymbol{x}}_{k-1|k-1}$, $\Sigma = \Sigma_{k-1|k-1}$ and generate independent realizations of ρ_1^l and $\mathcal{Q}^l$ for $l = 1, 2, \cdots, N_m$. Then, for each l, we generate the following set of sigma-points:

$$\boldsymbol{X}_i^l = \boldsymbol{\mu} + \Sigma^{1/2}\rho^l\mathcal{Q}^l\boldsymbol{e}_i \qquad 0 < i \le n. \tag{29}$$

Let $f_1(\boldsymbol{x}) = f(\boldsymbol{x})$, $f_2(\boldsymbol{x}) = f(\boldsymbol{x})f(\boldsymbol{x})^T$, and $\widetilde{f}_i(\boldsymbol{x}) = \frac{1}{2}(f_i(\boldsymbol{x}) + f_i(-\boldsymbol{x}))$ for $i = 1, 2$. Then, using (28), the integral in (8) are approximated as

$$\widehat{\boldsymbol{x}}_{k|k-1} = \widetilde{\boldsymbol{x}}_{k|k-1} \approx \frac{1}{N_m}\sum_{l=1}^{N_m}\left[f_1(\boldsymbol{\mu})\varpi_0^l + \frac{\eta}{(\eta-2)(\rho^l)^2}\sum_{j=1}^{n}\widetilde{f}_1(\boldsymbol{X}_i^l)\right], \tag{30a}$$

$$P_{k-1}^{ff} \approx \frac{1}{N_m}\sum_{l=1}^{N_m}\left[f_2(\boldsymbol{\mu})^T\varpi_0^l + \frac{\eta}{(\eta-2)(\rho^l)^2}\sum_{j=1}^{n}\widetilde{f}_2(\boldsymbol{X}_i^l)\right] - \widehat{\boldsymbol{x}}_{k|k-1}\widehat{\boldsymbol{x}}_{k|k-1}^T. \tag{30b}$$

Finally, we evaluate $\Sigma_{k|k-1} = \frac{\eta-2}{\eta}\cdot P_{k-1}^{ff} + Q_{k-1}$. For the measurement-update, we set $\boldsymbol{\mu} = \widehat{\boldsymbol{x}}_{k|k-1}$, $\Sigma = \Sigma_{k|k-1}$ and generate a new set of sigma-points using (29). Let $h_1(\boldsymbol{x}) = h(\boldsymbol{x})$, $h_2(\boldsymbol{x}) = \boldsymbol{x}h(\boldsymbol{x})^T$, $h_3(\boldsymbol{x}) = h(\boldsymbol{x})h(\boldsymbol{x})^T$ and $\widetilde{h}_i(\boldsymbol{x}) = \frac{1}{2}(h_i(\boldsymbol{x}) + h_i(-\boldsymbol{x}))$, $i = 1, 2, 3$. Now using (28), the integrals in (14), (15) and (16) are approximated as

$$\widehat{\boldsymbol{y}}_{k|k-1} \approx \frac{1}{N_m}\sum_{l=1}^{N_m}\left[h_1(\boldsymbol{\mu})\varpi_0^l + \frac{\eta}{(\eta-2)(\rho^l)^2}\sum_{j=1}^{n}\widetilde{h}_1(\boldsymbol{X}_i^l)\right], \tag{31a}$$

$$P_k^{xh} \approx \frac{1}{N_m}\sum_{l=1}^{N_m}\left[h_2(\boldsymbol{\mu})^T\varpi_0^l + \frac{\eta}{(\eta-2)(\rho^l)^2}\sum_{j=1}^{n}\widetilde{h}_2(\boldsymbol{X}_i^l)\right] - \widehat{\boldsymbol{x}}_{k|k-1}\widehat{\boldsymbol{y}}_{k|k-1}^T, \tag{31b}$$

$$P_k^{hh} \approx \frac{1}{N_m}\sum_{l=1}^{N_m}\left[h_3(\boldsymbol{\mu})^T\varpi_0^l + \frac{\eta}{(\eta-2)(\rho^l)^2}\sum_{j=1}^{n}\widetilde{h}_3(\boldsymbol{X}_i^l)\right] - \widehat{\boldsymbol{y}}_{k|k-1}\widehat{\boldsymbol{y}}_{k|k-1}^T. \tag{31c}$$

Finally, $\widehat{\boldsymbol{x}}_{k|k}$ and $\widehat{\Sigma}_{k|k}$ are evaluated using (19) and (20), respectively.

5. Simulation results

In this section, we compare the performance of the proposed filter with the conventional CKF [15], the SIF [10] and a recently developed robust CKF (RCKF) filter [2,6] which is a multivariate t-filter. We consider a benchmark scenario of tracking a target using bearings-only measurements [10,28]. The target is observed from a maneuvering platform and the angle between the target and the platform is measured at each instance k. We assume a nearly constant velocity motion for the target, i.e., the process model is described as follows [28]:

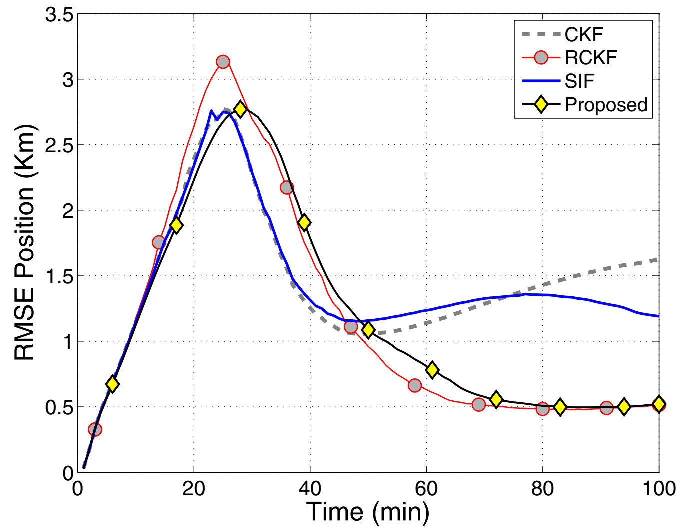

Fig. 1. Comparison of position RMSE of the proposed filter RSIF with CKF, SIF and RCKF for jointly t-distributed process and measurement noises.

$$\boldsymbol{x}_k = F\boldsymbol{x}_{k-1} + \boldsymbol{w}_{k-1}, \tag{32}$$

where $\boldsymbol{x}_k = [\zeta_k, \dot{\zeta}_k, \epsilon_k, \dot{\epsilon}_k]^T$, $F = F_1 \otimes I_2$, $[\zeta_k, \epsilon_k]$ are the position coordinates of the target; whereas, ζ_k and $\dot{\epsilon}_k$ are the velocities in ζ and ϵ directions, respectively. We have $F_1 = \begin{bmatrix}1 & T\\ 0 & 1\end{bmatrix}$ and the sampling time T is set to 1 min. The measurement model is describes as

$$y_k = \tan^{-1}\left(\frac{\epsilon_k - \epsilon_{p,k}}{\zeta_k - \zeta_{p,k}}\right) + v_k, \tag{33}$$

where $[\zeta_{p,\,k}, \epsilon_{p,\,k}]^T$ are the coordinates of the platform and are assumed to be known. The object follows a course of −2.36 rad at a constant speed of 27.7 Km/h, i.e., 15 knots. The platform follows a course of −1.39 rad at a constant speed of 9.26 Km/h. The platform undergoes a maneuver after 15 min after which it follows a course of 2.54 rad. The platform starts from the origin; whereas, the initial position of the target is set to [7 Km, 7 Km]. The initial error covariance matrix $P_{0|0}$ is set to diag([400 m, 4 m/s,400 m, 4 m/s]). The total simulation time is set to 100 min. The filtering parameters N_m and η are set to 100 and 4, respectively.

5.1. Experiment No. 1

In the first experiment, we generate the uncertainties $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ using t-distributed random processes, i.e., $\boldsymbol{w}_k \sim \mathrm{St}(\boldsymbol{0}, Q, \eta)$ and $\boldsymbol{v}_k \sim \mathrm{St}(\boldsymbol{0}, \sigma_\theta^2, \eta)$, where $Q = (Q_1 \otimes I_2)\sigma_w^2$, $Q_1 = \begin{bmatrix}T^{3/2} & T^{2/2}\\ T^{2/2} & T\end{bmatrix}$, $\sigma_w^2 = 1\times 10^{-6}\,\mathrm{m}^2/\mathrm{s}^3$ and $\sigma_\theta = 2^\circ$. We compare the performances using root-mean-square-error (RMSE) values of positions and velocities that are computed by taking the average of 1000 Monte-Carlo runs.

The results are depicted in Figs. 1 and 2. We note that the filters utilizing t-distributed process and measurement noise, i.e., the RCKF and the proposed RSIF significantly outperform the standard Gaussian assumption based CKF and SIF. Furthermore, we observe that the proposed RSIF performs better than robust CKF filter in terms of peak RMSE; however, RCKF performs slightly better than the proposed scheme during the transition period and both filters show similar performance in steady-state.

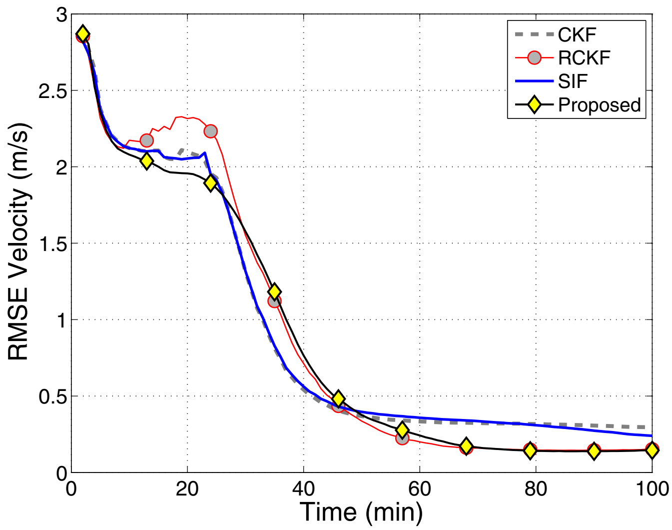

Fig. 2. Comparison of velocity RMSE of the proposed filter RSIF with CKF, SIF and RCKF for jointly *t*-distributed process and measurement noises.

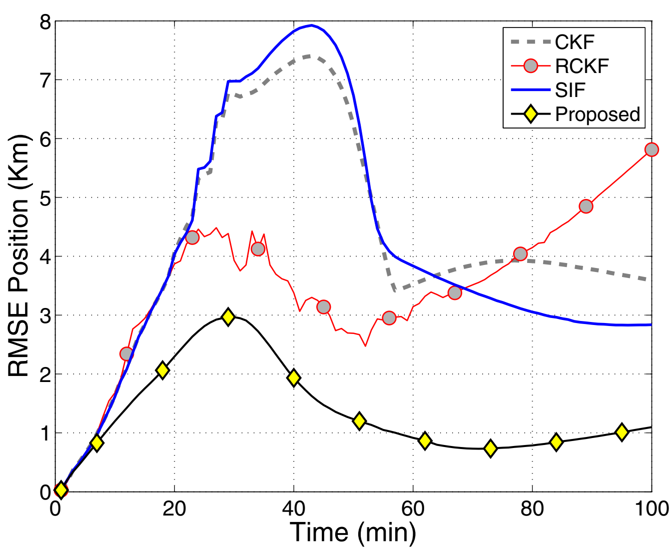

Fig. 3. Comparison of position RMSE of the proposed filter RSIF with CKF, SIF and RCKF for outliers contaminated noise generated using (34).

5.2. Experiment No. 2

In the second experiment, we use a standard model, as suggested in [2,4,6], to simulate the outliers contaminated noise processes, i.e., $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are generated according to the following probability scheme: (below w.p. stands for 'with probability')

$$\boldsymbol{w}_k \sim \begin{cases} \mathcal{N}(\mathbf{0}, Q) & \text{w.p. } 0.95 \\ \mathcal{N}(\mathbf{0}, 10Q) & \text{w.p. } 0.05 \end{cases}$$

$$\boldsymbol{v}_k \sim \begin{cases} \mathcal{N}(\mathbf{0}, \sigma_\theta^2) & \text{w.p. } 0.95 \\ \mathcal{N}(\mathbf{0}, 100\sigma_\theta^2) & \text{w.p. } 0.05 \end{cases} \tag{34}$$

where $\sigma_\theta = 1^\circ = 17.5$ mrad.

The RMSE values are depicted in Figs. 3 and 4. We note that the proposed scheme outperforms all other filters in both position and velocity RMSE values. In Table 3, we depict the values of the position and velocity RMSE averaged over the entire simulation range. We note that the position estimates of the proposed filter show an improvement of around 68.1%, 67.1% and 61% as compared to CKF, SIF and RCKF, respectively. Similarly, the velocity RMSE of the proposed filter is lesser by 50.7%, 51.2% and 51.4% when compared to CKF, SIF and RCKF, respectively. Thus, under heavy-tailed uncertainties, the proposed filter seems to be quite a promising substitute for the traditional counterparts.

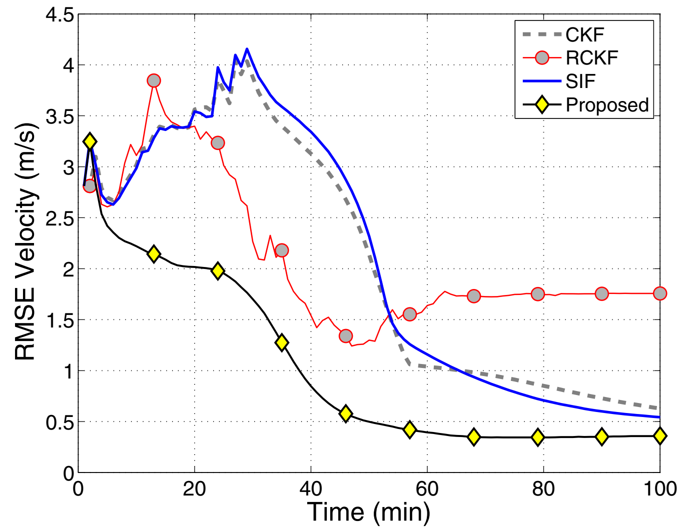

Fig. 4. Comparison of velocity RMSE of the proposed filter RSIF with CKF, SIF and RCKF for outliers contaminated noise generated using (34).

Table 3
RMSE values of the proposed RSIF filter, CKF, SIF and RCKF.

Filter	Position (Km)	Velocity (m/s)
Cubature Kalman Filter (CKF)	4.27	2.07
Stochastic Integration Filter (SIF)	4.14	2.09
Robust CKF	3.5	2.1
Proposed (Robust SIF)	1.36	1.02

6. Conclusions

In this work, we discussed the utilization of spherical-radial stochastic integration rules for nonlinear Bayesian filtering in the presence of multivariate *t*-distributed uncertainties. We specifically developed a robust third-degree stochastic integration filter (RSIF) for the estimation of state variables under heavy-tailed uncertainties. Under the influence of outliers, the performance of the proposed filter was compared with the standard cubature Kalman filter (CKF), the stochastic integration filter (SIF), and a robust variant of CKF in a nonlinear tracking scenario. It is observed that the proposed filter is highly robust to the presence of heavy-tailed uncertainties, and can outperform a number of existing solutions.

References

[1] A. Haug, Bayesian Estimation and Tracking: A Practical Guide, John Wiley and Sons, 2012.
[2] Y. Huang, Y. Zhang, N. Li, S. Naqvi, J. Chambers, A robust student *t*-based cubature filter, in: 19th International Conference on Information Fusion, 2016, pp. 9–16.
[3] G. Agamennoni, J. Nieto, E. Nebot, Approximate inference in state-space models with heavy-tailed noise, IEEE Trans. Signal Process. 60 (10) (2012) 5024–5037.
[4] M. Roth, E. Özkan, F. Gustafsson, A student *t*-filter for heavy tailed process and measurement noise, in: IEEE Intl. Conference on Acoustics, Speech and Signal Processing, 2013, pp. 5770–5774.
[5] R. Piché, S. Särkkä, J. Hartikainen, Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student *t*-distribution, in: IEEE Intl. Workshop on Machine Learning for Signal Processing, 2012, pp. 1–6.
[6] F. Tronarp, R. Hostettler, S. Särkkä, Sigma-point filtering for nonlinear systems with non-additive heavy-tailed noise, in: 19th International Conference on Information Fusion, 2016, pp. 1859–1866.
[7] R. Meinhold, N. Singpurwalla, Robustification of Kalman filter models, J. Am. Statis. Assoc. 84 (406) (1989) 479–486.

[8] Y. Huang, Y. Zhang, N. Li, J. Chambers, A robust Gaussian approximate fixed-interval smoother for nonlinear systems with heavy-tailed process and measurement noises, IEEE Signal Process. Lett. 23 (4) (2016) 468–472.
[9] B. Jia, M. Xin, Y. Cheng, High-degree cubature Kalman filter, Automatica 49 (2) (2013) 510–518.
[10] J. Duník, O. Straka, M. Šimandl, Stochastic integration filter, IEEE Trans. Autom. Contr. 58 (6) (2013) 1561–1566.
[11] A. Genz, J. Monahan, Stochastic integration rules for infinite regions, SIAM J. Sci. Comput. 19 (2) (1998) 426–439.
[12] S. Khalid, N. Rehman, S. Abrar, Higher-degree stochastic integration filtering, in: arXiv preprint, available at: arxiv.org/abs/1608.00337, 2016, pp. 1–5.
[13] I. Arasaratnam, S. Haykin, R. Elliott, Discrete-time nonlinear filtering algorithms using Gauss–Hermite quadrature, Proc. IEEE 95 (5) (2007) 953–977.
[14] A. García-Fernández, L. Svensson, M. Morelande, S. Särkkä, Posterior linearization filter: principles and implementation using sigma points, IEEE Trans. Signal Process. 63 (20) (2015) 5561–5573.
[15] I. Arasaratnam, S. Haykin, Cubature Kalman filters, IEEE Trans. Autom. Contr. 54 (6) (2009) 1254–1269.
[16] M. Al-Shabi, S. Gadsden, S. Habibi, Kalman filtering strategies utilizing the chattering effects of the smooth variable structure filter, Signal Process. 93 (2) (2013) 420–431.
[17] S. Gadsden, M. Al-Shabi, I. Arasaratnam, S. Habibi, Combined cubature Kalman and smooth variable structure filtering: a robust nonlinear estimation strategy, Signal Process. 96 (2014) 290–299.
[18] S. Julier, J. Uhlmann, Unscented filtering and nonlinear estimation, Proc. IEEE 92 (3) (2004) 401–422.
[19] B. Jia, M. Xin, Y. Cheng, Sparse–grid quadrature nonlinear filtering, Automatica 48 (2) (2012) 327–341.
[20] Y. Zhang, Y. Huang, N. Li, L. Zhao, Embedded cubature Kalman filter with adaptive setting of free parameter, Signal Process. 114 (2015) 112–116.
[21] J. Dunik, O. Straka, M. Simandl, E. Blasch, Random–point-based filters: analysis and comparison in target tracking, IEEE Trans. Aerosp. Electron. Syst. 51 (2) (2015) 1403–1421.
[22] S. Kotz, S. Nadarajah, Multivariate *t*-Distributions and Their Applications, Cambridge University Press, 2004.
[23] W. Shaw, K. Lee, Bivariate student *t*-distributions with variable marginal degrees of freedom and independence, J. Multivariate Anal. 99 (6) (2008) 1276–1287.
[24] M. Roth, Kalman Filters for Nonlinear Systems and Heavy-Tailed Noise, Division of Automatic Control, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, 2013 Ph.D. thesis.
[25] T. Lefebvre, H. Bruyninckx, J. De Schuller, Comment on "a new method for the nonlinear transformation of means and covariances in filters and estimators" [with authors' reply], IEEE Trans. Autom. Contr. 47 (8) (2002) 1406–1409.
[26] A. Genz, Fully symmetric interpolatory rules for multiple integrals over hyper--spherical surfaces, J. Comput. Appl. Math. 157 (1) (2003) 187–195.
[27] A. Genz, Methods for generating random orthogonal matrices, Monte Carlo and Quasi-Monte Carlo Methods (1998) 199–213.
[28] P. Leong, S. Arulampalam, T. Lamahewa, T. Abhayapala, A Gaussian-sum based cubature Kalman filter for bearings-only tracking, IEEE Trans. Aerosp. Electron. Syst. 49 (2) (2013) 1161–1176.