

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/171220/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Hao, Pengcheng, Karakus, Oktay and Achim, Alin 2024. Robust Kalman filters based on the sub-Gaussian α -stable distribution. *Signal Processing* 224 , 109574. 10.1016/j.sigpro.2024.109574

Publishers page: <http://dx.doi.org/10.1016/j.sigpro.2024.109574>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Robust Kalman Filters Based on the Sub-Gaussian α -stable Distribution

Pengcheng Hao^{a,*}, Oktay Karakuş^b, Alin Achim^a

^a*Visual Information Laboratory, University of Bristol, Bristol, U.K.*

^b*School of Computer Science and Informatics, Cardiff University, Cardiff, U.K.*

Abstract

Motivated by filtering tasks under a linear system with non-Gaussian heavy-tailed noise, various robust Kalman filters (RKF) based on different heavy-tailed distributions have been proposed. Although the sub-Gaussian α -stable (SG α S) distribution captures heavy tails well and is applicable in various scenarios, its potential has not yet been explored for RKF. The main hindrance is that there is no closed-form expression of its mixing density. This paper proposes a novel RKF framework, RKF-SG α S, where the process noise is assumed to be Gaussian and the heavy-tailed measurement noise is modelled by the SG α S distribution. The corresponding joint posterior distribution of the state vector and auxiliary random variables is approximated by the Variational Bayesian approach. Also, four different minimum mean square error (MMSE) estimators of the scale function are presented. The first two methods are based on the Importance Sampling (IS) and Gauss-Laguerre quadrature (GLQ), respectively. In contrast, the last two estimators combine a proposed Gamma series (GS) based method with the IS and GLQ estimators and hence are called GSIS and GSGL. Besides, the RKF-SG α S is compared with the state-of-the-art RKF under three kinds of heavy-tailed measurement noises, and the simulation results demonstrate its estimation accuracy and efficiency.

Keywords: sub-Gaussian α -stable distribution, Kalman filter, heavy-tailed noise, Variational Bayesian.

1. Introduction

Kalman filtering is a technique that uses a sequence of observations to estimate the unknown states of a system over time [1]. By judiciously defining the state vector of the relevant system, this algorithm can be employed and be beneficial in many applications. For example, in target tracking, the state vector represents the kinematics of the target [2]; in GPS navigation, Kalman filter (KF) is applied to GPS receiver position/velocity determination [3]; in robotics, the KF combines inertial, contact and kinematic data to estimate the robot's base pose, velocity and contact states [4]; in economics, it can capture interest rates, inflation, etc [5]. When the dynamic and observation models of the state-space model are both linear and Gaussian, the minimum mean squared error (MMSE) estimate can optimally be obtained by the KF. However, when the noise contains outliers and is hence heavy-tailed, the estimation precision of the KF degrades since

*Pengchang Hao is the Corresponding author.

Email addresses: ju18422@bristol.ac.uk (Pengcheng Hao), karakuso@cardiff.ac.uk (Oktay Karakuş), alin.achim@bristol.ac.uk (Alin Achim)

the Gaussianity assumption is violated.

To improve the performance of the KF under heavy-tailed noise, three main types of M-estimator-based robust KFs (RKF) have been proposed. These are the Huber-based RKF [6], which is a generalized maximum likelihood-type estimator that minimises the Huber function, the maximum correntropy Kalman filter [7], and the statistical similarity measure based KF [8]. Alternatively, one has proposed several outlier-detection-based RKFs. For example, [9, 10] detect the measurement outliers using the Chi-Square test, and variational Bayes strategies are employed in [11, 12]. However, these filters do not exploit the stochastic properties of the heavy-tailed noise, and hence their filtering performance is limited.

Different from M-estimator-based filters, another category of RKFs models the noise using heavy-tailed distributions. The Student's t (ST) distribution was the first model to be used for this purpose, and there were two types of approaches involving it. First, the ST-distribution-based RKFs assume that the measurements are contaminated by heavy-tailed noise. For example, in [13], the symmetric measurement noise is modelled as an ST distribution. Also, skewed noise is fitted by the skewed ST distribution in [14]. However, when the process noise is heavy-tailed, these filters are not applicable because of the Gaussian assumption which is violated. In a second group of techniques, RKFs are based on the assumption that both the one-step prediction and likelihood densities can be captured by ST distributions. In [15, 16], the ST-distribution-based filter considers that the heavy-tailed posterior probability density function (PDF) follows the ST distribution. However, its filtering accuracy is limited due to the requirement that the prediction and observation PDFs should be characterised by the same degree of impulsiveness. By contrast, in [17], the robust Student's t -based Kalman filter (RSTKF) eliminates this requirement and expresses the prediction and likelihood PDFs in hierarchical Gaussian forms. Hence, it exhibits higher computation complexity but produces more precise estimates. Despite all these improvements, two drawbacks of the RSTKF remain. First, due to the symmetric assumption, it only provides a rough estimation under skewed noise. Also, the degree of freedom (dof) value of the ST distribution is set manually for each filtering scenario, which degrades its overall performance.

Owing to these disadvantages, other heavy-tailed distributions have also been investigated for the purpose of designing RKFs. In [18, 19], an RKF based on the multivariate Laplace distribution was considered. As the Laplace distribution is only determined by its mean and covariance, the selection of the dof value is unnecessary. Nevertheless, without a shape parameter, the filter cannot adapt to varying types of noise. Hence, for heavy-tailed and/or skewed state and measurement noises, the RKF based on the Gaussian scale mixture (RKF-GSM) distribution was proposed in [20]. However, the MMSE estimate of the scale function is not available, and the maximum *a posteriori* probability (MAP) estimate introduces more errors. Also, the dof values are still manually set, which debases the estimation precision.

This work focuses on linear models with heavy-tailed sub-Gaussian α -stable ($\text{SG}\alpha\text{S}$) noise, and various RKFs can be used for the posterior estimation. However, the M-estimator-based and outlier-detection-based RKFs cannot exploit the stochastic characteristics of the $\text{SG}\alpha\text{S}$ noise, and the performance of the RKFs using heavy-tailed distributions might degrade due to the model error between the utilised heavy-tailed

distributions and the SG α S noise. Hence, a new RKF framework based on the SG α S distribution [21, 22] is considered in this work. The motivation for this choice is twofold. 1) The SG α S distribution belongs to the α -stable family of distributions and is hence theoretically motivated by the generalized Central Limit Theorem, which means many practical types of noise can be modelled by this density [23, 24]. For example, work by Lachmann in the early 20th century [25] shows that the SG α S distribution can well fit the noise in real audio data and filtering based on a SG α S distribution was proposed for music enhancement. Also, [26] and [27] utilise the SG α S distribution to inflation forecasting in macroeconomic analysis and acoustic source target tracking, respectively, where the isotropic complex symmetric α stable noise can be considered a SG α S variable [21]. Besides, the multiple access interference in ad hoc networks can be precisely represented by the SG α S distribution [28]. 2) Besides being a member of the more general family of α -stable distributions, the SG α S is also a special case of the GSM class of distributions [29, 30]. Hence, the design of an RKF framework based on the SG α S distribution can, in principle, be based on the structure of the RKF-GSM. However, as there is no closed-form expression for its mixing density, the MMSE estimation of the scale function cannot be achieved simply based on the RKF-GSM and hence remains challenging. Thus, despite its many appealing properties and various practical applications, the use of the SG α S distribution for the design of RKFs has not been investigated, and this constitutes the primary focus of this manuscript.

Specifically, the main contributions of this work consist in:

1. We propose a RKF framework based on the SG α S distribution (RKF-SG α S), where four different estimators are proposed to approximate the MMSE of the scale function. For the first two methods, the expectation of the scale function is approximated by the Importance Sampling (IS) [31] and Gauss-Laguerre quadrature (GLQ) methods [32], respectively. Two additional hybrid estimators, the gamma series IS (GSIS) and gamma series GLQ (GSGL), are proposed for situations when the proposed GS estimator diverges and is thus replaced with the IS and GLQ-based methods, respectively.
2. The properties of the four proposed RKF-SG α S variants based on the different estimators are thoroughly analysed and the RKF-SG α S is compared with the benchmark heavy-tailed-distribution-based RKFs in different measurement noise environments. Particularly, the parameters of the SG α S distribution and the benchmark GSM distributions are estimated based on the existing expectation maximisation (EM) and maximum likelihood estimation (MLE) methods, respectively, which reduces the model errors caused by the empirical parameters.

As a side contribution, for the RKF based on the slash distribution (RKF-SL)[20], which is one of the filters we benchmark against in the simulations section, we propose a novel MMSE estimate of the scale function in Appendix A.

The rest of the paper is organised as follows: we begin, in Section 2, with a description of the theoretical background on the Linear Discrete-Time State-Space Model and the SG α S distribution. In Section 3, we present the details of the proposed RKF-SG α S. The proposed framework is tested in target tracking scenarios in Section 4, whilst Section 5 concludes the paper with a summary.

2. Theoretical preliminary

In this section, we provide brief, necessary details, on the key concepts on which our main developments in subsequent sections build. We start by providing a list of the main mathematical symbols and notations used throughout the work, in Table 1.

Table 1: Mathematical Notations

<i>Notations</i>	<i>Definitions</i>
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$ $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	Multivariate Gaussian pdf with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
$S(\cdot; \alpha, \beta, \gamma, \delta)$	Univariate stable distribution pdf with the shape parameter α , the skewness parameter β , the scale parameter γ and the location parameter δ .
$G(\cdot; a, b)$	Gamma pdf with the shape parameter a and rate parameter b .
$\text{Be}(\cdot; a, b)$	Beta pdf with the shape parameter a and scale parameter b .
$\text{IG}(\cdot; a, b)$	Inverse-Gamma pdf with the shape parameter a and b .
$\text{IW}(\cdot; v, \boldsymbol{\Sigma})$	Inverse-Wishart pdf with the shape parameter a and scale matrix $\boldsymbol{\Sigma}$.
$\Gamma(\cdot), \psi(\cdot)$	Gamma function and Digamma function.
$\gamma(a, x)$	Lower Incomplete Gamma function defined as $\int_0^x t^{a-1} e^{-t} dt$.
$\mathbf{I}_n$	$n \times n$ identity matrix.
$\text{tr}\{\cdot\}, (\cdot)^T$	Trace operation and transpose operation.
$ \cdot $	Absolute value operation.
$\text{abs}(\mathbf{X})$	Element-wise absolute value operation.
$\text{sum}(\mathbf{X})$	Summation for all the elements.
$\mathbf{x} \sim p(\mathbf{x})$	The pdf of the variable $\mathbf{x}$ is $p(\mathbf{x})$.
$\stackrel{d}{=}, \stackrel{d}{\rightarrow}$	Equality and convergence in distribution.

2.1. Linear Discrete-Time State-Space Model

A linear discrete-time state-space model can be described as a probabilistic graph structure and includes two models. The dynamic noise describes the state transition over time, whilst the measurement model explains the relationship between the states and corresponding measurements. A generic linear discrete-time state-space model is written as

$$\begin{cases} \mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{w}_{k-1} \\ \mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \end{cases} \quad (1)$$

where $\mathbf{x}_k \in \mathbb{R}^n$ and $\mathbf{z}_k \in \mathbb{R}^m$ denote the hidden state and the measurement vectors at time k , respectively. Also the process and measurement noises are represented as $\mathbf{w}_k \in \mathbb{R}^n$ and $\mathbf{v}_k \in \mathbb{R}^m$, respectively. Additionally,

$\mathbf{F}_k \in \mathbb{R}^{n \times n}$ and $\mathbf{H}_k \in \mathbb{R}^{m \times n}$ are the state transition and measurement matrices. In this work, we assume $\mathbf{w}_k$ follows a Gaussian distribution, whilst $\mathbf{v}_k$ contains outliers and is thus heavy-tailed.

2.2. α -stable Distributions and the Sub-Gaussian Case

In this section, we first introduce the α -stable family of distributions whereby its sub-Gaussian subclass, the SG α S distribution, is also illustrated. Furthermore, the tail behaviour of this subclass is analysed.

In probability theory, the α -stable distribution is a family of probability distributions which is generally used to model heavy-tailed behaviour and outlier effects [33, 34, 23]. A stable random variable can be defined based on the generalized Central Limit Theorem [29]:

Definition 1. A random variable x has a stable distribution if it has a domain of attraction, i.e., if there is a sequence of i.i.d. random variables $y_1, y_2, \dots$ and sequence of positive numbers $(d_{n'})_{(n' \in \mathbf{N})}$ and real numbers $(D_{n'})_{(n' \in \mathbf{N})}$, such that

$$\frac{y_1 + y_2 + \dots + y_{n'}}{d_{n'}} + D_{n'} \xrightarrow{d} x$$

Also, the multivariate α -stable distribution extends the concept of the univariate stable distribution to high-dimensional cases. Thus, stable random vectors also exhibit the stability property and allow for the generalized Central Limit Theorem. However, we cannot directly apply the multivariate stable distribution to RKF frameworks due to the challenge of estimating the spectral measure of its characteristic function [29]. Instead, we focus on a more suitable subclass of the multivariate stable distribution, known as the SG α S distribution. As previously noted, the SG α S distribution is also a special case of the GSM distribution which can be written as

$$p(\mathbf{x}) = \int_0^{+\infty} \mathcal{N}(\mathbf{x}; \boldsymbol{\mu} + y\boldsymbol{\beta}, \boldsymbol{\Sigma}/\kappa(y))\pi(y)dy, \quad (2)$$

where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\beta}$ is the skewness vector. Also, $\boldsymbol{\Sigma}$ is the scale matrix, $y > 0$ is the mixing parameter, and $\kappa(y)$ and $\pi(y)$ are the scale function and mixing density, respectively. For the SG α S distribution [29], $\boldsymbol{\beta} = \mathbf{0}$, $\kappa(y) = 1/y$ and $\pi(y)$ can be represented as a totally skewed univariate stable distribution, i.e.,

$$p(\mathbf{x}) = \int_0^{+\infty} \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, y\boldsymbol{\Sigma})S(y; \alpha/2, 1, \cos(\pi\alpha/2)^{2/\alpha}, 0)dy. \quad (3)$$

Table 2 details the parameters of SG α S along with several GSM distributions investigated in this paper, where v refers to the shape parameters of the GSM distributions. For simplicity, we only consider the zero-mean symmetric heavy-tailed noise as in previous works [13, 35] where the mean and skewness vectors are zero for both the SG α S and the GSM distributions.

Subsequent to the succinct introduction of the SG α S distribution, we explain its tails behaviour, which is determined by its mixing density. Figure 1-(a) shows the mixing density is a Dirac delta distribution at 1 for $\alpha = 2$ and a totally skewed heavy-tailed stable distribution for $\alpha < 2$. Besides, smaller α values result in heavier tails as in Figure 1-(b). As the SG α S distribution is motivated by the generalized Central Limit Theorem, and has a hierarchical Gaussian form and flexible heavy tails, it has been employed in various practical applications.

Table 2: EXEMPLARY GSM DISTRIBUTIONS AND THEIR PARAMETERS

GSM distribution	$\kappa(y)$	$\pi(y)$	Constraints
SG α S	y^{-1}	$S(y; \alpha/2, 1, \cos(\pi\alpha/2)^{2/\alpha}, 0)$	$y > 0, 0 < \alpha \leq 2$
Student's t	y	$G(y; \frac{v}{2}, \frac{v}{2})$	$y > 0, v > 0$
Slash	y	$\text{Be}(y; \frac{v}{2}, 1)$	$0 < y < 1, v > 0$
Variance Gamma	y	$\text{IG}(y; \frac{v}{2}, \frac{v}{2})$	$y > 0, v > 0$

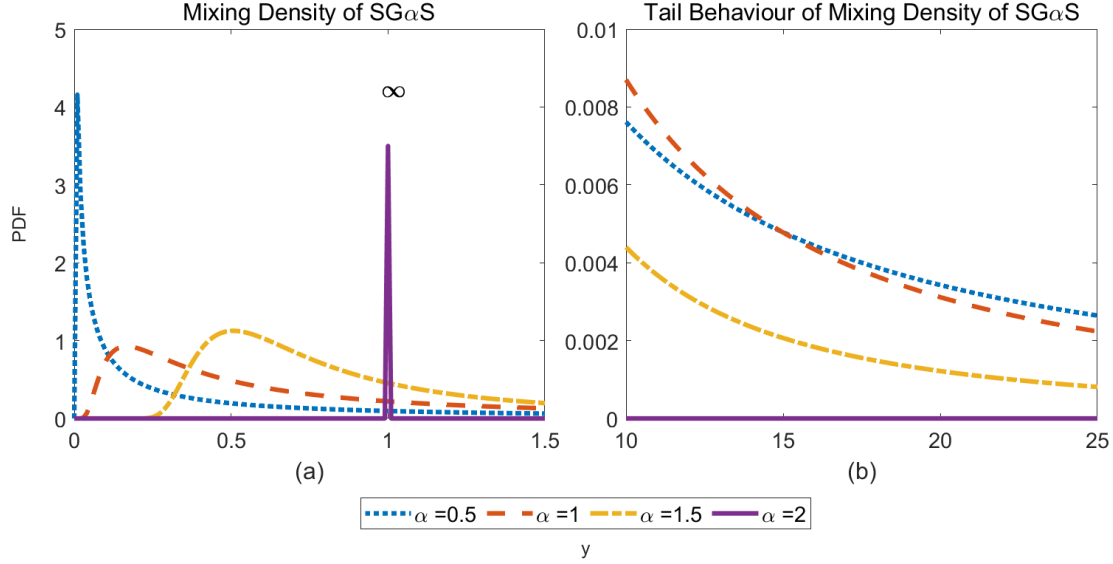


Figure 1: Mixing density and tail behaviour analysis. (a) and (b) plot the pdf and tail behaviours of the mixing density $S(y; \alpha/2, 1, \cos(\pi\alpha/2)^{2/\alpha}, 0)$, respectively.

3. RKF-SG α S

This section begins by introducing the proposed robust Kalman filtering framework based on the SG α S distribution in section 3.1. Subsequently, Section 3.2 details the four MMSE estimators used to estimate the scale function. Moreover, Section 3.3 describes the convergence test strategy employed for the RKF-SG α S.

3.1. Proposed Robust Kalman Filtering Framework

Assume the zero-mean signal and measurement noises are subject to Gaussian and SG α S distributions, respectively. Then, both are modelled as

$$p(\mathbf{w}_{k-1}) = \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1}) \quad (4)$$

$$p(\mathbf{v}_k) = \int_0^{+\infty} \mathcal{N}(\mathbf{v}_k; \mathbf{0}, \lambda_k \mathbf{R}_k) S(\lambda_k) d\lambda_k \quad (5)$$

where $\mathbf{Q}_{k-1}$ is the covariance matrix of the process noise at time $k-1$ whilst $\mathbf{R}_k$ and λ_k refer to the scale matrix and mixing parameter of the measurement noise at time k , respectively. Also, the mixing density $S(\lambda_k; \alpha/2, 1, \cos(\pi\alpha/2)^{2/\alpha}, 0)$ is simplified as $S(\lambda_k)$. Assume the posterior distribution of $\mathbf{x}_{k-1}$ at time $k-1$ is

approximated as a Gaussian distribution, i.e. $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) = \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1})$, where $\hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{P}_{k-1|k-1}$ are the mean vector and covariance matrix. The forecast pdf $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ and likelihood pdf $p(\mathbf{z}_k|\mathbf{x}_k)$ can be expressed as

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k|k-1}) \quad (6)$$

$$p(\mathbf{z}_k|\mathbf{x}_k) = \int_0^{+\infty} \mathcal{N}(\mathbf{z}_k; \mathbf{H}_k \mathbf{x}_k, \lambda_k \mathbf{R}_k) S(\lambda_k) d\lambda_k \quad (7)$$

where

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_{k-1} \quad (8)$$

denotes the forecast error covariance matrix at time $k-1$. According to (7), $p(\mathbf{z}_k|\mathbf{x}_k)$ follows a hierarchical Gaussian form of

$$p(\mathbf{z}_k|\mathbf{x}_k, \lambda_k, \mathbf{R}_k) = \mathcal{N}(\mathbf{z}_k; \mathbf{H}_k \mathbf{x}_k, \lambda_k \mathbf{R}_k) \quad (9)$$

where $\lambda_k \sim S(\lambda_k)$. Assume the true value of $\mathbf{R}_k$ is unknown, and we model its uncertainty using the inverse-Wishart distribution, i.e. $p(\mathbf{R}_k) = \text{IW}(\mathbf{R}_k; u_k, \mathbf{U}_k)$, where u_k and $\mathbf{U}_k$ are the dof parameter and scale matrix, respectively. According to equations (4)-(9), the joint posterior distribution $p(\boldsymbol{\Theta}|\mathbf{z}_{1:k})$ can be expressed as

$$\begin{aligned} p(\boldsymbol{\Theta}|\mathbf{z}_{1:k}) &\propto p(\mathbf{z}_k|\boldsymbol{\Theta})p(\boldsymbol{\Theta}|\mathbf{z}_{1:k-1}) \\ &= \mathcal{N}(\mathbf{z}_k; \mathbf{H}_k \mathbf{x}_k, \lambda_k \mathbf{R}_k) \times \mathcal{N}(\mathbf{x}_k; \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k|k-1}) \times S(\lambda_k) \times \text{IW}(\mathbf{R}_k; u_k, \mathbf{U}_k) \end{aligned} \quad (10)$$

where $\boldsymbol{\Theta}_k = \{\mathbf{x}_k, \lambda_k, \mathbf{R}_k\}$. As there is no analytic expression for $p(\boldsymbol{\Theta}_k|\mathbf{z}_{1:k})$, the Variational Bayesian approach [36] based on the EM algorithm is employed to approximate the posterior density from

$$p(\boldsymbol{\Theta}_k|\mathbf{z}_{1:k}) \approx q(\mathbf{x}_k)q(\lambda_k)q(\mathbf{R}_k). \quad (11)$$

Then, the density $q(\cdot)$ can be calculated by

$$\log q(\phi) = \mathbb{E}_{\boldsymbol{\Theta}_k^{(-\phi)}} [\log p(\boldsymbol{\Theta}_k, \mathbf{z}_{1:k})] + c_\phi \quad (12)$$

where $\phi \subset \boldsymbol{\Theta}_k$, $\phi \cup \boldsymbol{\Theta}_k^{(-\phi)} = \boldsymbol{\Theta}_k$ and c_ϕ is a constant number related to ϕ .

Next, the fixed-point iteration is employed to solve equation (12). At the $t+1$ -th iteration, $q^{(t+1)}(\phi)$ is approximated by employing $q^{(t)}(\phi)$ to calculate the expectation in (12). Proposition 1-3 present the updated formulation, but their proofs are not provided in this work as similar proofs have been made in previous work [20, 35, 37]. For completeness of the presentation however, these propositions are provided below.

Proposition 1 ([35] Proposition 2). *Let $\phi = \mathbf{x}_k$, then $q^{(t)}(\mathbf{x}_k)$ can be updated as the Gaussian distribution, i.e.,*

$$q^{(t+1)}(\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^{(t+1)}, \mathbf{P}_{k|k}^{(t+1)}),$$

155 where

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k-1} &= \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} \\
\mathbf{K}_k^{(t+1)} &= \mathbf{P}_{k|k-1} \mathbf{H}_k^T \left(\mathbf{H}_k \mathbf{P}_{k|k-1}^T \mathbf{H}_k^T + \tilde{\mathbf{R}}_k^{(t)} \right)^{-1} \\
\hat{\mathbf{x}}_{k|k}^{(t+1)} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k^{(t+1)} (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) \\
\mathbf{P}_{k|k}^{(t+1)} &= \left(\mathbf{I}_n - \mathbf{K}_k^{(t+1)} \mathbf{H}_k \right) \mathbf{P}_{k|k-1}.
\end{aligned} \tag{13}$$

156 where $\tilde{\mathbf{R}}_k^{(t)}$ is the modified measurement noise covariance matrix and can be written as

$$\tilde{\mathbf{R}}_k^{(t)} = \frac{[\mathbf{E}^{(t)}(\mathbf{R}_k^{-1})]^{-1}}{\mathbf{E}^{(t)}(\kappa(\lambda_k))} = \frac{[\mathbf{E}^{(t)}(\mathbf{R}_k^{-1})]^{-1}}{\mathbf{E}^{(t)}(\lambda_k^{-1})}. \tag{14}$$

157 **Proposition 2** ([35] Proposition 4). Let $\phi = \lambda_k$, $\log q^{(t+1)}(\lambda_k)$ can be expressed as,

$$\log q^{(t+1)}(\lambda_k) = -\frac{\eta}{2} \lambda_k^{-1} - \frac{m}{2} \log \lambda_k + \log S(\lambda_k) + C', \tag{15}$$

158 where C' is a constant number and

$$\begin{aligned}
\eta &= \text{tr} \left\{ \mathbf{B}_k^{(t+1)} \mathbf{E}^{(t)}(\mathbf{R}_k^{-1}) \right\} \\
\mathbf{B}_k^{(t+1)} &= \left(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(t+1)} \right) \left(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(t+1)} \right)^T + \mathbf{H}_k \mathbf{P}_{k|k}^{(t+1)} \mathbf{H}_k^T
\end{aligned} \tag{16}$$

159 **Proposition 3** ([37] Proposition 4). Let $\phi = \mathbf{R}_k$, $q^{(t+1)}(\mathbf{R}_k)$ is written as the inverse-Wishart distribution,

160 i.e.,

$$q^{(t+1)}(\mathbf{R}_k) = \text{IW} \left(\mathbf{R}_k; u_k^{(t+1)}, \mathbf{U}_k^{(t+1)} \right),$$

161 where

$$\begin{aligned}
u_k^{(t+1)} &= u_k + 1, \mathbf{U}_k^{(t+1)} = \mathbf{U}_k + \mathbf{D}_k^{(t+1)} \\
\mathbf{D}_k^{(t+1)} &= \mathbf{E}^{(t+1)}(\lambda_k^{-1}) \times \left[\mathbf{b}_k^{(t+1)} \left(\mathbf{b}_k^{(t+1)} \right)^T + \mathbf{H}_k \hat{\mathbf{P}}_{k|k}^{(t+1)} \mathbf{H}_k^T \right] \\
\mathbf{b}_k^{(t+1)} &= \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(t+1)}
\end{aligned} \tag{17}$$

162 Thence, we have

$$\mathbf{E}^{(t+1)}(\mathbf{R}_k^{-1}) = \left(u_k^{(t+1)} - m - 1 \right) \left(\mathbf{U}_k^{(t+1)} \right)^{-1} \tag{18}$$

163 **Remark 1.** In equations (14) and (17), the expectation of the scale function $\mathbf{E}(\kappa(\lambda_k)) = \mathbf{E}(\lambda_k^{-1})$ is required.

164 While Proposition 2 provides the pdf of λ_k , the calculation of the expectation is not straightforward due to

165 the lack of a closed-form expression of $S(\lambda_k)$ in equation (15).

166 3.2. MMSE Estimators of the Scale Function

167 To estimate the expectation of the scale function, this section introduces four estimators of $\mathbf{E}^{(t+1)}(\lambda_k^{-1})$,

168 including IS and GLQ-based estimators, along with their corresponding hybrid estimators with the incorpo-

169 ration of the GS based estimator. For simplicity, we replace λ_k with y and omit the iteration number $t + 1$

170 in (15), and then we have

$$E^{(t+1)}(\lambda_k^{-1}) = E(y^{-1}) \quad (19)$$

$$\log q(y) = -\frac{\eta}{2}y^{-1} - \frac{m}{2}\log y + \log S(y) + C' \quad (20)$$

$$q(y) \propto q'(y) = y^{-\frac{m}{2}} \exp\left(-\frac{\eta}{2y}\right) S(y), \quad (21)$$

171 where $q'(y)$ is a function proportional to the density $q(y)$. Then,

$$E(y^{-1}) = \frac{\int_0^{+\infty} y^{-1} q'(y) dy}{\int_0^{+\infty} q'(y) dy}. \quad (22)$$

172 3.2.1. Importance Sampling Based Estimator

173 Here we introduce the estimation method of $E(y^{-1})$ based on the IS technique. The IS-based estimator is
 174 elaborated in Theorem 1. The proof is not provided in this work, and Theorem 1 can be easily demonstrated
 175 based on the IS formulation as shown in [31].

176 **Theorem 1.** *Let y_i ($i = 1, \dots, N$) be samples from the target distribution $S(y)$. Then $E(y^{-1})$ can be*
 177 *estimated by*

$$E(y^{-1}) = \sum_{i=1}^N w_i y_i^{-1} \quad (23)$$

178 where

$$w_i = \frac{y_i^{-\frac{m}{2}} \exp\left(-\frac{\eta}{2y_i}\right)}{\sum_{i'=1}^N y_{i'}^{-\frac{m}{2}} \exp\left(-\frac{\eta}{2y_{i'}}\right)}. \quad (24)$$

179 **Remark 2.** For simplicity, the prior distribution $S(y)$ is selected as the importance proposal of the IS.
 180 Although the approximation of the $S(y)$ value in (21) is computationally expensive [38], it can be efficiently
 181 sampled [39]. Hence, based on the IS algorithm, we first obtain the samples of $S(y)$, and then the numerator
 182 and denominator integrals in (22) can be approximated.

183 3.2.2. Gauss–Laguerre Quadrature Based Estimator

184 An alternative estimator of $E(y^{-1})$, this time based on the GLQ, is discussed in this section. The details
 185 of this estimator are provided in Theorem 2. The proof of Theorem 2 is not provided in this work, but can
 186 be easily obtained based on the GLQ formulation as shown in [32].

187 **Theorem 2.** *Based on the GLQ method, $E(y^{-1})$ can be approximated by*

$$E(y^{-1}) = \frac{\sum_{l=1}^L w_l x_l f(x_l)}{\frac{\eta}{2} \sum_{l=1}^L w_l f(x_l)}. \quad (25)$$

188 where x_l , $l = 1, \dots, L$, is the l -th root of Laguerre polynomial $\mathcal{L}_L(x)$ and $f(x_l) = x_l^{\frac{m}{2}-2} S(\frac{\eta}{2x_l})$. Also, the
 189 weight w_l can be written as

$$w_l = \frac{x_l}{(L+1)^2 [\mathcal{L}_{L+1}(x_l)]^2}. \quad (26)$$

Remark 3. Considering the exponential function in $q'(y)$, the GLQ is a suitable method to approximate the integrals in (22). The GLQ-based estimator in Theorem 2 requires calculating $f(x_l)$ in which the approximation of $S(\frac{\eta}{2x_l})$ cannot be avoided. However, compared to the IS method, the GLQ requires fewer samples [32]. Thus, despite the large computational load of approximating $S(y)$ values [38], the GLQ is still applicable to the estimation of $E(y^{-1})$.

3.2.3. Gamma Series Based Estimator

Both the IS and GLQ-based estimators suffer from low efficiency due to their requirements of numerous particles and the approximation of $S(y)$, respectively. In contrast, a more efficient estimator based on the GS is explained in detail in this section. We first show how the density $q'(y)$ can be represented by a series based on the inverse Gamma density (cf. Lemma 1). Subsequently, a corresponding estimator of $E(y^{-1})$ is introduced as a result of Theorem 3.

Lemma 1. For the $SG\alpha S$ distribution with shape parameter α , let $\alpha_1 = \frac{\alpha}{2}$, then $q'(y)$ can be represented as an Inverse-Gamma-pdf-based series, i.e.,

$$q'(y) = \sum_{\xi=1}^{+\infty} c_{\xi} \frac{\Gamma(a_{\xi})}{b^{a_{\xi}}} \text{IG}(y; a_{\xi}, b) \quad (27)$$

where

$$a_{\xi} = \xi\alpha_1 + \frac{m}{2}, \quad b = \frac{\eta}{2}, \quad c_{\xi} = (-1)^{\xi+1} \frac{\Gamma(\xi\alpha_1 + 1)}{\pi\xi!} \sin(\xi\alpha_1\pi). \quad (28)$$

PROOF. For the $SG\alpha S$ with shape parameter α , let $\alpha_1 = \frac{\alpha}{2}$, then $S(y)$ can be represented as a series [40]

$$S(y) = -\frac{1}{\pi y} \sum_{\xi=1}^{+\infty} \frac{\Gamma(\xi\alpha_1 + 1)}{\xi!} (-y^{-\alpha_1})^{\xi} \sin(\xi\alpha_1\pi)$$

Then, replacing $S(y)$ with the expression in (21),

$$q'(y) = \sum_{\xi=1}^{+\infty} c_{\xi} y^{-a_{\xi}-1} \exp\left(-\frac{b}{y}\right) \quad (29)$$

where a_{ξ} , b and c_{ξ} are defined in (28) and (28). As $y^{-a_{\xi}-1} \exp\left(-\frac{b}{y}\right)$ is proportional to an IG distribution, equation (29) can be reformulated as in (27), which finalises the proof.

Theorem 3. For the $SG\alpha S$ distribution with shape parameter α , $E(y^{-1})$ can be represented as the ratio of two Gamma function-based series sums as

$$E(y^{-1}) = \frac{\sum_{\xi=1}^{+\infty} c_{\xi} \frac{\Gamma(a_{\xi}+1)}{b^{(a_{\xi}+1)}}}{\sum_{\xi=1}^{+\infty} c_{\xi} \frac{\Gamma(a_{\xi})}{b^{a_{\xi}}}} \quad (30)$$

iff both the series sums are convergent.

PROOF. Based on equation (27), we can represent the two integrals in (22) with two series respectively as

$$\int_0^{+\infty} q'(y) dy = \sum_{\xi=1}^{+\infty} c_{\xi} \frac{\Gamma(a_{\xi})}{b^{a_{\xi}}} \int_0^{+\infty} \text{IG}(y; a_{\xi}, b) dy \quad (31)$$

212

$$\int_0^{+\infty} y^{-1} q'(y) dy = \sum_{\xi=1}^{+\infty} c_{\xi} \frac{\Gamma(a_{\xi})}{b^{a_{\xi}}} \int_0^{+\infty} \frac{\text{IG}(y; a_{\xi}, b)}{y} dy \quad (32)$$

213 assuming the above series are convergent. As $\int_0^{+\infty} \text{IG}(y; a_{\xi}, b) dy = 1$, from (31), we have

$$\int_0^{+\infty} q'(y) dy = \sum_{\xi=1}^{+\infty} c_{\xi} \frac{\Gamma(a_{\xi})}{b^{a_{\xi}}}. \quad (33)$$

214 Following this, since $y \sim \text{IG}(y; a_{\xi}, b)$, $y^{-1} \sim \text{G}(y; a_{\xi}, b)$, then

$$\int_0^{+\infty} \frac{\text{IG}(y; a_{\xi}, b)}{y} dy = \int_0^{+\infty} y \text{G}(y; a_{\xi}, b) dy = \frac{a_{\xi}}{b}. \quad (34)$$

215 According to (32) and (34), we write

$$\int_0^{+\infty} y^{-1} q'(y) dy = \sum_{\xi=1}^{+\infty} c_{\xi} \frac{\Gamma(a_{\xi} + 1)}{b^{a_{\xi} + 1}} \quad (35)$$

216 Using the expressions in (33) and (35), we obtain the equation (30) $\square$.

217 **Remark 4.** The numerator and denominator series in equation (30) may not always converge. Hence, the
218 estimator cannot be directly applied to the RKF-SG α S, as it fails in cases when the estimator diverges.

219 3.2.4. Hybrid Estimators

220 Owing to the divergence of the GS-based estimator, two hybrid estimators, GSIS and GSGL, are further
221 developed. When either/both the numerator or/and denominator series in (30) diverges, the GS estimator
222 is replaced with the IS or GLQ-based methods.

223 Considering the convergence ranges of the above-mentioned series are not available, convergence analysis
224 requires testing. For this purpose, let

$$r_{\xi}^{(1)} = c_{\xi} \frac{\Gamma(a_{\xi} + 1)}{b^{(a_{\xi} + 1)}}, \quad r_{\xi}^{(2)} = c_{\xi} \frac{\Gamma(a_{\xi})}{b^{a_{\xi}}} \quad (36)$$

225 where $r_{\xi}^{(1)}$ and $r_{\xi}^{(2)}$ are the ξ -th elements of the numerator and denominator series, respectively. Then, the
226 convergence conditions of these two series can be written as [41]

$$\sum_{\xi=\bar{\xi}-\tau_1}^{\bar{\xi}} \left| \frac{r_{\xi}^{(j)}}{\sum_{\xi'=1}^{\xi} r_{\xi'}^{(j)}} \right| < \varepsilon_1, \quad j = 1, 2 \quad (37)$$

227 where the threshold ε_1 is a small positive number. $\tau_1 \geq 1$ is a small positive integer used to ensure the series
228 sums are stable when $\bar{\xi} - \tau_1 \leq \xi \leq \bar{\xi}$. Also, we set the largest value of $\bar{\xi}$ as Ξ . Following this, we can judge
229 the convergence with the increase of $\bar{\xi}$. If $\bar{\xi} < \Xi$ and the condition in (37) are satisfied, the estimation results
230 are reliable. By contrast, we consider the estimator to diverge when $\bar{\xi} = \Xi$ and the convergence inequalities
231 are violated. Then, either the IS or GLQ estimator is employed instead. The whole estimation method is
232 shown in Algorithm 1.

Algorithm 1: GSIS and GSGL estimation

Input: η, m, α
 $b \leftarrow \frac{\eta}{2};$
for $\bar{\xi} \leftarrow 1$ **to** Ξ **do**
 Calculate $a_{\bar{\xi}}$ by (28);
 Calculate $c_{\bar{\xi}}$ by (28);
 Calculate $r_{\bar{\xi}}^{(1)}, r_{\bar{\xi}}^{(2)}$ by (36);
 if $\bar{\xi} < \Xi$ **and inequalities in (37) are satisfied** **then**
 Estimate $E(y^{-1})$ by (30); Break;
 if $\bar{\xi} = \Xi$ **then**
 Estimate $E(y^{-1})$ by (23) or (25);
return $E(y^{-1})$

3.3. Convergence Test for Fixed-point Iteration Method

Under the RKF-SG α S framework, the convergence of the fixed-point iteration to solve equation (12) needs to be investigated. In previous work [42], $\mathbf{x}_{k|k}^{(t)}$ was solely tested. However, in this paper, we extend this approach by additionally detecting the convergence of the posterior covariance matrix $\mathbf{P}_{k|k}^{(t)}$ and the expectation of the scale function $E^{(t)}(\kappa(\lambda_k))$ to improve the detection reliability. Furthermore, we test the convergence of the latest $\tau_2 > 1$ items, as opposed to only the t -th item tested in [42]. The convergence conditions are expressed as

$$\sum_{t=\bar{t}-\tau_2}^{\bar{t}} \frac{\text{sum}(\text{abs}(\Phi^{(t)} - \Phi^{(t-1)}))}{\text{sum}(\text{abs}(\Phi^{(t)}))} < \varepsilon_2 \quad (38)$$

where $\Phi \in \{\mathbf{x}_{k|k}, \mathbf{P}_{k|k}, E(\kappa(\lambda_k))\}$ and a small positive number ε_2 is manually set. In addition, the fixed-point iteration has a maximum number of iterations, denoted by M , after which the algorithm will terminate and output the current estimation results. The implementation pseudo-code for the proposed RKF-SG α S is given in Algorithm 2.

4. NUMERICAL SIMULATIONS

4.1. Target Tracking Models and Noises

In this section, we introduce the target tracking model used in the experimental analysis. The target moves uniformly in a straight line, and the state-space model is given by (1)

$$\mathbf{F}_k = \begin{bmatrix} \mathbf{I}_2 & \Delta t \mathbf{I}_2 \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}, \quad \mathbf{H}_k = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \end{bmatrix} \quad (39)$$

with the observation interval of $\Delta t = 1$. Also, the initial target state is chosen from $\mathcal{N}(\mathbf{x}_{0|0}, \mathbf{P}_{0|0})$ randomly, where $\mathbf{x}_{0|0} = [0, 0, 10, 10]^T$, $\mathbf{P}_{0|0} = \text{diag}([25, 25, 2, 2])$. Besides, the process noise $\mathbf{w}_k$ follows the Gaussian

Algorithm 2: One Time Step of the Proposed RKF-SG α S

Input: $\mathbf{z}_k, \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}, \mathbf{F}_k, \mathbf{H}_k, \mathbf{Q}_k, u_k, \mathbf{U}_k, \alpha, m, \Xi, \varepsilon_1, \varepsilon_2, \tau_1, \tau_2, M, N, L$

Calculate $\mathbf{P}_{k|k-1}$ using (8);

Initialisation: $\mathbf{E}^{(0)}(\mathbf{R}_k^{-1}) = (u_k - m - 1)(\mathbf{U}_k)^{-1}$, $\mathbf{E}^{(0)}(\lambda_k^{-1}) = 1$;

for $t \leftarrow 0$ **to** $M - 1$ **do**

 Calculate $\tilde{\mathbf{R}}_k^{(t)}$ using (14);

 Calculate $\hat{\mathbf{x}}_{k|k}^{(t+1)}$ and $\mathbf{P}_{k|k}^{(t+1)}$ using (13);

 Calculate $\mathbf{B}_k^{(t+1)}$ and η using (16);

 Calculate $\mathbf{E}^{(t+1)}(\lambda_k^{-1})$ using one of the scale function estimators in section 3.2;

 Calculate $\mathbf{b}_k^{(t+1)}, \mathbf{D}_k^{(t+1)}, u_k^{(t+1)}$ and $\mathbf{U}_k^{(t+1)}$ using (17);

 Calculate $\mathbf{E}^{(t+1)}(\mathbf{R}_k^{-1})$ using (18);

if *The convergence condition in (38) is satisfied* **then**

 └ Terminate the iteration;

return $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k}^{(M)}, \mathbf{P}_{k|k} = \mathbf{P}_{k|k}^{(M)}$

250 distribution, of which the covariance matrix is given by

$$\mathbf{Q}_k = 0.1 * \overline{\mathbf{Q}}, \quad \overline{\mathbf{Q}} = \begin{bmatrix} \frac{\Delta t^3}{3} \mathbf{I}_2 & \frac{\Delta t^2}{2} \mathbf{I}_2 \\ \frac{\Delta t^2}{2} \mathbf{I}_2 & \Delta t \mathbf{I}_2 \end{bmatrix}. \quad (40)$$

251 Then, three kinds of heavy-tailed measurement noises are selected—the Gaussian Mixture (GM), ST and
 252 SG α S noises. Among them, the GM noise has been employed to test the filtering performance of RKF in
 253 many references [13, 14, 20]. In contrast, we also consider the ST and SG α S noises thanks to their capability
 254 to fit the practical noises. The GM noise [20, 35] is written as

$$\mathbf{GM} = \begin{cases} \mathcal{N}(\mathbf{0}, \overline{\mathbf{R}}) & \text{with probability } 0.9 \\ \mathcal{N}(\mathbf{0}, U\overline{\mathbf{R}}) & \text{with probability } 0.1, \end{cases} \quad (41)$$

255 where $\overline{\mathbf{R}} = 10\mathbf{I}_2$ is the nominal covariance matrix. Also, U is the augment factor which is taken from a vector
 256 $[5, 10, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, 10^8]$. Contrarily, the SG α S and ST noises have the same scale matrix $\overline{\mathbf{R}}$,
 257 and their shape parameters, α and v , take values from $[0.3, 0.5, 0.7, 0.9, 1.1, 1.3, 1.5, 1.7, 1.85]$ and $[0.3, 0.5,$
 258 $0.7, 0.9, 1.2, 1.7, 2.5, 3.5, 6]$, respectively. The whole tracking process lasts for 300 seconds, and 100 Monte
 259 Carlo runs are simulated.

260 4.2. Benchmark Filters

261 The RKF-SG α S is compared with 3 kinds of heavy-tailed-distribution-based RKF in the following sim-
 262 ulations. The first two filters are the RKF based on the variance Gamma distribution (RKF-VG) [20] and
 263 the RSTKF [17]. Lastly, our improved version of RKF-SL is also used as a comparison method. Besides, the
 264 standard KF with true noise covariance matrices (KFTNCM) [35] is employed for reference.

For the RKF-SG α S, the performance of the four variants is evaluated, including RKF-SG α S-IS, RKF-SG α S-GLQ, RKF-SG α S-GSIS, RKF-SG α S-GSGL. Then, for the GS estimator, $\Xi = 30$, $\varepsilon_1 = 10^{-2}$ and $\tau_1 = 4$. Moreover, we set $M = 50$, $\varepsilon_2 = 10^{-2}$ and $\tau_2 = 4$ for all the RKF-based filters.

4.3. Estimation of RKF Parameters

In this section, we employ the EM or MLE method to initialise the shape parameters of the GSM distributions, u_k , and $\mathbf{U}_k$. Compared with manually setting these RKF parameters as in the previous work [20], this can alleviate the influence of the model errors and provide a fairer comparison among these RKF. Considering the two-dimensional isotropic measurement noise, we simply assume that the covariance matrix of the GSM distributions is a scalar matrix. Following this, the parameter estimation of the two-dimensional GSM distributions can be addressed as in the univariate case. Specifically, we can first estimate the parameters of a univariate GSM distribution based on the one-dimensional squeezed measurement noise samples, and then the estimated shape and variance parameters equal the shape parameter and the diagonal elements of the covariance matrix of the two-dimensional GSM distribution, respectively.

For the SG α S distribution, the EM algorithm is employed [30]. Specifically, the EM iteration number is 200, and the estimates of the last 15 iterations are averaged for the final estimate. Besides, to sample the posterior distribution of the Weibull random variable under this EM parameter estimation framework, a rejection sampling method is used and the sample number is 2000. In contrast to the case of SG α S mentioned above, the MLE algorithm is utilised for the parameter estimation of the VG, SL and ST distributions. As we only consider the zero-mean symmetric noise, the values of both the mean and skewness vectors for all the distributions are assumed to be known. The measurement noise sample size is 1000, and 1000 Monte Carlo runs are taken. The shape parameter (α or ν) is estimated by averaging the results. Lastly, 1000 scale matrix estimation results are modelled by the inverse-Wishart distribution, and the MLE method is used to estimate its parameters u_k and $\mathbf{U}_k$.

4.4. Performance Evaluation of RKF-SG α S Variants

Based on the parameter estimation results, the properties of our proposed RKF-SG α S variants are investigated in this section. The particle number N and the root number L are taken from the vectors [10, 40, 100, 400, 1000, 4000, 10000] and [1, 2, 4, 6, 10, 20, 30], respectively. Also, the measurement noises come from SG α S distributions, and the parameter selection is explained in section 4.1.

Figure 2 shows the experimental results of RKF-SG α S-IS. (1) From Figure 2-(a), larger particle numbers allow for the broader effective ranges of α and more stable filtering results. The heavier-tailed distributions contain larger probability spaces, and then more samples are needed for the IS approximation. (2) The position estimation RMSEs in Figure 2-(b) are not influenced by N within the effective ranges of α . This suggests that the larger estimation errors of $E(\kappa(\lambda_k))$ caused by small N have little impact on the estimation of the position marginal distribution. (3) From Figure 2-(c), the average iteration number is inversely proportional to N , as the imprecise estimate of $E(\kappa(\lambda_k))$ causes the EM estimator to fluctuate around

the local optimum and cannot satisfy the convergence inequalities in (38). Thus, the convergence process is delayed. (4) From Figure 2-(d), the execution time is proportional to the sample size. Although large N can reduce the iteration number, the filter with small N is still more efficient because of the shorter implementation time on the IS.

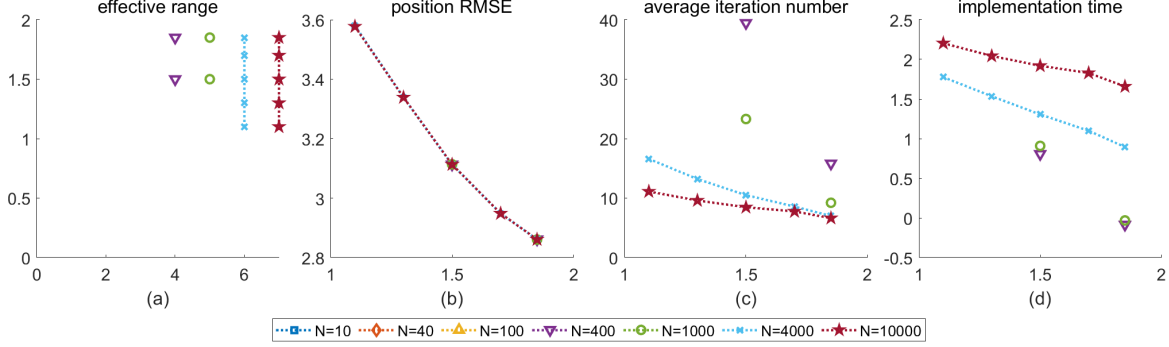


Figure 2: The target tracking results of the RKF-SG α S-IS under different particle numbers N . For (a), the x-axis is the particle number index and the y-axis represents the shape parameter value. In comparison, for (b)-(d), the x-axis is the shape parameter value, and the y-axis represents the estimation RMSE, iteration number and $\log(\text{time})$, respectively. (a) depicts the shape parameter ranges, where the filter can track the target steadily. By contrast, (b) show the position estimation RMSEs. Also, (c) describes the average fixed-point iteration numbers. Besides, the average implementation time of one Monte Carlo run is shown in (d).

The tracking performance of RKF-SG α S-GLQ is plotted in Figure 3. (1) The effective α range shown in Figure 3-(a) is related to the root number. For $L = 1$, the filter fails on the whole α range, because the GLQ cannot estimate the integral. By contrast, for $L = 2$, the effective range is large. However, when α is large, the target is lost. For the light-tailed case, the probability space of $S(y)$ concentrates on a small range, and then the GLQ produces rough estimates. In comparison, the filter is stable on the whole tested ranges when $L \geq 4$. (2) The position estimation (Figure 3-(b)) precision cannot be improved by larger L although the GLQ can produce more precise estimates of $E(\kappa(\lambda_k))$. (3) From Figure 3-(c), the convergence of the fixed-point iteration method is influenced by both α and L . When α is large, smaller L requires more iterations due to the fluctuation around the local optimum. By contrast, when $\alpha < 1.3$, different values of L need similar iteration numbers since the GLQ estimation with small L is still precise in the heavy-tailed cases. (4) From Figure 3-(d), the relation between the execution time and L is complicated. When α is small, the filter with small L is more efficient. Nevertheless, for large α , its efficiency degrades caused by the increased number of iterations.

The simulation results of the RKF-SG α S-GSGL are plotted in Figure 4. (1) For this hybrid-estimator-based filter, the ratio of the second method, GLQ, is added to Figure 4-(e). For small α , the ratio is 0. However, with the increase of α , the ratio of the GLQ gradually rises towards 1. This suggests that the GS estimator is more stable when the noise is heavy-tailed. (2) The effective range is related to L . For $L = 1$, although the RKF-SG α S-GLQ always fails in Figure 3-(a), the effective range of this hybrid filter is $\alpha \leq 0.7$,

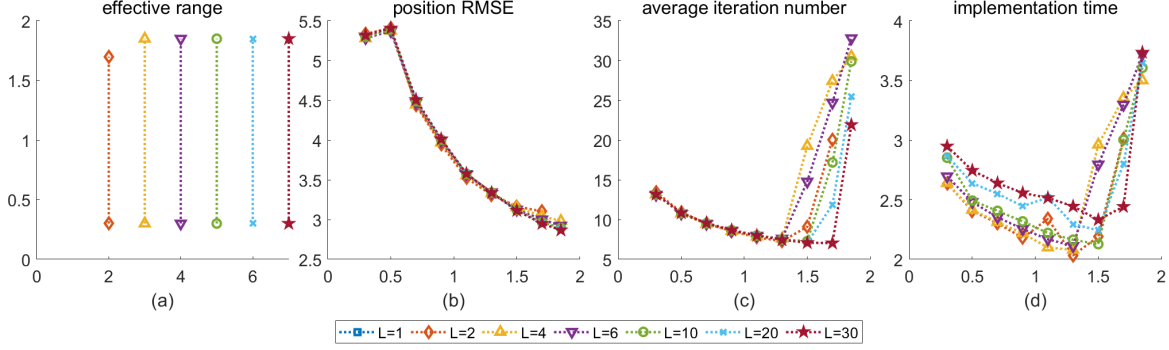


Figure 3: The target tracking results of the RKF-SG α S-GLQ under different root numbers L . For (a), the x-axis is the root number index and the y-axis represents the shape parameter value. In comparison, for (b)-(d), the x-axis is the shape parameter value, and the y-axis represents the estimation RMSE, iteration number and $\log(\text{time})$, respectively. (a) depicts the shape parameter ranges, where the filter can track the target steadily. By contrast, (b) show the position estimation RMSEs. Also, (c) describes the average fixed-point iteration numbers. Besides, the average implementation time of one Monte Carlo run is shown in (d).

where the GS estimator always converges. By contrast, for $L > 1$, the effective range is the same as that of the RKF-SG α S-GLQ. (3) The position estimation results are always precise within the effective ranges. (4) The iteration number and the execution-time lines overlap for small α , where the hybrid filter mainly relies on the GS estimator. However, from Figure 3-(d) and 4-(d), the RKF-SG α S-GSGL is more efficient than the RKF-SG α S-GLQ in heavy-tailed cases, which is caused by the smaller computation loads of the GS estimator. By contrast, with the increase of α , their results gradually become similar due to the rising ratio of the GLQ estimator.

The performance of RKF-SG α S-GSIS is evaluated in Figure 5. (1) The ratio of the second method, the IS method, is similar to that of the RKF-SG α S-GSGL in Figure 4-(e). (2) The hybrid filter is stable on the whole α range, even when N is small. The GS estimator is unstable for large α , whilst the IS method with few particles can produce precise results in this case (cf. Figure 2). So, the complementary advantages of the GS and IS strategies extend the effective range of the filter. (3) The hybrid filter can work in the α ranges where both the IS and GS-based filters fail. For example, from Figures 2-(a) and 4-(a), the IS estimator with $N = 100$ fails in the whole tested range, and GS estimator is effective for only $\alpha \leq 0.7$. However, Figure 5-(a) shows the hybrid filter with $N = 100$ can also work when $\alpha > 0.7$. This suggests that although the GS estimator cannot converge at every iteration when $\alpha > 0.7$, the IS method is effective when the GS diverges. (4) All the position estimation RMSEs are close. (5) Based on Figures 2-(c-d) and 5-(c-d), the hybrid filter is more efficient than the RKF-SG α S-IS when α is small because of the high ratio of the GS estimator as shown in Figure 5-(e). By contrast, their performances are similar when α is large.

Figure 6 provides a filtering performance comparison among these four variants of RKF-SG α S, where $N = 100, L = 2$. (1) From Figure 6-(a), the GSIS-based filter presents the broadest effective filtering range owing to the complementary advantages of the IS and IG estimators. (2) Within the effective scales, all the

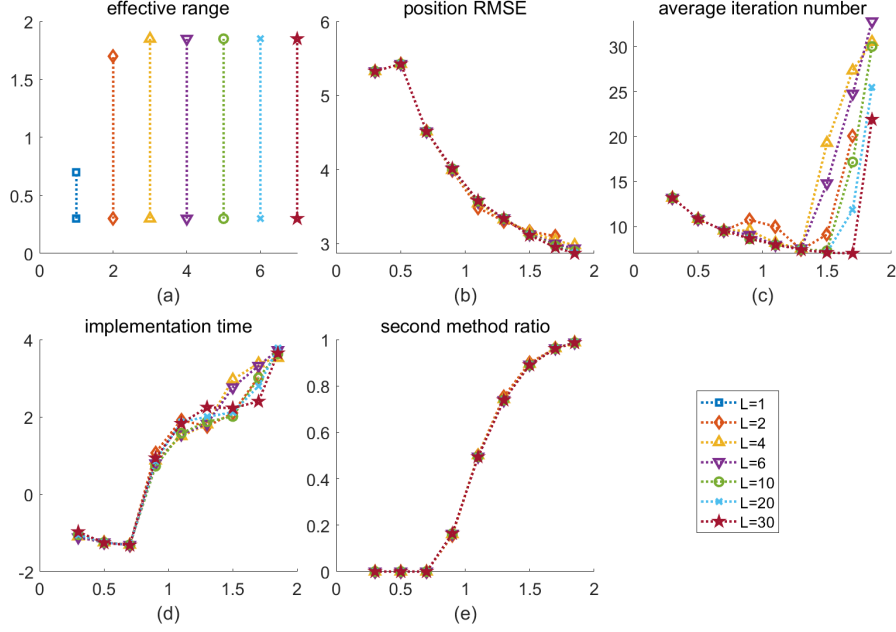


Figure 4: The target tracking results of the RKF-SG α S-GSGL under different root numbers L . For (a), the x-axis is the root number index and the y-axis represents the shape parameter value. In comparison, for (b)-(e), the x-axis is the shape parameter value, and the y-axis represents the estimation RMSE, iteration number, log(time) and ratio value, respectively. (a) depicts the shape parameter ranges, where the filter can track the target. By contrast, (b) shows the position estimation RMSEs. Also, (c) describes the average fixed-point iteration numbers. Then, the average implementation time of one Monte Carlo run is shown in (e). Besides, (f) paints the ratio of the iterations where the GS method is replaced with the GLQ estimator.

filters produce similar filtering results in Figure 6-(b). (3) Figure 6-(c) shows the RKF-SG α S-GSIS requires the smallest computational loads over the whole tested scale though its implementation time is similar to that of the GSGL-based filter for small α .

In summary, for the effective range, the performance of RKF-SG α S-IS is worse than the other filters, and numerous particles are needed for small α . Conversely, the position estimation results of all four proposed strategies are similar in the effective α ranges. However, benefiting from the complementary advantages of the GS and IS estimators, the RKF-SG α S-GSIS with few particles performs steadily and then becomes the most efficient.

4.5. RKF-SG α S vs. Benchmark Filters

Following the thorough analysis of the RKF-SG α S in the previous sections, the comparison between the RKF-SG α S-GSIS and the benchmark filters, listed in Section 4.2, is presented in this section. For efficiency, we set $N = 100$. Also, three kinds of measurement noises are employed, and their parameter selection is made as explained in section 4.1.

In Figure 7, the effective range results are shown. The RKF-VG fails when the shape value is small under the SG α S noises. On the contrary, all other filters behave stably over the whole tested range.

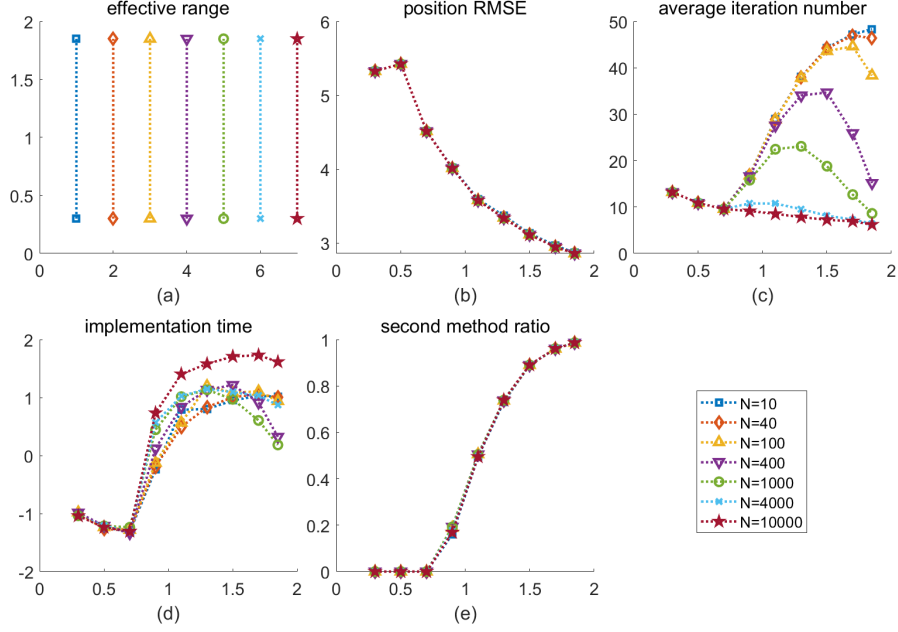


Figure 5: The target tracking results of the RKF-SG α S-GSIS under different particle numbers N . For (a), the x-axis is the root number index and the y-axis represents the shape parameter value. In comparison, for (b)-(e), the x-axis is the shape parameter value, and the y-axis represents the estimation RMSE, iteration number, $\log(\text{time})$ and ratio value, respectively. (a) depicts the shape parameter ranges, where the filter can track the target. By contrast, (b) shows the position estimation RMSEs. Also, (c) describes the average fixed-point iteration numbers. Then, the average implementation time of one Monte Carlo run is shown in (d). Besides, (e) depicts the ratio of the iterations where the GS method is replaced with the IS estimator.

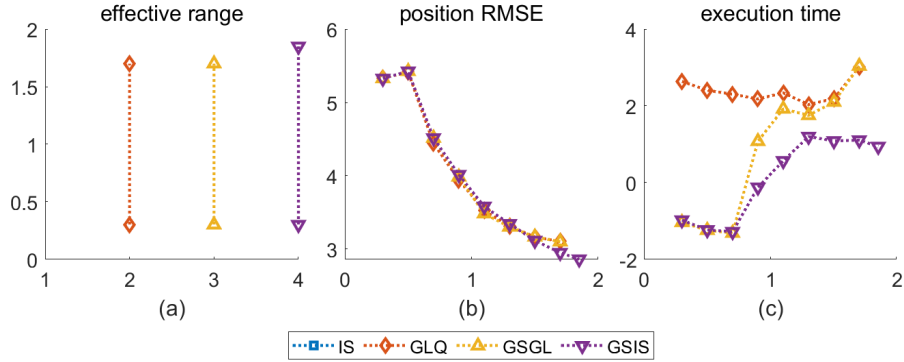


Figure 6: The target tracking results of the four RKF-SG α S variants. For (a), the x-axis is the variant index and the y-axis represents the shape parameter value. In comparison, for (b)-(c), the x-axis is the shape parameter value, and the y-axis represents the estimation RMSE and $\log(\text{time})$, respectively. (a) depicts the shape parameter ranges, where the filter can track the target. By contrast, (b) shows the position estimation RMSEs, and the average implementation time of one Monte Carlo run is shown in (c).

359 The position estimation results are depicted in Figure 8. Under light-tailed noise, all the filters produce
 360 similar results. However, under the heavy-tailed noises, their performances vary. Compared to RKF-VG, the
 361 RKF-SG α S always performs better under different measurement noises. Also, the RKF-SL obtains a more

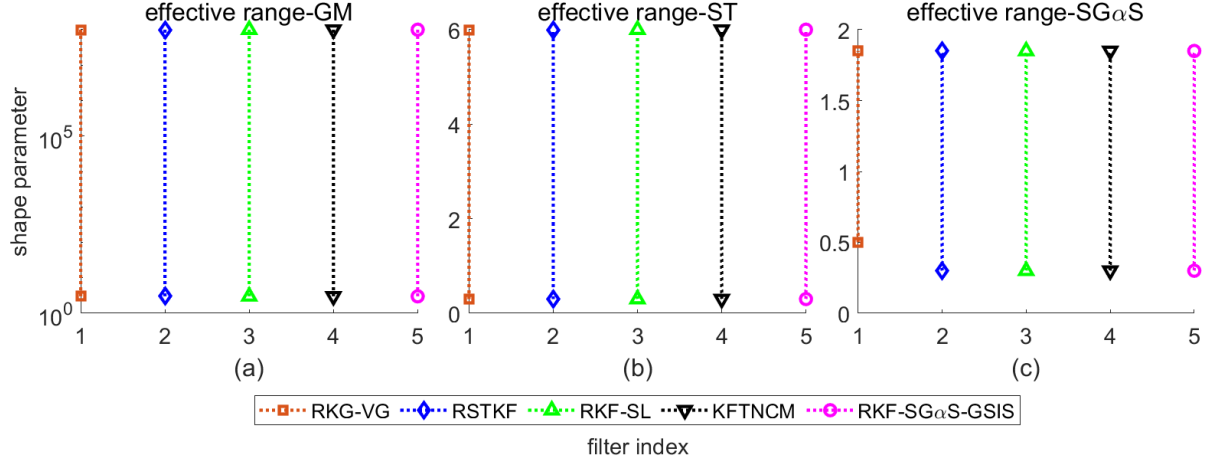


Figure 7: The effective shape parameter ranges of different filters under the three kinds of measurement noises. The titles of the subplots are the noise distributions, the x-axis is the filter index, and the y-axis represents the shape parameter value. Also, for simplicity, the parameter U of GM distribution is defined as the shape parameter.

precise estimation than RKF-SGαS under the GM noise, but worse results under SGαS noises. Further, the RSTKF and RKF-SGαS achieve comparable estimates under GM and ST noises. Nonetheless, the RKF-SGαS becomes superior under the heavy-tailed SGαS noise.

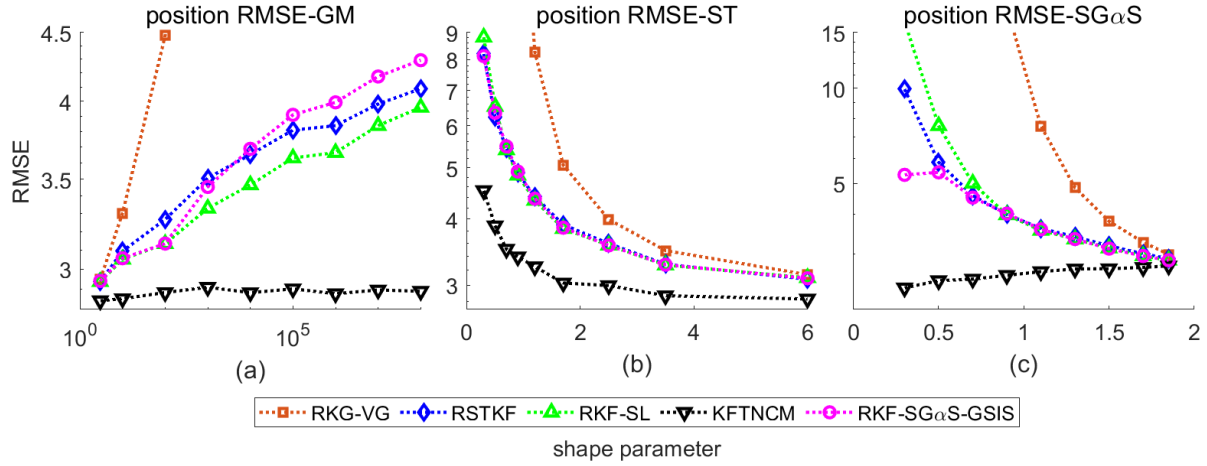


Figure 8: The position estimation RMSEs of different filters under the three kinds of measurement noises. The titles of the subplots follow the format: position-noise distributions, the x-axis is the shape parameter value, and the y-axis represents position estimation RMSEs. Also, for simplicity, the U of GM distribution is defined as the shape parameter.

Figure 9 plots the execution time of different filters. Under light-tailed noise, the computational cost of the RKF-SGαS is higher. Nevertheless, their performance becomes similar in the heavy-tailed cases because the high ratio of the GS estimator increases the filtering efficiency.

Figure 10 presents the filtering error dynamics of the position and velocity estimation, for SGαS noise with characteristic exponent $\alpha = 0.3$. The RKF-VG fails in this scenario, so no corresponding filtering performance

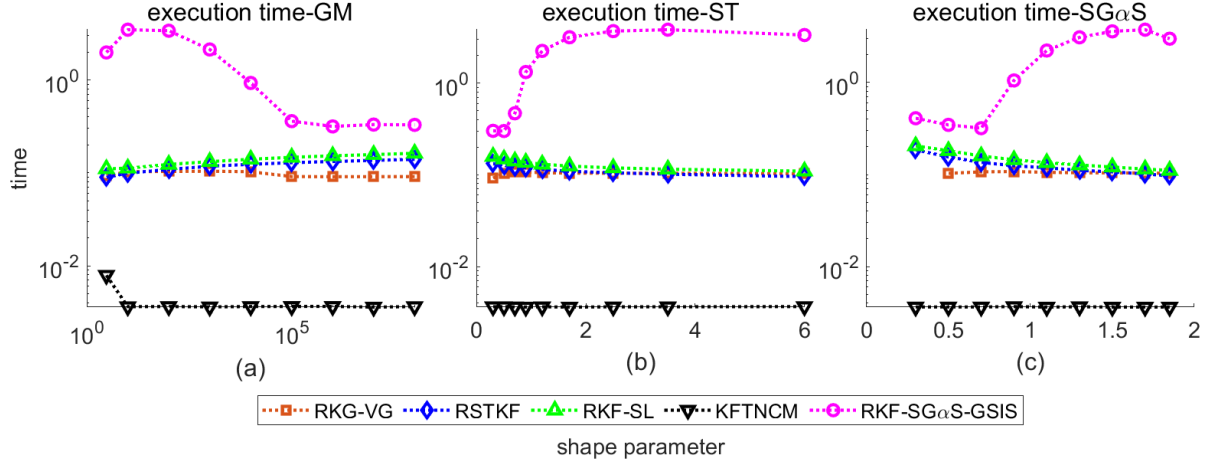


Figure 9: The average execution time of one Monte Carlo run of different filters under the three kinds of measurement noises. The titles of the subplots are the noise distributions, the x-axis is the shape parameter value, and the y-axis represents the implementation time. Also, for simplicity, the U of GM distribution is defined as the shape parameter.

is plotted. Also, all the other filters produce stable estimation results over time, and the RKF-SG α S achieves the best results.

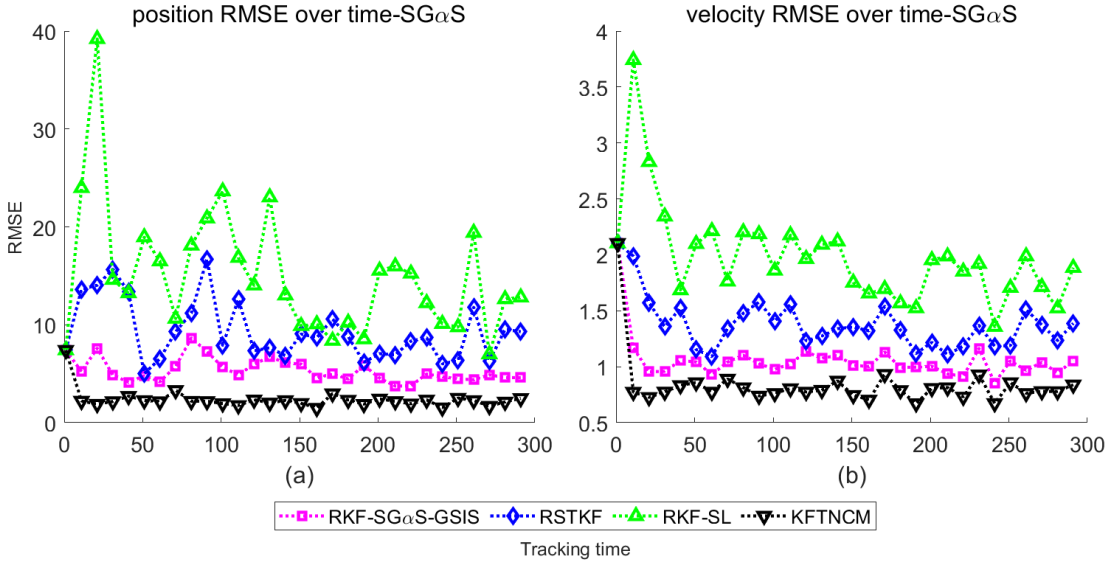


Figure 10: The filtering error dynamics over time. For (a) and (b), the x-axis is the tracking times, and the y-axis represents the position/velocity estimation RMSEs.

5. Conclusion

In this paper, considering linear models with the SG α S noise, a robust Kalman filter framework based on the SG α S distribution is proposed. First, we present four RKF-SG α S variants based on the different MMSE

estimators of the scale function. Then their properties are investigated. The experimental results indicate that the RKF-SG α S-GSIS obtains the most efficient performance thanks to the complementary advantages of the GS and IS estimators. Second, the RKF-SG α S is compared with several heavy-tailed-distribution-based RKF. From the simulations, the RKF-SG α S produces similar estimates under the GM and ST noises but obtains the best performance under the heavy-tailed SG α S noise. Considering the importance of the SG α S noise, this superiority over benchmark filters is significant. Besides, although our RKF-SG α S has higher computational complexity than the benchmark filters, their execution times become closer in particularly heavy-tailed cases.

In the future, we will consider applying the proposed framework to the state-space models with both heavy-tailed signal and measurement noises. In this work, only the measurement noise is considered heavy-tailed since the parameter estimation of the one-step prediction noise is more difficult compared to that of the measurement noise. However, it is straightforward to apply the RKF-SG α S to dynamic models with heavy-tailed process noise.

Code Availability

All the code needed to reproduce the results presented in this work are available at: <https://github.com/PengchengH/Robust-Kalman-Filters-Based-on-the-Sub-Gaussian-alpha-stable-Distribution>.

Acknowledgements

This work was supported by the Chinese Scholarship Council (CSC).

Appendix A. Improved RKF-SL

Here, we give the MMSE estimator of the scale function under the RKF-SL framework. As a special case of the GSM distribution, it has been applied to the RKF in the previous work [20, 43], whilst only the MAP estimate of the scale function was employed. To enhance the precision, we derived the MMSE estimate and Proposition 4 is presented with details.

Proposition 4. *For a slash distribution with dof parameter v , $E(\kappa(y))$ can be estimated by,*

$$E(\kappa(y)) = E(y) = \frac{\gamma(a+1, b)}{b\gamma(a, b)}, \quad (\text{A.1})$$

where

$$a = \frac{m+v}{2} \quad \text{and} \quad b = \frac{\eta}{2}. \quad (\text{A.2})$$

PROOF. For a slash distribution, the mixing density follows $\text{Be}(y; \frac{v}{2}, 1)$ and $0 < y < 1, v > 0$. Then according to (20), we have

$$q'(y) \propto y^{\frac{m+v-2}{2}} \exp\left(-\frac{\eta}{2}y\right).$$

Thus $q'(y)$ is proportional to a Gamma pdf when $0 < y < 1$. As $\kappa(y) = y$, we have

$$E(\kappa(y)) = E(y) = \frac{\int_0^1 yq'(y)dy}{\int_0^1 q'(y)dy} \quad (\text{A.3})$$

Calculate a and b according to (A.2), then

$$\begin{aligned} \int_0^1 yq'(y)dy &= \frac{\gamma(a+1, b)}{b^{a+1}} \\ \int_0^1 q'(y)dy &= \frac{\gamma(a, b)}{b^a}. \end{aligned} \quad (\text{A.4})$$

The expressions in (A.3) and (A.4), give the expression in (A.1) which finalises the proof $\square$.

References

- [1] R. E. Kalman, A new approach to linear filtering and prediction problems, Transactions of the ASME–Journal of Basic Engineering 82 (1) (1960) 35–45. doi:10.1115/1.3662552.
- [2] R. S. Baheti, Efficient approximation of Kalman filter for target tracking, IEEE Transactions on Aerospace and Electronic Systems AES-22 (1) (1986) 8–14. doi:10.1109/TAES.1986.310687.
- [3] D.-J. Jwo, S.-H. Wang, Adaptive fuzzy strong tracking extended Kalman filtering for GPS navigation, IEEE Sensors Journal 7 (5) (2007) 778–789. doi:10.1109/JSEN.2007.894148.
- [4] R. Hartley, M. Ghaffari, R. M. Eustice, J. W. Grizzle, Contact-aided invariant extended Kalman filtering for robot state estimation, The International Journal of Robotics Research 39 (4) (2020) 402–430. doi:10.1177/0278364919894385.
- [5] Expectations anchoring indexes for brazil using Kalman filter: Exploring signals of inflation anchoring in the long term, International Economics 165 (2020) 72–91.
- [6] C. D. Karlgaard, H. Schaub, Huber-based divided difference filtering, Journal of Guidance, Control, and Dynamics 30 (3) (2007) 885–891. doi:10.2514/1.27968.
- [7] R. Izanloo, S. A. Fakoorian, H. S. Yazdi, D. Simon, Kalman filtering based on the maximum correntropy criterion in the presence of non-Gaussian noise, 2016 Annual Conference on Information Science and Systems (CISS) (2016). doi:10.1109/ciss.2016.7460553.
- [8] Y. Huang, Y. Zhang, Y. Zhao, P. Shi, J. A. Chambers, A novel outlier-robust Kalman filtering framework based on statistical similarity measure, IEEE Transactions on Automatic Control 66 (6) (2021) 2677–2692. doi:10.1109/tac.2020.3011443.
- [9] G. Gao, S. Gao, G. Hong, X. Peng, T. Yu, A robust INS/SRS/CNS integrated navigation system with the chi-square test-based robust Kalman filter, Sensors 20 (20) (2020) 5909.

- [10] Z. Yu, Q. Zhang, K. Yu, N. Zheng, A state-domain robust chi-square test method for GNSS/INS integrated navigation, *Journal of Sensors* 2021 (2021) 1–8.
- [11] H. Wang, H. Li, J. Fang, H. Wang, Robust Gaussian Kalman filter with outlier detection, *IEEE Signal Processing Letters* 25 (8) (2018) 1236–1240.
- [12] H. Li, D. Medina, J. Vilà-Valls, P. Closas, Robust variational-based Kalman filter for outlier rejection with correlated measurements, *IEEE Transactions on Signal Processing* 69 (2020) 357–369.
- [13] G. Agamennoni, J. I. Nieto, E. M. Nebot, Approximate inference in state-space models with heavy-tailed noise, *IEEE Transactions on Signal Processing* 60 (10) (2012) 5024–5037. doi:10.1109/tsp.2012.2208106.
- [14] H. Nurminen, T. Ardeshiri, R. Piche, F. Gustafsson, Skew- t filter and smoother with improved covariance matrix approximation, *IEEE Transactions on Signal Processing* 66 (21) (2018) 5618–5633. doi:10.1109/tsp.2018.2865434.
- [15] M. Roth, E. Ozkan, F. Gustafsson, A Student’s t filter for heavy tailed process and measurement noise, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (2013). doi:10.1109/icassp.2013.6638770.
- [16] Y. Huang, Y. Zhang, N. Li, J. Chambers, Robust Student’s t based nonlinear filter and smoother, *IEEE Transactions on Aerospace and Electronic Systems* 52 (5) (2016) 2586–2596. doi:10.1109/taes.2016.150722.
- [17] Y. Huang, Y. Zhang, N. Li, Z. Wu, J. A. Chambers, A novel robust Student’s t -based Kalman filter, *IEEE Transactions on Aerospace and Electronic Systems* 53 (3) (2017) 1545–1554. doi:10.1109/taes.2017.2651684.
- [18] G. Wang, C. Yang, X. Ma, A novel robust nonlinear Kalman filter based on multivariate Laplace distribution, *IEEE Transactions on Circuits and Systems II: Express Briefs* 68 (7) (2021) 2705–2709. doi:10.1109/tcsii.2021.3052351.
- [19] J. Neri, P. Depalle, R. Badeau, Approximate inference and learning of state space models with Laplace noise, *IEEE Transactions on Signal Processing* 69 (2021) 3176–3189. doi:10.1109/tsp.2021.3075146.
- [20] Y. Huang, Y. Zhang, P. Shi, Z. Wu, J. Qian, J. A. Chambers, Robust Kalman filters based on Gaussian scale mixture distributions with application to target tracking, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 49 (10) (2019) 2082–2096. doi:10.1109/tsmc.2017.2778269.
- [21] G. Samorodnitsky, M. S. Taqqu, Stable non-Gaussian random processes (2017). doi:10.1201/9780203738818.
- [22] J. P. Nolan, Models for heavy tailed data, in: *Univariate stable distributions*, Springer Cham, 2020, pp. 1–333.

- [23] G. Laguna-Sanchez, M. Lopez-Guerrero, On the use of alpha-stable distributions in noise modeling for PLC, *IEEE Transactions on Power Delivery* 30 (4) (2015) 1863–1870. doi:10.1109/tpwrd.2015.2390134.
- [24] S. Luan, M. Zhao, Y. Gao, Z. Zhang, T. Qiu, Generalized covariance for non-Gaussian signal processing and GC-music under alpha-stable distributed noise, *Digital Signal Processing* 110 (2021) 102923. doi:10.1016/j.dsp.2020.102923.
- [25] M. J. Lombardi, S. J. Godsill, Monte Carlo Bayesian filtering and smoothing for TVAR signals in symmetric α -stable noise, in: *2004 12th European Signal Processing Conference, IEEE, 2004*, pp. 865–872.
- [26] P. V. Bidarkota, J. H. McCulloch, Optimal univariate inflation forecasting with symmetric stable shocks, *Journal of Applied Econometrics* 13 (6) (1998) 659–670.
- [27] X. Zhong, A. Premkumar, A. Madhukumar, Particle filtering for acoustic source tracking in impulsive noise with alpha-stable process, *IEEE Sensors Journal* 13 (2) (2012) 589–600.
- [28] N. Jaoua, E. Duflos, P. Vanheeghe, L. Clavier, F. Septier, Impulsive interference mitigation in ad hoc networks based on alpha-stable modeling and particle filtering, in: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2011*, pp. 3548–3551.
- [29] S. Kring, S. T. Rachev, M. Höchstätter, F. J. Fabozzi, Estimation of α -stable sub-Gaussian distributions for asset returns, in: *Risk assessment: Decisions in banking and finance, Springer, 2009*, pp. 111–152.
- [30] M. Teimouri, S. Rezakhah, A. Mohammadpour, Parameter estimation using the EM algorithm for symmetric stable random variables and sub-Gaussian random vectors., *J. Stat. Theory Appl.* 17 (3) (2018) 439–461.
- [31] S. T. Tokdar, R. E. Kass, Importance sampling: A review, *WIREs Computational Statistics* 2 (1) (2009) 54–60. doi:10.1002/wics.56.
- [32] N. Kovvali, *Theory and applications of Gaussian quadrature methods*, Springer Cham, 2011.
- [33] A. Achim, E. Kuruoglu, J. Zerubia, SAR image filtering based on the heavy-tailed Rayleigh model, *IEEE Transactions on Image Processing* 15 (9) (2006) 2686–2693. doi:10.1109/tip.2006.877362.
- [34] A. Achim, A. Basarab, G. Tzagkarakis, P. Tsakalides, D. Kouamé, Reconstruction of ultrasound RF echoes modeled as stable random variables, *IEEE Transactions on Computational Imaging* 1 (2) (2015) 86–95. doi:10.1109/TCI.2015.2463257.
- [35] C. Xue, Y. Huang, F. Zhu, Y. Zhang, J. A. Chambers, An outlier-robust Kalman filter with adaptive selection of elliptically contoured distributions, *IEEE Transactions on Signal Processing* 70 (2022) 994–1009. doi:10.1109/tsp.2022.3151199.

- [36] D. G. Tzikas, A. C. Likas, N. P. Galatsanos, The variational approximation for Bayesian inference, *IEEE Signal Processing Magazine* 25 (6) (2008) 131–146. doi:10.1109/msp.2008.929620.
- [37] H. Zhu, G. Zhang, Y. Li, H. Leung, An adaptive Kalman filter with inaccurate noise covariances in the presence of outliers, *IEEE Transactions on Automatic Control* 67 (1) (2022) 374–381. doi:10.1109/tac.2021.3056343.
- [38] J. P. Nolan, Numerical calculation of stable densities and distribution functions, *Communications in Statistics. Stochastic Models* 13 (4) (1997) 759–774. doi:10.1080/15326349708807450.
- [39] A. Weron, R. Weron, Computer simulation of Lévy α -stable variables and processes, in: *Chaos—The Interplay Between Stochastic and Deterministic Behaviour: Proceedings of the XXXIst Winter School of Theoretical Physics Held in Karpacz, Poland 13–24 February 1995*, Springer, 2005, pp. 379–392.
- [40] P. HOUGAARD, Survival models for heterogeneous populations derived from stable distributions, *Biometrika* 73 (2) (1986) 387–396. doi:10.1093/biomet/73.2.387.
- [41] S. Singh, W. Richards, J. Zinecker, D. Wilton, Accelerating the convergence of series representing the free space periodic green’s function, *IEEE Transactions on Antennas and Propagation* 38 (12) (1990) 1958–1962. doi:10.1109/8.60985.
- [42] Y. Huang, Y. Zhang, Y. Zhao, J. A. Chambers, A novel robust Gaussian–Student’s t mixture distribution based Kalman filter, *IEEE Transactions on Signal Processing* 67 (13) (2019) 3606–3620. doi:10.1109/tsp.2019.2916755.
- [43] M. Bai, Y. Huang, B. Chen, L. Yang, Y. Zhang, A novel mixture distributions-based robust Kalman filter for cooperative localization, *IEEE Sensors Journal* 20 (24) (2020) 14994–15006. doi:10.1109/jsen.2020.3012153.