

## Article

# An Improved Invariant Kalman Filter for Lie Groups Attitude Dynamics with Heavy-Tailed Process Noise

Jiaolong Wang <sup>1,\*</sup> , Chengxi Zhang <sup>2</sup> , Jin Wu <sup>3</sup> and Ming Liu <sup>3</sup>

<sup>1</sup> Key Laboratory of Advanced Process Control for Light Industry, Institute of Automation, Jiangnan University, Wuxi 214111, China

<sup>2</sup> School of Electronics and Information Engineering, Harbin Institute of Technology, Shenzhen 518055, China; zhangchengxi@hit.edu.cn

<sup>3</sup> Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, China; jwucp@connect.ust.hk (J.W.); eelium@ust.hk (M.L.)

\* Correspondence: jlwang@jiangnan.edu.cn

**Abstract:** Attitude estimation is a basic task for most spacecraft missions in aerospace engineering and many Kalman type attitude estimators have been applied to the guidance and navigation of spacecraft systems. By building the attitude dynamics on matrix Lie groups, the invariant Kalman filter (IKF) was developed according to the invariance properties of symmetry groups. However, the Gaussian noise assumption of Kalman theory may be violated when a spacecraft maneuvers severely and the process noise might be heavy-tailed, which is prone to degrade IKF's performance for attitude estimation. To address the attitude estimation problem with heavy-tailed process noise, this paper proposes a hierarchical Gaussian state-space model for invariant Kalman filtering: The probability density function of state prediction is defined based on student's *t* distribution, while the conjugate prior distributions of the scale matrix and degrees of freedom (dofs) parameter are respectively formulated as the inverse Wishart and Gamma distribution. For the constructed hierarchical Gaussian attitude estimation state-space model, the Lie groups rotation matrix of spacecraft attitude is inferred together with the scale matrix and dof parameter using the variational Bayesian iteration. Numerical simulation results illustrate that the proposed approach can significantly improve the filtering robustness of invariant Kalman filter for Lie groups spacecraft attitude estimation problems with heavy-tailed process uncertainty.



**Citation:** Wang, J.; Zhang, C.; Wu, J.; Liu, M. An Improved Invariant Kalman Filter for Lie Groups Attitude Dynamics with Heavy-Tailed Process Noise. *Machines* **2021**, *9*, 182. <https://doi.org/10.3390/machines9090182>

Academic Editors: Peter Odry, Akos Odry and Jan Awrejcewicz

Received: 20 July 2021

Accepted: 23 August 2021

Published: 27 August 2021

**Keywords:** spacecraft attitude estimation; variational Bayesian iteration; invariant Kalman filter; matrix Lie groups; heavy-tailed noise

## 1. Introduction

The navigation and control operations in aerospace engineering usually require adequate accurate knowledge of spacecraft orientation in space, and attitude determination from vector observations is usually employed in astronautically applications [1–4]. The widely used Kalman type filters can infer the state in Euclidean vector space by fusing attitude dynamic and observation sensor according to their probabilities [5–8]. Recently, the building of attitude dynamics on matrix Lie groups has been investigated actively for the navigation and control of spacecraft targets because it can significantly improve the estimation and control performance to make full use of the geometrical properties of Lie groups models [9,10].

For attitude estimation, there have already been significant research interests in the estimation and control of spacecraft targets such as the invariant observers and filters for attitude dynamics on the *special orthogonal group*  $SO(3)$ , or the navigation systems defined on the *special Euclidean group*  $SE(3)$  [11,12]. Accounting for the geometrical invariance of symmetry attitude dynamics, the resulted attitude description with rotation matrix on  $SO(3)$  can avoid the trouble of singularities and unwinding encountered when using unit

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

quaternion [13] and, to some extent, also contribute to better estimation performance for attitude determination algorithms [14]. Exploiting the mapping relation between matrix Lie groups and Lie algebra, invariant Kalman filter (IKF) can convert the attitude dynamics on  $SO(3)$  to Euclidean vector space and mimic classical Kalman filtering steps [15]. IKF is based on the invariance and log-linearity of symmetry dynamics, which then leads to better convergence on a much bigger set of state trajectories [16]. In these available researches, the process noise of attitude dynamical model was assumed to be a concentrated Gaussian distribution [17,18] and the statistics parameter are also assumed to be accurately given in advance as required in classical Kalman theory [19,20]. However, due to the presence of server maneuvers in space missions, there might be some unpredictable disturbances or outliers that would contaminate the Gaussian distribution and violate the required assumptions for Kalman filtering on matrix Lie groups.

For Euclidean space filtering problems, some adaptive methods are commonly used to deal with the trouble of inaccurate/unknown noise parameters [21–23]. Some adaptive techniques have been designed to estimate online unknown/inaccurate noise parameters, mainly including innovation-based adaptive methods and multiple model methods [24–26]; nevertheless, the convergence of innovation-based estimation methods cannot be guaranteed while the parallel-running multiple filter banks might increase the computational complexities. The variational Bayesian (VB) iteration-based adaptive methods have also been employed to jointly determine the system state and inaccurate noise statistics [27–29]. However, these adaptive methods all assume the noise distribution to be Gaussian and cannot be applied to the filtering cases with heavy-tailed noise. For heavy-tailed system noises, the Student's  $t$  distribution assumes the distribution of the latent state and system noises as a joint Student's  $t$ -distribution, and it has been proved to be more suitable to represent the non-Gaussian probability density function (PDF) [30,31].

To deal with the heavy-tailed process noise caused by severe maneuvering of rigid bodies, a robust Student's  $t$ -based hierarchical Gaussian system model for the SINS/GPS integration is presented by modeling the prior probability density function as a Student's  $t$  distribution and the conjugate prior distributions [21]. To improve the filtering performance of Student's  $t$ -based methods for cases with both heavy-tailed process and observation noises, a robust Student's  $t$ -based Kalman filter is proposed employing the variational Bayesian-based iterative approach in [32]; in the work of [33], by applying the Student's  $t$  approximate distributions to posterior filtering and smoothing probability density functions, the robust nonlinear Student's  $t$ -based filter and smoother are also presented as the generalization and extension of the linear Student's  $t$  filtering methods. Note that the above researches provide a sound theoretical foundation for attitude estimation problems with heavy-tailed noises, but they are originally designed in Euclidean space systems for filtering of vector variable states. To the best of our knowledge, for attitude estimation problem in  $SO(3)$ , until now there is still no special study on the robust Student's  $t$ -based adaptive methods for invariant Kalman filtering on matrix Lie groups.

Note that although the available invariant Kalman filter also obeys the basic theory of classical Kalman filter, the above researches about robust Student's  $t$ -based methods cannot be directly applied to matrix Lie groups systems. The uncertainty definition of concentrated Gaussian distribution on matrix Lie groups requires the assumption of noisy errors being small enough [34,35], which is the condition for first order approximation to the Baker–Campbell–Hausdorff (BCH) formula [36,37]. For attitude estimation on  $SO(3)$  in aerospace engineering, the presence of heavy-tailed noises and disturbances imposes a critical constraint on the application of robust Student's  $t$ -based methods to improve the performance of invariant Kalman filter.

According to the above discussions, it is rather meaningful to investigate the robust Student's  $t$ -based invariant Kalman filtering method especially for the matrix Lie groups attitude estimation problems in the presence of heavy-tailed outliers in system process noise. In this work, for attitude estimation problem with heavy-tailed process noise, this paper proposes a hierarchical Gaussian state-space model for invariant Kalman filtering, and the

Lie groups rotation matrix of spacecraft attitude is inferred together with the scale matrix and dof parameter using the variational Bayesian iteration. Numerical simulation results illustrate that the proposed approach can significantly improve the filtering robustness of invariant Kalman filter for Lie groups spacecraft attitude estimation problems with heavy-tailed process uncertainty.

## 2. Primaries and Problem Definition

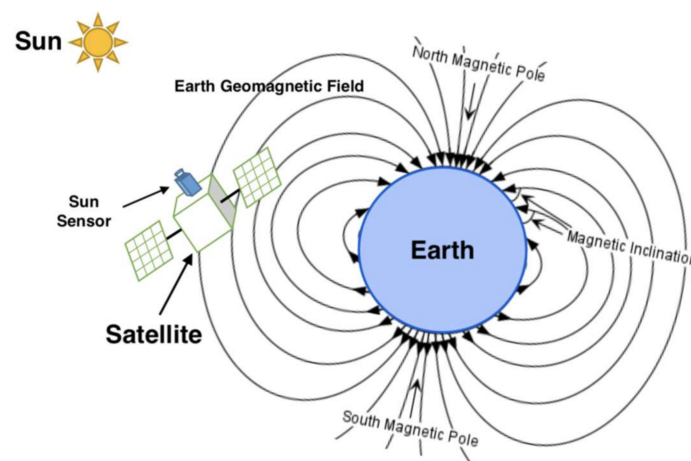
### 2.1. The Attitude Estimation System on the Special Orthogonal Group $SO(3)$

Attitude determination from two vector observations is an important task in most aerospace engineering. The attitude dynamic on *special orthogonal group*  $SO(3) \subset \mathbb{R}^{3 \times 3}$  with the associated observation vectors can be formulated as the following system model [1,14]

$$R_k = \text{Exp}_G(w_k) R_{k-1} \Omega_{k-1} \quad (1)$$

$$Y_k = \begin{pmatrix} y'_k \\ y''_k \end{pmatrix} = \begin{pmatrix} R_k^T b' + v'_k \\ R_k^T b'' + v''_k \end{pmatrix}, \quad (2)$$

where the *special orthogonal group*  $SO(3) := \{RR^T = I_{3 \times 3}, \det(R) = 1\}$  is the group of  $3 \times 3$  rotation matrices and  $R_k \in SO(3)$  is the mapping rotation from the body frame of the spacecraft to the earth-fixed frame at time instant  $k$ ;  $\Omega_k \in SO(3)$  denotes the control rotation;  $Y_k \in \mathbb{R}^6$  is a composition of the vector observations  $y'_k, y''_k \in \mathbb{R}^3$  with their respective observation vector  $b', b'' \in \mathbb{R}^3$ ;  $v'_k \sim \mathcal{N}^c(v'_k; 0_{3 \times 1}, \Sigma_{v'})$ , and  $v''_k \sim \mathcal{N}^c(v''_k; 0_{3 \times 1}, \Sigma_{v''})$  are Gaussian white and mutually independent observation noises with  $\mathcal{N}^c$  denoting the Gaussian distribution. Note that the vector observation based attitude determination is a common application scenario because the Sun sensor, magnetometer, horizon sensor, and the gravitational accelerometer are usually equipped by most spacecrafts in astronautical engineering [1]. A graphical presentation of target application scenario is given in Figure 1.



**Figure 1.** The vector observations-based attitude determination in astronautical application (permission granted from [1]).

Note that the additive forms of vector noises in Euclidean space are not supported in Lie Groups space. The vector process noise  $w_k \in \mathbb{R}^3$  is injected through Lie exponential operator  $\text{Exp}_G(\cdot)$  and, for any  $w = [w_1 \ w_2 \ w_3]^T \in \mathbb{R}^3$ , the projection between the vector noise  $w = [w_1 \ w_2 \ w_3]^T \in \mathbb{R}^3$  and the associated groups element  $R \in SO(3)$  are

$$\text{Exp}_G(\cdot) : \mathbb{R}^3 \rightarrow SO(3), \text{Exp}_G^{-1}(\cdot) : SO(3) \rightarrow \mathbb{R}^3 \quad (3)$$

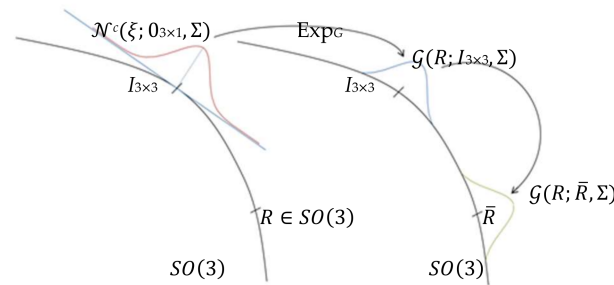
$$\text{Exp}_G(w) := \exp(w_{\times}) = \sum_{m=0}^{\infty} \frac{w_{\times}^m}{m!} = I_{3 \times 3} + \frac{\sin\|w\|}{\|w\|} w_{\times} + 2 \frac{\sin(\|w\|/2)^2}{\|w\|^2} w_{\times}^2, \quad (4)$$

$$w_{\times} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}_{\times} = \begin{pmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{pmatrix} \in \mathbb{R}^{3 \times 3}, \quad (5)$$

where  $\exp(\cdot)$  is the matrix exponential operator and  $w_{\times}$  is the skew-symmetric matrix. The concept of *concentrated Gaussian distribution* (CGD) was employed to define the probability distribution for random variables on matrix Lie groups [19,20,37]. For attitude estimation, in *concentrated Gaussian distribution* the groups variable  $R \in SO(3)$  with mean  $\bar{R} \in SO(3)$  and covariance  $\Sigma \in \mathbb{R}^{3 \times 3}$  is defined as

$$R \sim \mathcal{G}(R; \bar{R}, \Sigma) \in SO(3) \text{ if } \xi \triangleq \text{Exp}_G^{-1}(\bar{R}R^{-1}) \sim \mathcal{N}^c(\xi; 0_{3 \times 1}, \Sigma) \in \mathbb{R}^3 \quad (6)$$

where  $\mathcal{G}(\cdot, \cdot)$  denotes the *concentrated Gaussian distribution* (CGD). Note that, in the definition of CGD, the eigenvalues of  $\Sigma$  are assumed to be small enough, i.e., almost all the mass of the distribution should be concentrated in a small neighborhood around the mean of the rotation variable [20,37]. A graphical illustration of the relation between the *concentrated Gaussian distribution* on matrix Lie groups  $SO(3)$  and the classical Gaussian distribution relation is presented in Figure 2.



**Figure 2.** The mappings between the probability definition for Lie group rotation matrix  $R \in SO(3)$  and the classical Gaussian distribution vector  $w \in \mathbb{R}^3$ .

## 2.2. Model Projection Based on the Invariance Property of Attitude Estimation System

For attitude estimation on  $SO(3)$ , the usual Kalman filtering methods in Euclidean vector space could not be applied directly. However, by properly defining the system variable, the Lie groups dynamic model (1) can be projected into the usual Euclidean space based on the invariance property. Let  $\hat{R}_{k-1}$  be an estimate for  $R_{k-1}$  at instant  $k-1$  and  $\hat{R}_k = \hat{R}_{k-1}\Omega_{k-1}$  be the prediction for  $R_k$ . Define the *invariant error*  $\xi_k = \text{Exp}_G^{-1}(\hat{R}_k R_k^{-1})$ , then the evolution model of *invariant error*  $\xi_k$  obeys

$$\begin{aligned} \xi_k &= \text{Exp}_G^{-1}(\hat{R}_k R_k^{-1}) = \text{Exp}_G^{-1}(\hat{R}_{k-1}\Omega_{k-1}(\text{Exp}_G(w_k)R_{k-1}\Omega_{k-1})^{-1}) \\ &= \text{Exp}_G^{-1}(\hat{R}_{k-1}R_{k-1}^{-1}\text{Exp}_G(-w_k)) = \text{Exp}_G^{-1}(\text{Exp}_G(\xi_{k-1})\text{Exp}_G(-w_k)) \\ &= \text{BCH}(\xi_{k-1}, -w_k) = \xi_{k-1} - w_k + O(|\xi_{k-1}|^2, |w_k|^2, |\xi_{k-1}||w_k|), \end{aligned} \quad (7)$$

where the  $\text{BCH}(\cdot, \cdot)$  of the last step is a first order form of the Baker–Campbell–Hausdorff (BCH) formula [34–36]. Note that if the norm of the error term  $\xi_k$  and  $w_k$  are small enough so that the second order term in (7) could be negligible, then the matrix Lie groups attitude dynamic (1) for rotation matrix  $R_k$  could be approximately converted into the following linear state space model in Euclidean space for the *invariant error*  $\xi_k$

$$\xi_k \doteq \xi_{k-1} - w_k. \quad (8)$$

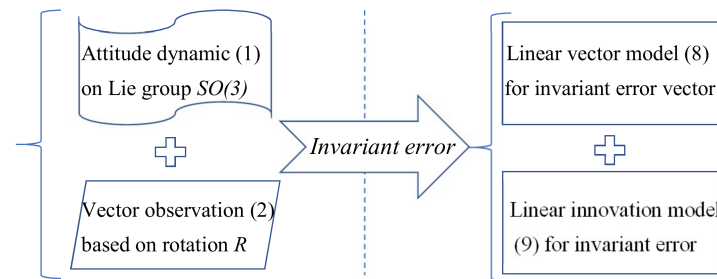
As for observation model (2), the observation sequence  $Y_k$  can also be reformulated into a function of the *invariant error*  $\xi_k$ , i.e., the new innovation sequence  $z_k \in \mathbb{R}^6$  [14]



$$z_k = \hat{R}_k Y_k - \begin{pmatrix} b' \\ b'' \end{pmatrix} = \begin{pmatrix} \hat{R}_k y'_k - b' \\ \hat{R}_k y''_k - b'' \end{pmatrix} = \hat{R}_k R_k^T \begin{pmatrix} b' \\ b'' \end{pmatrix} - \begin{pmatrix} b' \\ b'' \end{pmatrix} + \hat{R}_k \begin{pmatrix} v'_k \\ v''_k \end{pmatrix} := H_\zeta \zeta_k + V_k, \quad (9)$$

where  $H = \begin{pmatrix} b'_\times \\ b''_\times \end{pmatrix} \in \mathbb{R}^{6 \times 3}$  and  $V_k = \hat{R}_{k|k-1} \begin{pmatrix} v'_k \\ v''_k \end{pmatrix} \in \mathbb{R}^6$  are the projected observation matrix and noise, respectively; note that, since the observation noises  $v'_k, v''_k$  are isotropic and mutually independent noises and with  $\hat{R}_k \hat{R}_k^T = I_{3 \times 3}$ , we have  $\Sigma_V = \text{Cov}(V_k V_k^T) = \text{diag}(\Sigma_{v'}, \Sigma_{v''})$ .

Obviously, based on the definition of the *invariant error*  $\zeta_k$ , the Lie groups dynamics (1) and observation model (2) could be converted as the Euclidean space model (8) and (9) for the *invariant error*  $\zeta_k$  (as in Figure 3). Note that both the true rotation state  $R_{k-1}$  and the control rotation  $\Omega_{k-1}$  do not appear in the converted model (8) and (9); therefore, the evolving motion of  $\zeta_k$  is actually independent of the true trajectory of rotation  $R_{k-1}$  and control rotation  $\Omega_{k-1}$ , which is called the *invariance* property of symmetry model [15,16].



**Figure 3.** Projecting the Lie groups model into vector space based on the invariant error.

### 2.3. The Invariant Kalman Filter for Attitude Estimation

As the consequence of invariance property, for the attitude estimation system, the probability distribution of rotation state  $R_k$  based on the observation  $Y_k$  is equivalent to that of the *invariant error*  $\zeta_k$  in  $\mathbb{R}^3$ , constituting the basis of invariant Kalman filter. Then, if the process noise obeys classical Gaussian distribution, i.e.,  $w_k \sim \mathcal{N}^c(w_k; 0_{3 \times 1}, \Sigma_w) \in \mathbb{R}^3$ , the filtering steps for attitude estimation system on Lie group  $SO(3)$  can be implemented by mimicking that of the classical Kalman filter for *invariant error*  $\zeta_k$  in Euclidean space.

For the attitude dynamic (1) and associated observation (2), invariant Kalman filter (IKF) aims to recursively predict and correct the estimate for the rotation state. Given the initial estimate  $\hat{R}_{0|0} = \text{Exp}_G(\hat{\zeta}_{0|0}) R_0$  for true  $R_0$  with error  $\hat{\zeta}_{0|0} \sim \mathcal{N}^c(\hat{\zeta}_{0|0}; 0_{3 \times 1}, \Sigma_{0|0})$ , as the prediction for prior error estimate  $\hat{\zeta}_{k|k-1}$  based on (8), the prior error rotation estimate  $\hat{R}_{k|k-1}$  based on the dynamic model (1) can be implemented as follows

$$\hat{R}_{k|k-1} = \hat{R}_{k-1|k-1} \Omega_{k-1}, \quad (10)$$

where  $\hat{R}_{k|k-1}$  and  $\hat{R}_{k-1|k-1}$  are the prior and posterior error estimate for true rotation state  $R_k$ , respectively;  $\hat{\zeta}_{k|k-1}$  and  $\hat{\zeta}_{k-1|k-1}$  are associated with the prior and posterior error estimate for true error  $\zeta_k$ . The covariance propagation of rotation group  $R_k$  is same as that of the classical Kalman filter for invariant error  $\zeta_k$  based on the evolution model (8),

$$\Sigma_{k|k-1} = \Sigma_{k-1|k-1} + \Sigma_w, \quad (11)$$

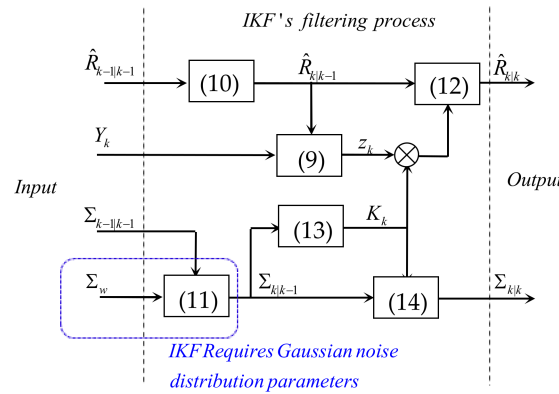
where  $\Sigma_{k-1|k-1}$  stands for the covariance of the posterior estimate  $\hat{\zeta}_{k|k-1}$  and  $\hat{R}_{k|k-1}$ . Therefore, based on the converted innovation (9), the correction steps of IKF for the rotation state  $\hat{R}_{k|k-1}$  can be implemented by updating the estimate  $\hat{R}_{k|k-1}$  with innovation sequence  $z_k$

$$\hat{R}_{k|k} = \text{Exp}_G(K_k z_k) \hat{R}_{k|k-1}, \quad (12)$$

$$K_k = \Sigma_{k|k-1} H^T / (H \Sigma_{k|k-1} H^T + \Sigma_V), \quad (13)$$

$$\Sigma_{k|k} = \Sigma_{k|k-1} - K_k H \Sigma_{k|k-1}, \quad (14)$$

where  $K_k$  is the Kalman gain matrix, and the posterior estimate  $\hat{R}_{k|k}$  is then updated with Lie exponential operation of the correction term  $K_k z_k$  as in (12) [16,17]. A graphical presentation of the filtering diagram of invariant Kalman filter is presented in Figure 4.



**Figure 4.** The filtering diagram of the invariant Kalman filter for attitude estimation.

#### 2.4. The Attitude Estimation Problem with the Trouble of Heavy-Tailed Process Noise

Obviously, the invariant Kalman filter (IKF) is a general extension of the Euclidean space Kalman filter to matrix Lie groups systems based on the invariance property of symmetry dynamics [16]. However, for the attitude estimation in aerospace engineering, there is still some difficulty for practical implementation of the IKF-based attitude estimation because the presence of heavy-tailed process noise in attitude dynamic might introduce great challenge and trouble to the theory of IKF:

Firstly, in the filtering step (10)~(14) of IKF, the process noise of attitude was assumed to be a concentrated Gaussian distribution and the statistics parameter is also required to be accurately known (as in Figure 4). However, due to the presence of severely maneuvering operations in space tasks, there might be some unpredictable disturbances or substantial outliers that would contaminate the Gaussian distribution and violate the assumptions for Kalman filtering on matrix Lie groups. Therefore, the theoretical foundation for optimal Kalman filtering is sure to be destroyed, which will degrade the final estimation.

Secondly, for the invariant projection of the Lie groups dynamics (1) into the linear vector form (8) in Section 2.2, the first order approximation to BCH formula is employed based on the assumption that both  $\xi_k$  and  $w_k$  were small enough and the second and higher order residual could be dropped from (7) to (8). However, for aerospace engineering applications, the unpredictable disturbances or substantial outliers caused by severely maneuvering operations might violate this condition and lead to inaccurate approximation to the system model, which is prone to a potential source of error and influences the estimation accuracy of IKF.

For attitude determination in astronautical engineering, it is rather meaningful to further study Lie groups estimation problems with heavy-tailed process noise and investigate new robust adaptive version of invariant Kalman filtering methods to improve the filtering precisions for cases with heavy-tailed process noises.

### 3. Robust Student's t Based Invariant Kalman Filter for Attitude Estimation on SO(3)

#### 3.1. Probability View of Attitude Estimation with Heavy-Tailed Process Noise

For attitude estimation, if the process noise  $w_k$  in dynamic model (1) is heavy-tailed and contaminated by larger outliers, the Gaussian distribution and *concentrated Gaussian distribution* cannot accurately describe the probability distribution of heavy-tailed process

noise. In this work, to account for the larger outliers due to severely maneuvering, the probability of the heavy-tailed noise is described as the student's t distribution

$$p(w_k) = \mathcal{S}_t(w_k; 0_{3 \times 1}, \Sigma_w, \tau_k) = \int_0^{+\infty} \mathcal{N}^c(w_k; 0_{3 \times 1}, \Sigma_w / \gamma_w) \Gamma\left(\gamma_w; \frac{\tau_w}{2}, \frac{\tau_w}{2}\right) d\gamma_w \quad (15)$$

where  $p(w_k)$  denotes the probability density of the process noise  $w_k$ ;  $\mathcal{S}_t(\cdot; \mu, \Sigma, \tau)$  denotes the student's t distribution with mean vector  $\mu$ , scale matrix  $\Sigma$ , and degree of parameter  $\tau$ ;  $\gamma_w$  is an auxiliary random variable in (15); and  $\Gamma(\gamma; \alpha, \beta)$  represents the Gamma probability density function of the scalar  $\gamma$  with  $\alpha$  and  $\beta$  as the shape and rate parameter, i.e.,

$$\Gamma(\gamma; \alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} \gamma^\alpha e^{-\beta\gamma}, & \gamma > 0 \\ 0, & \gamma \leq 0 \end{cases} \quad (16)$$

where  $e$  is the natural constant and  $\Gamma(\cdot)$  represents the Gamma function. Obviously, the probability density function of the student's t distribution can be regarded as the infinite mixture of the classical Gaussian probability density function [21,30].

If given the Gaussian approximation of the posterior probability density function  $p(\xi_{k-1}|z_{1:k-1})$  and concentrated Gaussian approximation of posterior  $p(R_{k-1}|Y_{1:k-1})$  for instant  $k-1$ , the prior probability density function of  $p(\xi_k|z_{1:k-1})$  and  $p(R_k|Y_{1:k-1})$  for time instant  $k$  can be obtained according to the Chapman–Kolmogorov equation [38,39],

$$p(\xi_k|z_{1:k-1}) = \int p_w(\xi_k - \xi_{k-1}) p(\xi_{k-1}|z_{1:k-1}) d\xi_{k-1} = \int p_w(\xi_k - \xi_{k-1}) \mathcal{N}^c(\xi_{k-1}; \hat{\xi}_{k-1|k-1}, \Sigma_{k-1|k-1}) d\xi_{k-1} \quad (17)$$

$$p(R_k|Y_{1:k-1}) = \int p_w(\text{Exp}_G^{-1}(R_k \Omega_k^{-1} R_{k-1})) p(R_{k-1}|Y_{1:k-1}) dR_{k-1} = \int p_w(\text{Exp}_G^{-1}(R_k \Omega_k^{-1} R_{k-1})) \mathcal{G}(R_{k-1}; \hat{R}_{k-1|k-1}, \Sigma_{k-1|k-1}) dR_{k-1} \quad (18)$$

where  $p(\xi_{k-1}|z_{1:k-1})$  denotes the posterior probability density function of the  $\xi_{k-1}$  based on the innovation sequence  $z_k$  from time instant 1 to  $k-1$  and  $p(R_{k-1}|Y_{1:k-1})$  denotes the posterior probability density function of rotation  $R_k$  based on observation  $Y_k$  from instant 1 to  $k-1$ . Note that resorting to the equivalence of the  $R_k$  and  $\xi_k$  in probability density function,  $p_w(\xi_k - \xi_{k-1}) = p_w(\text{Exp}_G^{-1}(R_k \Omega_k^{-1} R_{k-1}))$  is the student's t distribution  $p(w_k)$  for system process noise and so we have  $p(R_k|Y_{1:k-1}) = p(\xi_k|z_{1:k-1})$ . Further, it is evident that the prior probability  $p(\xi_k|z_{1:k-1})$  for instant  $k$  is not Gaussian but a student's t distribution with mean vector parameter  $\hat{\xi}_{k|k-1}$ , scale matrix  $\Sigma_{k|k-1}$  and parameter  $\tau_k$ , i.e.,

$$p(R_k|Y_{1:k-1}) = p(\xi_k|z_{1:k-1}) \triangleq \mathcal{S}_t(\xi_k; \hat{\xi}_{k|k-1}, \Sigma_{k|k-1}, \tau_k) = \int_0^{+\infty} \mathcal{N}^c(\xi_k; \hat{\xi}_{k|k-1}, \Sigma_{k|k-1} / \gamma_k) \Gamma\left(\gamma_k; \frac{\tau_k}{2}, \frac{\tau_k}{2}\right) d\gamma_k \quad (19)$$

where  $\hat{\xi}_{k|k-1}$ ,  $\Sigma_{k|k-1}$ , and  $\tau_k$  are the mean vector, scale matrix, and degree of freedom parameter of the prior student's t distribution  $p(R_k|Y_{1:k-1})$ ; and  $\gamma_k$  is the auxiliary random variable. Employing (15)–(19), the prior student's t distribution  $p(R_k|Y_{1:k-1})$  can be reformulated into a hierarchical composition of the Gaussian distributions

$$p(\xi_k|\gamma_k, z_{1:k-1}) = \mathcal{N}^c(\xi_k; \hat{\xi}_{k|k-1}, \Sigma_{k|k-1} / \gamma_k), \quad (20)$$

$$p(\gamma_k) = \Gamma\left(\gamma_k; \frac{\tau_k}{2}, \frac{\tau_k}{2}\right). \quad (21)$$

Attitude estimation aims to propagate the posterior probability density distribution based on the observation  $Y_k$ . Since the observation noise in (2) and its equivalent form  $V_k$  in (9) are zero mean Gaussian white sequence with covariance  $\Sigma_V$ , the likelihood probability density function of the innovation sequence  $z_k$  based on the error  $\xi_k$  is equal to that of observation  $Y_k$  given the rotation  $R_k$ , i.e.,

$$p(Y_k|R_k) = p(z_k|\xi_k) = \mathcal{N}^c(z_k; H_\xi^T \xi_k, \Sigma_V) \quad (22)$$

where  $p(Y_k|R_k)$  is the probability density function of  $Y_k$  given rotation  $R_k$ , and  $p(z_k|\xi_k)$  is the probability density function of  $z_k$  based on error  $\xi_k$ .

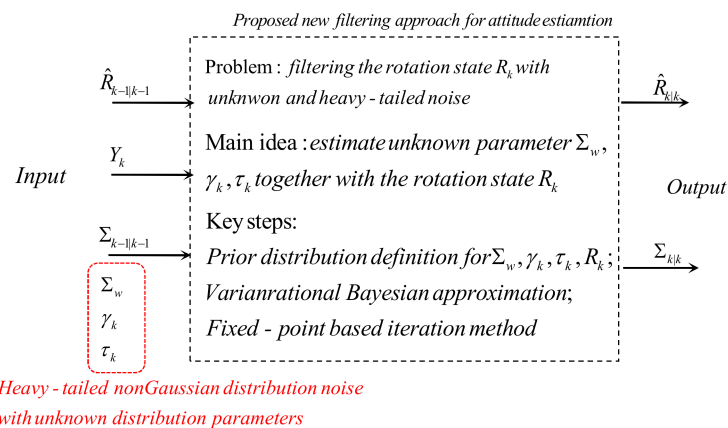
Note that the accurate propagation of the probability density function requires accurate knowledge of scale matrix  $\Sigma_{k|k-1}$  and degree of freedom  $\tau_k$ . However, for real engineering applications in the presence of unpredictable disturbances or substantial outliers, these two distribution parameters are usually not correct or unavailable in advance. Therefore, from the view of probability density, the main troubles of attitude estimation with heavy-tailed process noise are threefold:

- (1) With the student's  $t$  distribution process noise, the probability density function  $p(\xi_k|z_{1:k-1})$  in classical invariant Kalman filter depends on the auxiliary random variable  $\gamma_k$  and becomes the hierarchical form  $p(\xi_k|\gamma_k, z_{1:k-1})$  in (20) and (21).
- (2) For the unpredictable disturbances or outliers induced by severely maneuvering operations, the accurate scale matrix  $\Sigma_{k|k-1}$  and degree of freedom  $\gamma_k$  are usually unavailable but essential to propagate the posterior estimates.

As a consequence, the heavy-tailed attitude estimation problem cannot be solved with the invariant Kalman filter in Section 2 since the distribution parameter  $\Sigma_{k|k-1}$  and  $\tau_k$  should be estimated together with the invariant error  $\xi_k$  and the rotation state  $R_k$ .

### 3.2. Prior Probability Definition for the Parameters of Student's $t$ Distribution

To solve Lie groups filtering problems with Student's  $t$  distribution process noise, this work aims to estimate the rotation group state  $R_k$  together with  $\Sigma_{k|k-1}$  and  $\tau_k$  by variational Bayesian iterations as in Figure 5. The prior distributions of parameter  $\Sigma_{k|k-1}$  and  $\tau_k$  based on all historical observation sequences should be conjugate prior distributions so that their posterior distribution would maintain the same form as their prior distribution.



**Figure 5.** The main idea and key innovation of this work to deal with the heavy-tailed attitude estimation problem on matrix Lie group  $SO(3)$ .

According to (20) and (21), the estimated scale matrix  $\Sigma_{k|k-1}$  is proportional to the covariance matrix of Gaussian distribution. Therefore, the commonly used inverse Wishart distribution and the Gamma distribution of Bayesian statistics are employed as the conjugate prior distributions for parameters  $\Sigma_{k|k-1}$  and  $\gamma_k$ , respectively.

The distribution of scale matrix  $\Sigma_{k|k-1}$  is defined as inverse Wishart distribution [29,39,40] which is often used as conjugate prior for the covariance matrix of the Gaussian variable. For attitude estimation problem, the inverse Wishart function of  $\Sigma_{k|k-1}$  can be written as

$$p(\Sigma_{k|k-1}) = \text{IW}(\Sigma_{k|k-1}; \lambda_k, \Psi_k) = \frac{|\Psi_k|^{\lambda_k/2} |\Sigma_{k|k-1}|^{-(\lambda_k+3+1)/2}}{2^{3\lambda_k/2} (\lambda_k/2)} e^{-0.5 \text{tr}(\Psi_k \Sigma_{k|k-1}^{-1})}, \quad (23)$$

where IW denotes the inverse Wishart distribution;  $\lambda_k$  denotes the degrees of freedom;  $\Psi_k$  is the  $3 \times 3$  inverse scale matrix;  $|\cdot|$  is the determinant operator; and  $\Upsilon_3(\cdot)$  is the 3-variate gamma function [29]. For the scale matrix  $\Sigma_{k|k-1} \sim \text{IW}(\Sigma_{k|k-1}; \lambda_k, \Psi_k)$ , the expectation of the matrix inverse is  $E(\Sigma_{k|k-1}^{-1}) = (\lambda_k - 3 - 1)\Psi_k^{-1}$  (for  $\lambda_k > 3 + 1$ ) and so we have

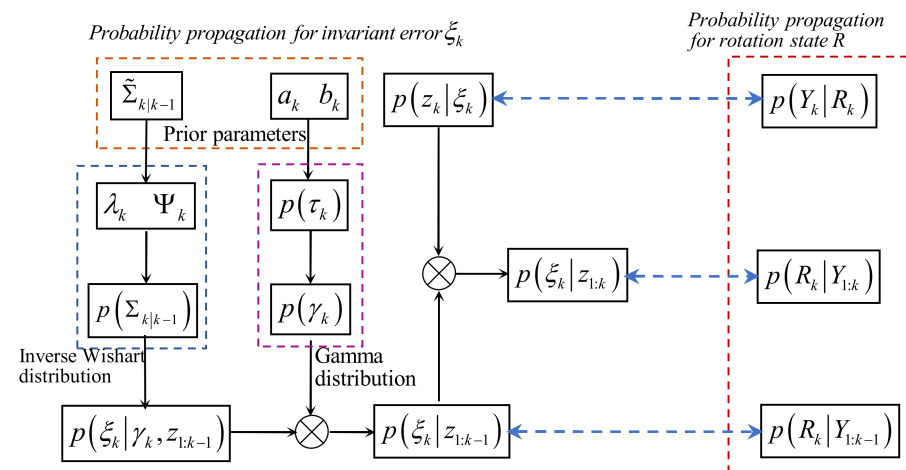
$$E(\Sigma_{k|k-1}) = \frac{\Psi_k}{(\lambda_k - 3 - 1)}, \text{ if } \lambda > 3 + 1. \quad (24)$$

The distribution parameter  $\tau_k$  can be defined as the Gamma distribution [21],

$$p(\tau_k) = \Gamma(\tau_k; a_k, b_k), \quad (25)$$

where  $a_k, b_k$  are respectively the shape and rate parameter of the Gamma distribution  $\tau_k$ , respectively.

In summary, Equations (20)~(25) actually constitute a new student's t-based hierarchical Gaussian state-space model for the heavy-tailed attitude estimation problem, whose diagram is presented in Figure 6. The initialization of the required distribution parameter is based on the nominal parameter  $\Sigma_{k|k-1}$ , which will be discussed later on. Using the above distribution definition, the rotation state  $R_k$  and  $\xi_k$ , the auxiliary random variable  $\gamma_k$ , the scale matrix  $\Sigma_{k|k-1}$ , and the parameter  $\tau_k$  will be jointly estimated based on the newly constructed hierarchical Gaussian state-space model by variational Bayesian iterations, which then leads to the new robust student's t-based invariant Kalman filter for the heavy-tailed attitude estimation problem.



**Figure 6.** Probability view of the proposed filtering method for Student's t based hierarchical Gaussian state-space model.

### 3.3. Variational Beayesian Approximations of Posterior Probability Density Function

To jointly estimate the rotation state  $R_k$ , auxiliary random variable  $\gamma_k$ , scale matrix  $\Sigma_{k|k-1}$ , and parameter  $\tau_k$  requires calculating the posterior probability density function  $p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k | z_{1:k})$ . Since the accurate analytical expression of  $p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k | z_{1:k})$  is not available, this work tries to obtain its factored approximate based on variational Bayesian iterations [38]. i.e.,

$$p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k | z_{1:k}) \approx q(\xi_k)q(\gamma_k)q(\Sigma_{k|k-1})q(\tau_k), \quad (26)$$

where  $q(\xi_k)$ ,  $q(\gamma_k)$ ,  $q(\Sigma_{k|k-1})$ , and  $q(\tau_k)$  represent, respectively, the approximate posterior probability density function of  $\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k$ ;  $q(\xi_k)$ ,  $q(\gamma_k)$ ,  $q(\Sigma_{k|k-1})$  and  $q(\tau_k)$ , which can be determined by minimizing the Kullback–Leibler divergence (KLD) [38,41] between the factored approximate  $q(\xi_k)q(\gamma_k)q(\Sigma_{k|k-1})q(\tau_k)$  and the true  $p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k | z_{1:k})$  [29,38], i.e.,

$$\{q(\xi_k), q(\gamma_k), q(\Sigma_{k|k-1}), q(\tau_k)\} = \arg \min \text{KLD}(q(\xi_k)q(\gamma_k)q(\Sigma_{k|k-1})q(\tau_k) \| p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k | Y_{1:k})), \quad (27)$$



where  $\text{KLD}(q(x)||p(x)) \triangleq \int q(x) \log \frac{q(x)}{p(x)} dx$  denotes the KLD between  $q(x)$  and  $p(x)$ . According to the variational Bayesian iteration, these approximate posterior  $q(\xi_k)$ ,  $q(\gamma_k)$ ,  $q(\Sigma_{k|k-1})$ , and  $q(\tau_k)$  to (27) satisfy the following equations [21,38]

$$\log q(\xi_k) = E_{-\xi_k} [\log p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k, z_{1:k})] + c_{\xi_k}, \quad (28)$$

$$\log q(\gamma_k) = E_{-\gamma_k} [\log p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k, z_{1:k})] + c_{\gamma_k}, \quad (29)$$

$$\log q(\Sigma_{k|k-1}) = E_{-\Sigma_{k|k-1}} [\log p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k, z_{1:k})] + c_{\Sigma_{k|k-1}}, \quad (30)$$

$$\log q(\tau_k) = E_{-\tau_k} [\log p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k, z_{1:k})] + c_{\tau_k}, \quad (31)$$

where  $\log(\cdot)$  denotes the natural logarithm function;  $E_{-\xi_k}[\cdot]$  represents the mathematical expectation operation for all variables except  $\xi_k$ , while  $E_{-\gamma_k}[\cdot]$ ,  $E_{-\Sigma_{k|k-1}}[\cdot]$ , and  $E_{-\tau_k}[\cdot]$  are with similar definitions;  $c_{\xi_k}$  is the constant with respect to  $\xi_k$ , while  $c_{\gamma_k}$ ,  $c_{\Sigma_{k|k-1}}$ , and  $c_{\tau_k}$  are similarly defined. Obviously, Equations (28)~(31) are coupled and so no analytical solutions to the above equations exist. Therefore, only the approximate solutions are available and in this work the fixed-point iterations are employed to determine these parameters [38].

According to the conditional independence properties, the joint probability density function  $p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k, z_{1:k})$  can be factored as follows

$$p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k, z_{1:k}) = p(z_k|\xi_k) p(\xi_k|z_{1:k-1}, \gamma_k) p(\gamma_k) p(\Sigma_{k|k-1}) p(\tau_k) p(z_{1:k-1}) \quad (32)$$

Then, using (20)~(26), we have

$$\begin{aligned} p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k, z_{1:k}) &= \mathcal{N}(z_k; H\xi_k, \Sigma_V) \mathcal{N}^c(\xi_k; \hat{\xi}_{k|k-1}, \Sigma_{k|k-1} / \gamma_k) \\ &\times \Gamma(\gamma_k; \frac{\tau_k}{2}, \frac{\tau_k}{2}) \text{IW}(\Sigma_{k|k-1}; \lambda_k, \Psi_k) \Gamma(\tau_k; a_k, b_k) p(Y_{1:k-1}) \end{aligned} \quad (33)$$

Using (16) and (24),  $\log p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k, z_{1:k})$  can be formulated as

$$\begin{aligned} \log p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k, z_{1:k}) &= -\frac{1}{2}(z_k - H\xi_k)^T \Sigma_V^{-1} (z_k - H\xi_k) - \frac{1}{2}(5 + \lambda_k) \log |\Sigma_{k|k-1}| - \frac{1}{2} \text{tr}(\Psi_k \Sigma_{k|k-1}^{-1}) \\ &- \frac{1}{2} \gamma_k (\xi_k - \hat{\xi}_{k|k-1})^T \Sigma_{k|k-1}^{-1} (\xi_k - \hat{\xi}_{k|k-1}) + \frac{\tau_k + 1}{2} \log \gamma_k - \frac{\tau_k \gamma_k}{2} + \frac{\tau_k}{2} \log \frac{\tau_k}{2} - \log \Gamma(\frac{\tau_k}{2}) \\ &+ (a_k - 1) \log \tau_k - b_k \tau_k + \text{constant}. \end{aligned} \quad (34)$$

### 3.4. Fixed-Point Iteration of the System State and Distribution Parameters

Based on the conditional independence property of the invariant error  $\xi_k$  and parameters  $\gamma_k$ ,  $\Sigma_{k|k-1}$ , and  $\tau_k$ , the Equations (28)~(31) can be rewritten with above Equation (34) and fixed-point methods can be used to obtain the approximate estimates to the invariant error  $\xi_k$  as well as the parameters  $\gamma_k$ ,  $\Sigma_{k|k-1}$ , and  $\tau_k$  [21].

#### 3.4.1. Fixed-Point Iteration of the Invariant Error $\xi_k$

With (34) and the independence property of  $\gamma_k$  and  $\Sigma_{k|k-1}$ , (28) can be written as

$$\begin{aligned} \log q^{(i+1)}(\xi_k) &= E_{-\xi_k}^{(i+1)} [\log p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k, z_{1:k})] + c_{\xi_k} \\ &= -\frac{1}{2}(z_k - H\bar{\xi}_k)^T \Sigma_V^{-1} (z_k - H\bar{\xi}_k) - \frac{1}{2} E_{-\xi_k}^{(i)} \left[ \gamma_k (\xi_k - \hat{\xi}_{k|k-1})^T \Sigma_{k|k-1}^{-1} (\xi_k - \hat{\xi}_{k|k-1}) \right] + c_{\xi_k} \\ &= -\frac{1}{2}(z_k - H\bar{\xi}_k)^T \Sigma_V^{-1} (z_k - H\bar{\xi}_k) - \frac{1}{2} E_{\gamma_k}^{(i)} [\gamma_k] (\xi_k - \hat{\xi}_{k|k-1})^T E_{\Sigma_{k|k-1}}^{(i)} [\Sigma_{k|k-1}^{-1}] (\xi_k - \hat{\xi}_{k|k-1}) + c_{\xi_k} \\ &= -\frac{1}{2}(z_k - H\bar{\xi}_k)^T \Sigma_V^{-1} (z_k - H\bar{\xi}_k) - \frac{1}{2} (\xi_k - \hat{\xi}_{k|k-1})^T (\bar{\Sigma}_{k|k-1}^{(i)})^{-1} (\xi_k - \hat{\xi}_{k|k-1}) + c_{\xi_k} \end{aligned} \quad (35)$$

where  $q^{(i+1)}(\cdot)$  is the iterative approximation to  $q(\cdot)$  at  $i + 1$ th iteration;  $E_{-\xi_k}^{(i)}[\cdot]$  represents the mathematical expectation at  $i$ th iteration for variables except the invariant error  $\xi_k$

while  $E_{\gamma_k}^{(i)}[\cdot]$  and  $E_{\Sigma_{k|k-1}}^{(i)}[\cdot]$  denote the mathematical expectation at  $i$ th iteration for  $\gamma_k$  and  $\Sigma_{k|k-1}$ , respectively;  $\widehat{\Sigma}_{k|k-1}^{(i)}$  is the modified prior error covariance matrix calculated as

$$\widehat{\Sigma}_{k|k-1}^{(i)} = \left\{ E_{\Sigma_{k|k-1}}^{(i)} \left[ \Sigma_{k|k-1}^{-1} \right] \right\}^{-1} / E_{\gamma_k}^{(i)}[\gamma_k]. \quad (36)$$

Then,  $q^{(i+1)}(\xi_k)$  can also be updated as a Gaussian distribution, i.e.,

$$q^{(i+1)}(\xi_k) = \mathcal{N}^c \left( \xi_k; \hat{\xi}_{k|k}^{(i+1)}, \Sigma_{k|k}^{(i+1)} \right), \quad (37)$$

where  $\hat{\xi}_{k|k}^{(i+1)}$  and  $\Sigma_{k|k}^{(i+1)}$  can be determined by following the update steps

$$\hat{\xi}_{k|k}^{(i+1)} = \hat{\xi}_{k|k-1} + K_k^{(i+1)} (z_k - H \hat{\xi}_{k|k-1}), \quad (38)$$

$$K_k^{(i+1)} = \widehat{\Sigma}_{k|k-1}^{(i)} H^T \left( H \widehat{\Sigma}_{k|k-1}^{(i)} H^T + \Sigma_V \right)^{-1}, \quad (39)$$

$$\Sigma_{k|k}^{(i+1)} = \widehat{\Sigma}_{k|k-1}^{(i)} - K_k^{(i+1)} H \widehat{\Sigma}_{k|k-1}^{(i)}, \quad (40)$$

where  $K_k^{(i+1)}$  denotes the calibrated Kalman gain matrix at the  $i + 1$ th iteration.

#### 3.4.2. Fixed-Point Iteration of the Auxiliary Random Variable $\gamma_k$

Using (34) and conditional independency, the  $\log q^{(i+1)}(\gamma_k)$  in (29) can be rewritten as

$$\begin{aligned} \log q^{(i+1)}(\gamma_k) &= E_{-\gamma_k}^{(i+1)} \left[ \log p \left( \xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k, z_{1:k} \right) \right] + c_{\gamma_k} \\ &= \frac{E_{\tau_k}^{(i)}[\tau_k] + 1}{2} \log \gamma_k - \frac{1}{2} \left\{ E_{-\gamma_k}^{(i)} \left[ \left( \xi_k - \hat{\xi}_{k|k-1} \right)^T \Sigma_{k|k-1}^{-1} \left( \xi_k - \hat{\xi}_{k|k-1} \right) \right] + E_{\tau_k}^{(i)}[\tau_k] \right\} \gamma_k + c_{\gamma_k} \\ &= \frac{E_{\tau_k}^{(i)}[\tau_k] + 1}{2} \log \gamma_k - \frac{1}{2} \left\{ \text{tr} \left( E_{\xi_k}^{(i)} \left[ \left( \xi_k - \hat{\xi}_{k|k-1} \right) \left( \xi_k - \hat{\xi}_{k|k-1} \right)^T \right] E_{\Sigma_{k|k-1}}^{(i)} \left[ \Sigma_{k|k-1}^{-1} \right] \right) + E_{\tau_k}^{(i)}[\tau_k] \right\} \gamma_k + c_{\gamma_k}. \end{aligned} \quad (41)$$

where  $E_{-\gamma_k}^{(i)}[\cdot]$  represents the mathematical expectation at  $i$ th iteration for variables except the  $\gamma_k$  while  $E_{\tau_k}^{(i)}[\cdot]$  denotes the mathematical expectation at  $i$ th iteration for the parameter  $\tau_k$ . Note that the approximate  $\hat{\xi}_{k|k}^{(i+1)}$  and  $\Sigma_{k|k}^{(i+1)}$  updated in (35)~(40) can be used to propagate Equation (41) and the term  $E_{\xi_k}^{(i)} \left[ \left( \xi_k - \hat{\xi}_{k|k-1} \right) \left( \xi_k - \hat{\xi}_{k|k-1} \right)^T \right]$  can be reformulated as

$$\begin{aligned} E_{-\gamma_k}^{(i)} \left[ \left( \xi_k - \hat{\xi}_{k|k-1} \right) \left( \xi_k - \hat{\xi}_{k|k-1} \right)^T \right] &= E_{-\gamma_k}^{(i)} \left[ \left( \xi_k - \hat{\xi}_{k|k}^{(i+1)} + \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right) \left( \xi_k - \hat{\xi}_{k|k}^{(i+1)} + \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right)^T \right] \\ &= E_{-\gamma_k}^{(i+1)} \left[ \left( \xi_k - \hat{\xi}_{k|k}^{(i+1)} \right) \left( \xi_k - \hat{\xi}_{k|k}^{(i+1)} \right)^T \right] + \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right) \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right)^T = \Sigma_{k|k}^{(i+1)} + \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right) \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right)^T \end{aligned} \quad (42)$$

Then (41) could be deduced into

$$\begin{aligned} \log q^{(i+1)}(\gamma_k) &= \left( \frac{E_{\tau_k}^{(i)}[\tau_k] + 3}{2} - 1 \right) \log \gamma_k \\ &- \frac{1}{2} \left\{ \text{tr} \left( \Sigma_{k|k}^{(i+1)} E_{\Sigma_{k|k-1}}^{(i)} \left[ \Sigma_{k|k-1}^{-1} \right] \right) + \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right) \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right)^T E_{\Sigma_{k|k-1}}^{(i)} \left[ \Sigma_{k|k-1}^{-1} \right] \right\} \gamma_k + c_{\gamma_k}. \end{aligned} \quad (43)$$

which is actually also a Gamma distribution as (16), i.e.,

$$q^{(i+1)}(\gamma_k) = \Gamma \left( \gamma_k; \eta^{(i+1)}, \theta^{(i+1)} \right), \quad (44)$$

where  $\eta^{(i+1)}$  and  $\theta^{(i+1)}$  are the shape and rate parameter determined by

$$\eta^{(i+1)} = \frac{E_{\tau_k}^{(i)}[\tau_k] + 3}{2}, \quad (45)$$

$$\theta^{(i+1)} = \frac{1}{2} \left\{ \text{tr} \left( \left( \Sigma_{k|k}^{(i+1)} + U_k^{(i)} + \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k}^{(i)} \right) \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k}^{(i)} \right)^T \right) E_{\Sigma_{k|k-1}}^{(i)} \left[ \Sigma_{k|k-1}^{-1} \right] \right) + E_{\tau_k}^{(i)}[\tau_k] \right\}. \quad (46)$$

### 3.4.3. Fixed-Point Iteration of the Prior Estimate for Covariance Matrix $\Sigma_{k|k-1}$

With (34), (37), (44) and the independence of  $\xi_k$  and  $\gamma_k$ , (30) can be rewritten as

$$\begin{aligned} \log q^{(i+1)}(\Sigma_{k|k-1}) &= E_{-\Sigma_{k|k-1}}^{(i+1)} \left[ \log p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k, z_{1:k}) \right] + c_{\Sigma_{k|k-1}} \\ &= -\frac{1}{2} E_{-\Sigma_{k|k-1}}^{(i)} \left[ \gamma_k \left( \xi_k - \hat{\xi}_{k|k-1} \right)^T \Sigma_{k|k-1}^{-1} \left( \xi_k - \hat{\xi}_{k|k-1} \right) \right] - \frac{1}{2} (5 + \lambda_k)^T \log |\Sigma_{k|k-1}| - \frac{1}{2} \text{tr} \left( \Psi_k \Sigma_{k|k-1}^{-1} \right) + c_{\Sigma_{k|k-1}} \\ &= -\frac{1}{2} (5 + \lambda_k)^T \log |\Sigma_{k|k-1}| - \frac{1}{2} \text{tr} \left\{ \left( \Psi_k + E_{\xi_k}^{(i+1)} [\xi_k] B_k^{(i+1)} \right) \Sigma_{k|k-1}^{-1} \right\} + c_{\Sigma_{k|k-1}} \end{aligned} \quad (47)$$

where the term  $B_k^{(i+1)}$  is calculated as

$$\begin{aligned} B_k^{(i+1)} &= E_{-\Sigma_{k|k-1}}^{(i+1)} \left[ \left( \xi_k - \hat{\xi}_{k|k-1} \right) \left( \xi_k - \hat{\xi}_{k|k-1} \right)^T \right] \\ &= E_{-\Sigma_{k|k-1}}^{(i+1)} \left[ \left( \xi_k - \hat{\xi}_{k|k}^{(i+1)} + \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right) \left( \xi_k - \hat{\xi}_{k|k}^{(i+1)} + \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right)^T \right] \\ &= \Sigma_{k|k}^{(i+1)} + \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right) \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right)^T, \end{aligned} \quad (48)$$

Therefore, using (24) and (47),  $q^{(i+1)}(\Sigma_{k|k-1})$  is also an inverse Wishart distribution as

$$q^{(i+1)}(\Sigma_{k|k-1}) = \text{IW}(\Sigma_{k|k-1}; \hat{\lambda}_k^{(i+1)}, \hat{\Psi}_k^{(i+1)}), \quad (49)$$

where  $\hat{\lambda}_k^{(i+1)}, \hat{\Psi}_k^{(i+1)}$  are given as

$$\hat{\lambda}_k^{(i+1)} = \lambda_k + 1, \quad (50)$$

$$\hat{\Psi}_k^{(i+1)} = \Psi_k + E_{\xi_k}^{(i+1)} [\xi_k] B_k^{(i+1)}. \quad (51)$$

### 3.4.4. Fixed-Point Iteration of the Prior Estimate for Parameter $\tau_k$

With (34), (37), (44), and (49), (33) can be rewritten as

$$\begin{aligned} \log q^{(i+1)}(\tau_k) &= E_{-\tau_k}^{(i+1)} \left[ \log p(\xi_k, \gamma_k, \Sigma_{k|k-1}, \tau_k, z_{1:k}) \right] + c_{\tau_k} \\ &= \frac{\tau_k}{2} E_{\gamma_k}^{(i+1)} [\log \gamma_k] - \frac{1}{2} E_{\gamma_k}^{(i+1)} [\gamma_k] \tau_k + \frac{\tau_k}{2} \log \frac{\tau_k}{2} - \log \Gamma \left( \frac{\tau_k}{2} \right) + (a_k - 1) \log \tau_k - b_k \tau_k + c_{\tau_k}. \end{aligned} \quad (52)$$

With Stirling's equation  $\log \Gamma \left( \frac{\tau_k}{2} \right) \approx \left( \frac{\tau_k - 1}{2} \right) \log \left( \frac{\tau_k}{2} \right) - \frac{\tau_k}{2}$ , (52) can be approximated as

$$\log q^{(i+1)}(\tau_k) = \left( a_k + \frac{1}{2} - 1 \right) \log \tau_k - \left( b_k - \frac{1}{2} - \frac{1}{2} E_{\gamma_k}^{(i+1)} [\log \gamma_k] + \frac{1}{2} E_{\gamma_k}^{(i+1)} [\gamma_k] \right) \tau_k + c_{\tau_k}. \quad (53)$$

Therefore, using (16) and (53),  $q^{(i+1)}(\tau_k)$  can be updated as a Gamma distribution, i.e.,

$$q^{(i+1)}(\tau_k) = \Gamma \left( \tau_k; \hat{a}_k^{(i+1)}, \hat{b}_k^{(i+1)} \right), \quad (54)$$

where  $\hat{a}_k^{(i+1)}, \hat{b}_k^{(i+1)}$  are given as

$$\hat{a}_k^{(i+1)} = a_k + \frac{1}{2}, \quad (55)$$

$$\hat{b}_k^{(i+1)} = b_k - \frac{1}{2} - \frac{1}{2} E_{\gamma_k}^{(i+1)} [\log \gamma_k] + \frac{1}{2} E_{\gamma_k}^{(i+1)} [\gamma_k]. \quad (56)$$

### 3.5. The Variational Bayesian Iteration Based Robust Student's $t$ Invariant Kalman Filter

To propagate the approximate posterior distributions, using (44), (49) and (54), the expectations  $E_{\gamma_k}^{(i+1)} [\log \gamma_k]$ ,  $E_{\gamma_k}^{(i+1)} [\gamma_k]$ ,  $E_{\Sigma_{k|k-1}}^{(i+1)} [\Sigma_{k|k-1}^{-1}]$ , and  $E_{\tau_k}^{(i+1)} [\tau_k]$  can be calculated as follows

$$E_{\gamma_k}^{(i+1)} [\gamma_k] = \eta^{(i+1)} / \theta^{(i+1)}, \quad (57)$$

$$E_{\gamma_k}^{(i+1)} [\log \gamma_k] = \psi \left( \eta^{(i+1)} \right) - \log \theta^{(i+1)}, \quad (58)$$

$$E_{\Sigma_{k|k-1}}^{(i+1)} [\Sigma_{k|k-1}^{-1}] = \left( \hat{\lambda}_k^{(i+1)} - 4 \right) \left( \hat{\Psi}_k^{(i+1)} \right)^{-1}, \quad (59)$$

$$E_{\tau_k}^{(i+1)} [\tau_k] = \hat{a}_k^{(i+1)} / \hat{b}_k^{(i+1)}, \quad (60)$$

where  $\psi(\cdot)$  represents a digamma function [42]. After  $N$  times of fixed-point iterations, the required posterior probability density function of  $\xi_k$  could be approximated as

$$q(\xi_k) \approx \mathcal{N}^c(\xi_k; \hat{\xi}_{k|k}^{(N)}, \Sigma_{k|k}^{(N)}) = \mathcal{N}^c(\xi_k; \hat{\xi}_{k|k}, \Sigma_{k|k}), \quad (61)$$

where  $\hat{\xi}_{k|k}, \Sigma_{k|k}$  are respectively the posterior estimate for invariant error  $\xi_k$  and the posterior estimate error covariance parameter. Then, resorting to the equivalence of probability distribution, the distributing of the propagated rotation state  $R_k$  can be updated as the concentrated Gaussian distribution with mean being  $\hat{R}_{k|k}$  and covariance being  $\Sigma_{k|k}$ , i.e.,

$$q(R_k) = \mathcal{G}(R_k; \hat{R}_{k|k}, \Sigma_{k|k}), \quad (62)$$

$$\hat{R}_{k|k} = \exp\left(\left[K_k^{(N)}(z_k - H\hat{\xi}_{k|k-1})\right]^\wedge\right) \hat{R}_{k|k-1} = \exp\left(\left[K_k^{(N)}(z_k - H\hat{\xi}_{k|k-1})\right]^\wedge\right) \hat{R}_{k-1|k-1} \Omega_{k-1} \quad (63)$$

where the iterative Kalman gain matrix  $K_k^{(N)}$  is calculated in (39). The Equations (62) and (63) are actually the projection of the Euclidean space Equation (38) into Lie group space.

Therefore, the filtering steps of the proposed approach include the steps (36)~(40), (42)~(46), (48)~(51), and (54)~(63). The detailed implementation of the proposed method for attitude estimation is presented in Algorithm 1. For the heavy-tailed attitude estimation, the nominal  $\tilde{\Sigma}_{k|k-1}$  could be selected as (11) to initialize the filter parameter  $\Psi_k$  and  $\lambda_k$ ; the invariant error is defined as the Lie exponential error of  $\hat{R}_{k-1|k-1}$  and  $R_k$  so it could be assumed to be zero at the initialization to propagate the prior estimate  $\hat{\xi}_{k|k-1} = \hat{\xi}_{k-1|k-1}$ .

In invariant Kalman filter, the prior error estimate  $\hat{R}_{k|k-1}$  is actually based on all the historical observations  $Y_{1:k-1}$  before time instant  $k$ . In this work, the covariance parameter  $\Sigma_{k|k-1}$  is also inferred using the historical observations  $Y_{1:k-1}$  until time instant  $k$ . As for the inverse Wishart distribution of  $\Sigma_{k|k-1}$ , at time instant  $k$  the degree of parameter is set as  $\lambda_k = k + 4$  and the inverse scale matrix parameter is set as  $\Psi_k = k\tilde{\Sigma}_{k|k-1}$ .

In practical numerical integration of the attitude dynamic model, if there exists severe maneuvering, the system process noise will have a heavy-tailed distribution and the degree of freedom parameter  $\tau_k$  depends on the maneuvering intensity. According to our experience, the initial estimates for the shape parameter  $a_k$  and rate parameter  $b_k$  could be set as  $a_k/b_k \in [3, 10]$  for the iteratively updating  $\hat{a}_k, \hat{b}_k$  in Algorithm 1.

---

**Algorithm 1.** The filtering steps of one time instant in the proposed approach for attitude estimation.

---

**Inputs:**  $\hat{R}_{k-1|k-1}, \Sigma_{k-1|k-1}, \hat{\xi}_{k-1|k-1}, \Omega_{k-1}, H, Y_k, \Sigma_w, \Sigma_v, \lambda_k, \Psi_k, a_k, b_k, N$

1. Predict the nominal invariant error  $\hat{\xi}_{k|k-1}$  and rotation  $\hat{R}_{k|k-1}$  according to (8) and (10)

$$\hat{\xi}_{k|k-1} = \hat{\xi}_{k-1|k-1}, \hat{R}_{k|k-1} = \hat{R}_{k-1|k-1} \Omega_{k-1}$$

2. Predict the nominal prior error covariance  $\tilde{\Sigma}_{k|k-1}$  according to (11)

$$\tilde{\Sigma}_{k|k-1} = \Sigma_{k-1|k-1} + \Sigma_w$$

3. Calculate the converted innovation according to (9)

$$z_k = \hat{R}_{k|k-1} Y_k - \begin{pmatrix} b' \\ b'' \end{pmatrix}$$

4. Initialize the prior parameters

$$\lambda_k^{(0)} = \lambda_k, \hat{\Psi}_k^{(0)} = \Psi_k = (\lambda_k - 4)\tilde{\Sigma}_{k|k-1}, E_{\gamma_k}^{(0)}[\gamma_k] = 1$$

5. Calculate initial expectations according to (24) and (25)

$$E_{\Sigma_{k|k-1}}^{(0)}[\Sigma_{k|k-1}^{-1}] = (\hat{\lambda}_k^{(0)} - 4)(\hat{\Psi}_k^{(0)})^{-1}, E_{\tau_k}^{(0)}[\tau_k] = a_k/b_k$$

**for**  $i = 0:N - 1$

6. Update  $q^{(i+1)}(\xi_k) = \mathcal{N}^c(\xi_k; \hat{\xi}_{k|k}^{(i+1)}, \Sigma_{k|k}^{(i+1)})$  according to (36)~(40)

$$\hat{\Sigma}_{k|k-1}^{(i)} = \left\{ E_{\Sigma_{k|k-1}}^{(i)}[\Sigma_{k|k-1}^{-1}] \right\}^{-1} / E_{\gamma_k}^{(i)}[\gamma_k], K_k^{(i+1)} = \hat{\Sigma}_{k|k-1}^{(i)} H^T \left( H \hat{\Sigma}_{k|k-1}^{(i)} H^T + \Sigma_v \right)^{-1},$$

$$\Sigma_{k|k}^{(i+1)} = \hat{\Sigma}_{k|k-1}^{(i)} - K_k^{(i+1)} H \hat{\Sigma}_{k|k-1}^{(i)} K_k^{(i+1)}, \hat{\xi}_{k|k}^{(i+1)} = \hat{\xi}_{k|k-1} + K_k^{(i+1)} (z_k - H \hat{\xi}_{k|k-1})$$


---

**Algorithm 1:** *Cont.*

7. Update  $q^{(i+1)}(\gamma_k) = \Gamma(\gamma_k; \eta^{(i+1)}, \theta^{(i+1)})$  according to (44)~(46)

$$\eta^{(i+1)} = \frac{E_{\tau_k}^{(i)}[\tau_k] + 3}{2},$$

$$\theta^{(i+1)} = \frac{1}{2} \left\{ \text{tr} \left( \left( \Sigma_{k|k}^{(i+1)} + \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right) \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right)^T \right) E_{\Sigma_{k|k-1}^{(i)}} \left[ \Sigma_{k|k-1}^{-1} \right] \right) + E_{\tau_k}^{(i)}[\tau_k] \right\}$$

8. Update  $q^{(i+1)}(\Sigma_{k|k-1}) = \text{IW}(\Sigma_{k|k-1}; \hat{\lambda}_k^{(i+1)}, \hat{\Psi}_k^{(i+1)})$  according to (48)~(51)

$$B_k^{(i+1)} = \Sigma_{k|k}^{(i+1)} + \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right) \left( \hat{\xi}_{k|k}^{(i+1)} - \hat{\xi}_{k|k-1} \right)^T, \hat{\lambda}_k^{(i+1)} = \lambda_k + 1, \hat{\Psi}_k^{(i+1)} = \Psi_k + E_{\xi_k}^{(i+1)}[\xi_k] B_k^{(i+1)}$$

9. Update expectations  $E_{\gamma_k}^{(i+1)}[\log \gamma_k]$ ,  $E_{\gamma_k}^{(i+1)}[\gamma_k]$  and  $E_{\Sigma_{k|k-1}^{(i+1)}}[\Sigma_{k|k-1}^{-1}]$  according to (57)~(59)

$$E_{\gamma_k}^{(i+1)}[\gamma_k] = \eta^{(i+1)} / \theta^{(i+1)}, \quad E_{\gamma_k}^{(i+1)}[\log \gamma_k] = \psi(\eta^{(i+1)}) - \log \theta^{(i+1)},$$

$$E_{\Sigma_{k|k-1}^{(i+1)}}[\Sigma_{k|k-1}^{-1}] = \left( \hat{\lambda}_k^{(i+1)} - 4 \right) \left( \hat{\Psi}_k^{(i+1)} \right)^{-1}$$

10. Update  $q^{(i+1)}(\tau_k) = \Gamma(\tau_k; \hat{a}_k^{(i+1)}, \hat{b}_k^{(i+1)})$  according to (55) and (56)

$$\hat{a}_k^{(i+1)} = a_k + \frac{1}{2}, \hat{b}_k^{(i+1)} = b_k - \frac{1}{2} - \frac{1}{2} E_{\gamma_k}^{(i+1)}[\log \gamma_k] + \frac{1}{2} E_{\gamma_k}^{(i+1)}[\gamma_k]$$

11. Calculate the expectation  $E_{\tau_k}^{(i+1)}[\tau_k]$  according to (60)

$$E_{\tau_k}^{(i+1)}[\tau_k] = \hat{a}_k^{(i+1)} / \hat{b}_k^{(i+1)}$$

**end for**

12. Update the posterior rotation group and its covariance according to (63)

$$\hat{R}_{k|k} = \exp \left( \left[ K_k^{(N)} (z_k - H \hat{\xi}_{k|k-1}) \right]^\wedge \right) \hat{R}_{k|k-1}, \Sigma_{k|k} = \Sigma_{k|k}^{(N)}$$

**Outputs:**  $\hat{R}_{k|k}, \Sigma_{k|k}$

The number of  $N$  iterations is also a crucial parameter for proposed approach and, generally, it should be set to a value larger than the system dimension to guarantee the local convergence of variational Bayesian iterations; nevertheless, a too larger value is sure to increase the computational cost for algorithm implementation. The balance between precision and cost should be considered according to the detailed applications.

#### 4. Numerical Simulations

To further demonstrate the performance of the variational Bayesian iteration-based invariant Kaman filter, the attitude estimation system (6) (7) is simulated with parameter  $\Sigma_{0|0} = 0.5236^2 I_{3 \times 3}$ ,  $\Sigma_w = 0.01745^2 I_{3 \times 3}$ ,  $b' = [1, 0, 0]^T$ ,  $b'' = [0, 1, 0]^T$ ,  $\Sigma_{v'} = 0.0873^2 I_{3 \times 3}$ ,  $\Sigma_{v''} = 0.0873^2 I_{3 \times 3}$ . The real attitude trajectories are generated for 5000 s and the heavy tailed process noise is generated with heavy tailed outliers according to [42], i.e.,

$$p(w_k) = \begin{cases} \mathcal{N}^c(w_k; 0_{3 \times 1}, \Sigma_w) & \text{w.p. } 1 - \alpha\% \\ \mathcal{N}^c(w_k; 0_{3 \times 1}, \beta \Sigma_w) & \text{w.p. } \alpha\% \end{cases}, \quad (64)$$

which means that  $\alpha$  percentage of process noise is drawn from a Gaussian distribution with covariance  $\Sigma_w$  while the covariance of  $\alpha$  percentage of process noise is severely increased by scale parameter  $\beta$  due to the presence of outliers.

As a comparison, the filtering methods of the proposed approach (RSIKF), IKF [14,16] the Student's t filter (STF) [43], the Huber-based Kalman filter (HKF) [44], and the QeIKF [24,25] are conducted together. For HKF, the tuning parameter is set to 1.345 and the number of iterations is selected as 10; in STF the degree of freedom parameter is selected as 3. In the proposed RSIKF, the filter parameters are set to  $\lambda_k = k + 4$ ,  $a_k = 2$ ,  $b_k = 6$ ,  $N = 10$ . All above filtering methods are implemented with MATLAB codes on a computer with Intel Core i7-7700 CPU at 3.60 GHz. To compare their performance, the error variable



in Lie algebra of 5000 random runs is used to calculate the root mean square error  $RMSE_k$  during the 5000 runs' filtering time and average root mean square error  $ARMSE$ , i.e.,

$$RMSE_k \triangleq \sqrt{\frac{1}{5000} \sum_{l=1}^{5000} \|\tilde{\zeta}_{k,l} - \zeta_{k,l}^*\|^2} \quad (65)$$

$$ARMSE \triangleq \sqrt{\frac{1}{5000 \times 5000} \sum_{k=1}^{5000} \sum_{l=1}^{5000} \|\hat{\zeta}_{k,l} - \zeta_{k,l}\|^2} \quad (66)$$

where  $\zeta_{k,l}$  denotes the true estimate at the  $k$ -th time instant of the  $l$ -th simulation run and  $\hat{\zeta}_{k,l}$  is the corresponding estimate for the true  $\zeta_{k,l}$ ;  $\|\cdot\|$  denotes the Euclidean vector norm.

In this work, three cases of heavy tailed process noises are simulated with  $\alpha = 5, 20$  and  $\beta = 100, 1000$  to investigate the filtering performance of different methods. The  $ARMSE$  result data of different filtering methods are given in Figure 7 while the  $RMSE_k$  result data during the first 500 runs' filtering processes are presented in Figures 8 and 9.

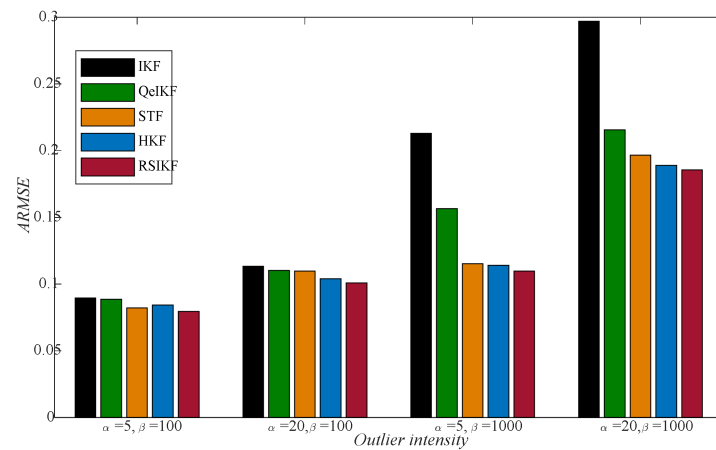


Figure 7. The  $ARMSE$  result of RSIKF with different  $N$  for all cases of  $\alpha$  and  $\beta$ .

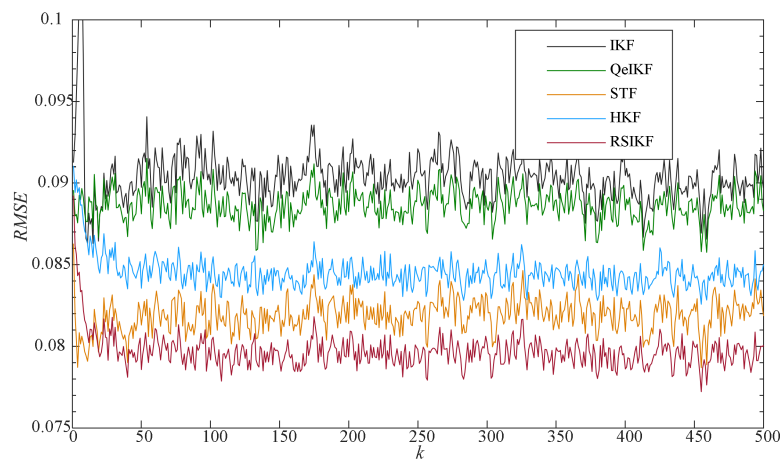
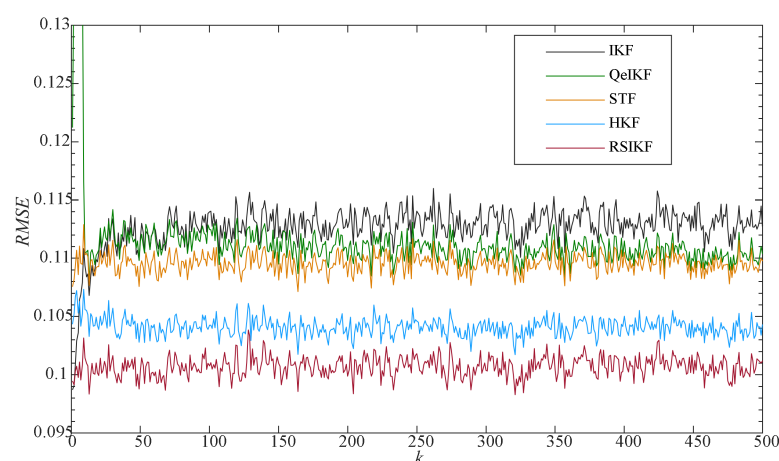
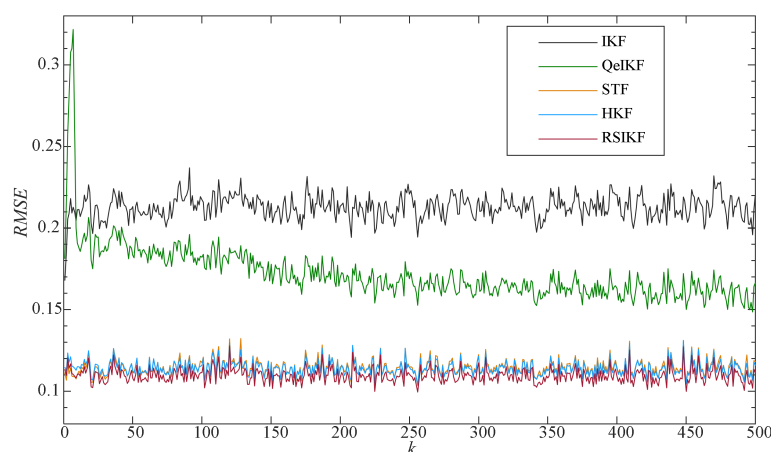


Figure 8. The  $RMSE_k$  result of different methods for the case of  $\alpha = 5$  and  $\beta = 10$ .

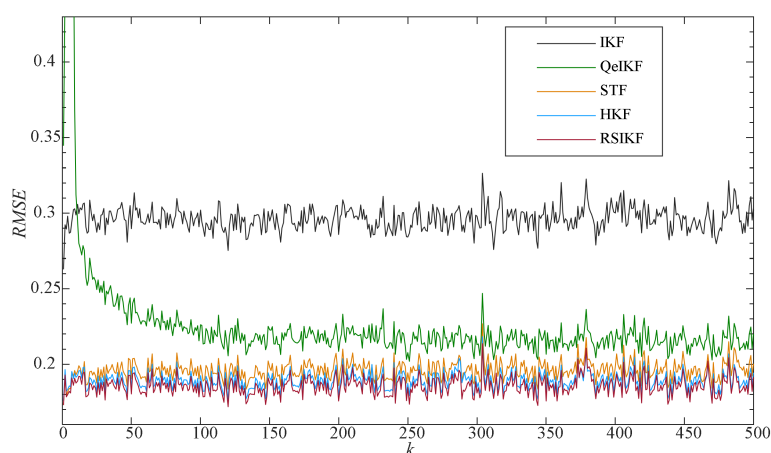


**Figure 9.** The  $RMSE_k$  result of different methods for the case of  $\alpha = 20$  and  $\beta = 10$ .

The  $ARMSE$  result of IKF in Figure 7 and  $RMSE_k$  data of IKF in Figures 8–11 are always the largest in all methods for all cases, which certifies that the estimation performance of IKF is severely corrupted by heavy tailed outliers.



**Figure 10.** The  $RMSE_k$  result of different methods for the case of  $\alpha = 5$  and  $\beta = 100$ .



**Figure 11.** The  $RMSE_k$  result of different methods for the case of  $\alpha = 5$  and  $\beta = 100$ .

In STF, the noise distribution is assumed as a given Student's  $t$  distribution, and the  $ARMSE$  result in Figure 7 and  $RMSE_k$  in Figures 8–11 is smaller than that of IKF. However,

since it still requires accurate prior knowledge of the noise distribution parameter, the unknown or inaccurate parameters still mislead the estimate results.

The QeKF aims to estimate, online, the unknown noise covariance parameter and also output better results than IKF and STF; however, it assumes the distribution as a Gaussian one so its  $ARMSE$  result in Figure 7 and  $RMSE_k$  in Figure 8 for the case  $\alpha = 5$  become worse than the case of  $\alpha = 20$  in Figure 9; while, as shown in Figures 10 and 11, its filtering precision would be significantly degraded by large outliers with  $\beta = 1000$ .

HKF shows some robustness to heavy tailed outliers IKF, STK, and QeKF for the cases of  $\beta = 1000$  as in Figures 10 and 11. Note that, the estimate given by HKF is rather conservative and its robustness is at the cost of inferior precision.

In all simulation cases of  $\alpha$  and  $\beta$ , the  $ARMSE$  and  $RMSE_k$  result of proposed RSIKF is always the least one in Figures 7–11, demonstrating its superiority filtering performance to the other available methods. For RSIKF, different setting of  $N$  from 1~15 are conducted with all cases of  $\alpha$  with the  $ARMSE$  result given in Figure 12; obviously,  $N = 10$  already guarantees the iterative convergence and a larger  $N$  would significantly increase the computational cost.

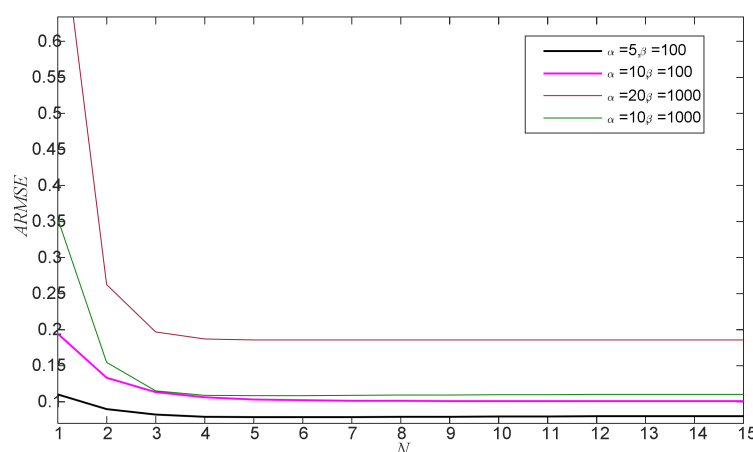


Figure 12. The  $ARMSE$  result of RSIKF with different  $N$  for the cases of  $\alpha$  and  $\beta$ .

## 5. Conclusions

For aerospace, satellite, and robotics engineering, the matrix Lie groups attitude estimation problem with heavy-tailed process noise is investigated. An improved robust Student's  $t$  invariant Kalman filtering is proposed based on a hierarchical Gaussian state-space model. The probability density function of state prediction is defined based on student's  $t$  distribution, while the conjugate prior distributions of the scale matrix and degrees of freedom (dofs) parameter are respectively formulated as the inverse Wishart and Gamma distribution. For attitude estimation state-space model, the Lie groups rotation matrix of spacecraft attitude is inferred together with the scale matrix and dof parameter using the variational Bayesian iteration. The new approach can improve the filtering robustness of the invariant Kalman filter for Lie groups spacecraft attitude estimation problems with heavy-tailed process uncertainty. Some further applications of the proposed methods can also be extended to other problems such as state smoother, parameter identification, state observers, and so on.

**Author Contributions:** Conceptualization, methodology, writing, software, validation, J.W. (Jiaolong Wang) and C.Z.; formal analysis, investigation, resources, J.W. (Jin Wu) and M.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is funded by the Fundamental Research Funds for Central Universities with number JUSRP121022 and National Natural Science Foundation of China with number 62003112.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available as the data also forms part of an ongoing study.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Wu, J.; Shan, S. Dot Product Equality Constrained Attitude Determination from Two Vector Observations: Theory and Astronautical Applications. *Aerospace* **2019**, *6*, 102. [\[CrossRef\]](#)
2. Phisannupawong, T.; Kamsing, P.; Torteeka, P.; Channumsin, S.; Sawangwit, U.; Hematulin, W.; Jarawan, T.; Somjit, T.; Yooyen, S.; Delahaye, D.; et al. Vision-Based Spacecraft Pose Estimation via a Deep Convolutional Neural Network for Noncooperative Docking Operations. *Aerospace* **2020**, *7*, 126. [\[CrossRef\]](#)
3. Soken, H.E.; Sakai, S.-I.; Asamura, K.; Nakamura, Y.; Takashima, T.; Shinohara, I. Filtering-Based Three-Axis Attitude Determination Package for Spinning Spacecraft: Preliminary Results with Arase. *Aerospace* **2020**, *7*, 97. [\[CrossRef\]](#)
4. Carletta, S.; Teofilatto, P.; Farissi, M.S. A Magnetometer-Only Attitude Determination Strategy for Small Satellites: Design of the Algorithm and Hardware-in-the-Loop Testing. *Aerospace* **2020**, *7*, 3. [\[CrossRef\]](#)
5. Louédec, M.; Jaulin, L. Interval Extended Kalman Filter-Application to Underwater Localization and Control. *Algorithms* **2021**, *14*, 142. [\[CrossRef\]](#)
6. Pan, C.; Qian, N.; Li, Z.; Gao, J.; Liu, Z.; Shao, K. A Robust Adaptive Cubature Kalman Filter Based on SVD for Dual-Antenna GNSS/MIMU Tightly Coupled Integration. *Remote Sens.* **2021**, *13*, 1943. [\[CrossRef\]](#)
7. Zheng, L.; Zhan, X.; Zhang, X. Nonlinear Complementary Filter for Attitude Estimation by Fusing Inertial Sensors and a Camera. *Sensors* **2020**, *20*, 6752. [\[CrossRef\]](#)
8. Deibe, Á.; Antón Nacimiento, J.A.; Cardenal, J.; López Peña, F. A Kalman Filter for Nonlinear Attitude Estimation Using Time Variable Matrices and Quaternions. *Sensors* **2020**, *20*, 6731. [\[CrossRef\]](#)
9. Guo, H.; Liu, H.; Hu, X.; Zhou, Y. A Global Interconnected Observer for Attitude and Gyro Bias Estimation with Vector Measurements. *Sensors* **2020**, *20*, 6514. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Ayala, V.; Román-Flores, H.; Torreblanca Todco, M.; Zapana, E. Observability and Symmetries of Linear Control Systems. *Symmetry* **2020**, *12*, 953. [\[CrossRef\]](#)
11. Bonnabel, S.; Martin, P.; Salaun, E. Invariant Extended Kalman Filter: Theory and application to a velocity-aided attitude estimation problem. In Proceedings of the IEEE Conference on Decision & Control, Shanghai, China, 15–18 December 2009. [\[CrossRef\]](#)
12. Vasconcelos, J.; Cunha, R.; Silvestre, C.; Oliveira, P. A nonlinear position and attitude observer on SE(3) using landmark measurements. *Syst. Control Lett.* **2010**, *59*, 155–166. [\[CrossRef\]](#)
13. Chaturvedi, N.; Sanyal, A.; Mcclamroch, A. Rigid-body attitude control using rotation matrices for continuous singularity-free control laws. *IEEE Control Syst. Mag.* **2011**, *31*, 30–51.
14. Barrau, A.; Bonnabel, S. Intrinsic filtering on Lie groups with applications to attitude estimation. *IEEE Trans. Autom. Contr.* **2014**, *60*, 436–449. [\[CrossRef\]](#)
15. Barrau, A.; Bonnabel, S. The invariant extended Kalman filter as a stable observer. *IEEE Trans. Autom. Contr.* **2017**, *62*, 1797–1812. [\[CrossRef\]](#)
16. Barrau, A.; Bonnabel, S. Invariant Kalman filtering. *Annu. Rev. Control Robot. Auton. Syst.* **2018**, *1*, 237–257. [\[CrossRef\]](#)
17. Batista, P.; Silvestre, C.; Oliveira, P. A GES attitude observer with single vector observations. *Automatica* **2012**, *49*, 388–395. [\[CrossRef\]](#)
18. Chirikjian, G.; Kobilarov, M. Gaussian approximation of non-linear measurement models on lie groups. In Proceedings of the IEEE Conference on Decision and Control, Los Angeles, CA, USA, 15–17 December 2014.
19. Barfoot, T.; Furgale, P. Associating uncertainty with three- dimensional poses for use in estimation problems. *IEEE Trans. Robot.* **2014**, *30*, 679–693. [\[CrossRef\]](#)
20. Said, S.; Manton, J. Extrinsic mean of Brownian distributions on compact lie groups. *IEEE Trans. Inf. Theory* **2012**, *58*, 3521–3535. [\[CrossRef\]](#)
21. Huang, Y.; Zhang, Y. A New Process Uncertainty Robust Student's t Based Kalman Filter for SINS/GPS Integration. *IEEE Access* **2017**, *5*, 14391–14404. [\[CrossRef\]](#)
22. Karasalo, M.; Hu, X. An optimization approach to adaptive Kalman filtering. *Automatica* **2011**, *47*, 1785–1793. [\[CrossRef\]](#)
23. Wang, J.; Wang, J.; Zhang, D.; Shao, X.; Chen, G. Kalman filtering through the feedback adaption of prior error covariance. *Signal. Process.* **2018**, *152*, 47–53. [\[CrossRef\]](#)
24. Feng, B.; Fu, M.; Ma, H.; Xia, Y.; Wang, B. Kalman filter with recursive covariance Estimation—sequentially estimating process noise covariance. *IEEE Trans. Ind. Electron.* **2014**, *61*, 6253–6263. [\[CrossRef\]](#)

25. Zanni, L.; Le Boudec, J.; Cherkaoui, R.; Paolone, M. A prediction-error covariance estimator for adaptive Kalman filtering in step-varying processes: Application to power-system state estimation. *IEEE Trans. Contr. Syst. Technol.* **2017**, *25*, 1683–1697. [[CrossRef](#)]
26. Mohamed, A.; Schwarz, K. Adaptive Kalman Filtering for INS/GPS. *J. Geod.* **1999**, *73*, 193–203. [[CrossRef](#)]
27. Ardeshiri, T.; Özkan, E.; Orguner, U.; Gustafsson, F. Approximate Bayesian smoothing with unknown process and measurement noise covariance. *IEEE Signal Process. Lett.* **2015**, *22*, 2450–2454. [[CrossRef](#)]
28. Assa, A.; Plataniotis, K. Adaptive Kalman filtering by covariance sampling. *IEEE Signal Process. Lett.* **2017**, *24*, 1288–1292. [[CrossRef](#)]
29. Huang, Y.; Zhang, Y.; Wu, Z.; Li, N.; Chambers, J. A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices. *IEEE Trans. Autom. Contr.* **2018**, *63*, 594–601. [[CrossRef](#)]
30. Tronarp, F.; Karvonen, T.; Särkkä, S. Student's t-Filters for Noise Scale Estimation. *IEEE Signal Process. Lett.* **2019**, *26*, 352–356. [[CrossRef](#)]
31. Dong, P.; Jing, Z.; Leung, H.; Shen, K.; Wang, J. Student-t mixture labeled multi-Bernoulli filter for multi-target tracking with heavy-tailed noise. *Signal Process.* **2018**, *152*, 331–339. [[CrossRef](#)]
32. Huang, Y.; Zhang, Y.; Li, N.; Wu, Z.; Chambers, J. A novel robust student's t-based Kalman filter. *IEEE Trans. Aerosp. Electron. Syst.* **2017**, *53*, 1545–1554. [[CrossRef](#)]
33. Huang, Y.; Zhang, Y.; Li, N.; Chambers, J. Robust student's t based nonlinear filter and smoother. *IEEE Trans. Aerosp. Electron. Syst.* **2016**, *52*, 2586–2596. [[CrossRef](#)]
34. Ćesić, J.; Marković, I.; Petrović, I. Mixture Reduction on Matrix Lie Groups. *IEEE Signal Process. Lett.* **2017**, *24*, 1719–1723. [[CrossRef](#)]
35. Ćesić, J.; Marković, I.; Bukal, M.; Petrović, I. Extended Information Filter on Matrix Lie Groups. *Automatica* **2017**, *82*, 226–234. [[CrossRef](#)]
36. Kang, D.; Jang, C.; Park, F. Unscented Kalman Filtering for Simultaneous Estimation of Attitude and Gyroscope Bias. *IEEE/ASME Trans. Mechatron.* **2019**, *24*, 350–360. [[CrossRef](#)]
37. Bourmaud, G.; Mégret, R.; Arnaudon, M.; Giremus, A. Continuous-Discrete Extended Kalman Filter on Matrix Lie Groups Using Concentrated Gaussian Distributions. *J. Math. Imaging Vis.* **2015**, *51*, 209–228. [[CrossRef](#)]
38. Tzikas, D.; Likas, A.; Galatsanos, N. The Variational Approximation for Bayesian Inference. *IEEE Signal Process. Mag.* **2008**, *25*, 131–146. [[CrossRef](#)]
39. Pavliotis, G.A. *Applied Stochastic Processes*; Springer: London, UK, 2009.
40. Hagan, T.; Forster, J.J. *Kendall's Advanced Theory of Statistics*; Arnold: London, UK, 2004.
41. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin, Germany, 2007.
42. Huang, Y.; Zhang, Y.; Li, N.; Chambers, J. A Robust Gaussian Approximate Fixed-Interval Smoother for Nonlinear Systems With Heavy-Tailed Process and Measurement Noises. *IEEE Signal Process. Lett.* **2016**, *23*, 468–472. [[CrossRef](#)]
43. Roth, M.; Ozkan, E.; Gustafsson, F. A Student's t filter for heavy tailed process and measurement noise. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 5770–5774.
44. Karlgaard, C.D.; Schaub, H. Huber-Based Divided Difference Filtering. *J. Guid. Control Dyn.* **2012**, *30*, 885–891. [[CrossRef](#)]