Received February 24, 2020, accepted March 28, 2020, date of publication April 6, 2020, date of current version April 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2986022

Maximum Correntropy Square-Root Cubature Kalman Filter for Non-Gaussian Measurement Noise

JINGJING HE[1], CHANGKU SUN[1,2], BAOSHANG ZHANG[1,2], AND PENG WANG[1,2]

[1]State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China
[2]Science and Technology on Electro-Optic Control Laboratory, Luoyang Institute of Electro-Optical Equipment, Luoyang 471009, China

Corresponding author: Peng Wang (wang_peng@tju.eud.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant NSFC-51875407, and in part by the Aeronautical Science Foundation of China under Grant ASFC-20175148006.

ABSTRACT Cubature Kalman filter (CKF) is widely used for non-linear state estimation under Gaussian noise. However, the estimation performance may degrade greatly in presence of heavy-tailed measurement noise. Recently, maximum correntropy square-root cubature Kalman filter (MCSCKF) has been proposed to enhance the robustness against measurement outliers. As is generally known, the square-root algorithms have the benefit of low computational complexity and guaranteed positive semi-definiteness of the state covariances. Therefore, MCSCKF not only possesses the advantages of square-root cubature Kalman filter (SCKF), but also is robust against the heavy-tailed measurement noise. Nevertheless, MCSCKF is prone to the numerical problems. In this paper, we propose a new maximum correntropy square-root cubature Kalman filter (NMCSCKF) based on a cost function which is obtained by a combination of weighted least squares (WLS) to handle the Gaussian process noise and maximum correntropy criterion (MCC) to handle the heavy-tailed measurement noise. Compared to MCSCKF, the proposed method is more time-efficient and most importantly, it avoids the numerical problem. A univariate non-stationary growth model and a multi-rate vision/IMU integrated attitude measurement model are used to demonstrate the superior performance of the proposed method.

INDEX TERMS Square-root cubature Kalman filter, maximum correntropy criterion, vision/IMU integrated measurement.

I. INTRODUCTION

Measuring the attitude of moving objects is important in many fields such as aerospace and industry manufacturing. Sensors fusion systems combine multi-sensors' advantages to overcome single sensor's limitations. Visual sensors have slow output with stable accuracy, whereas Inertial sensors have fast output with error accumulation. By fusing visual measurements with inertial data, accurate and fast attitude estimation can be realized [1], [2].

In the past few years, state estimation problem has drawn many scholars' attention due to its wide use [3], [4]. Kalman filter (KF) is the most used linear estimator under minimum mean square error (MMSE) criterion. To solve the non-linear problems, many non-linear estimators are developed. Extended Kalman filter (EKF) is a direct nonlinear extension of KF, which uses Taylor series expansions to linearize the nonlinear system. Nevertheless, it may lead to low estimation accuracy and even filter divergence [5]. The unscented Kalman filter (UKF) [5] and cubature Kalman filter (CKF) [6] use sigma points to approximate the probability distribution to solve the nonlinear problem, which are widely applied to non-linear systems. The aforementioned filters are derived under Gaussian assumption. Their estimation accuracy may decline significantly when there exist measurement outliers from unreliable sensors [7]. To enhance the robustness against large outliers, many approaches have been proposed over the past few years. Particle filters (PFs) use Monte Carlo random sampling method to approximate the probability distribution of states with many random particles [8]. The high computational cost severely limits PFs' use in practical applications. Variational Bayesian (VB) methods have been embedded into KF to improve the estimation accuracy in presence of heavy-tailed measurement noise.

The associate editor coordinating the review of this manuscript and approving it for publication was Yuan Zhuang.

By using VB method to approximate the posterior state at each time step, VB-based Kalman filters (VBFs) can deal with state estimation problem under non-Gaussian noise effectively [9]–[11]. Huber-based filters, using Huber function that combines minimum ℓ_1 and ℓ_2-norm, also exhibit good robustness against measurement outliers [12], [13].

Recently, information theoretic learning has been gaining more attention for its effectiveness in robust state estimation [14]–[17]. By modifying the optimization criterion using information theoretic quantities (e.g., entropy), high-order statistics of data can be captured. Particularly, KF filters with optimization criteria based on maximum correntropy criterion (MCC) have been proven to cope with heavy-tailed measurement noise successfully. Maximum correntropy Kalman filter (MCKF) for linear system was first proposed in [18] and then extended to non-linear system using UKF [19]. Based on the form of MCKF and its non-linear extension, maximum correntropy square-root cubature Kalman filter (MCSCKF) was newly proposed in [20]. As is generally known, the square-root algorithms have reduced computational complexity, numerical stability and guaranteed positive semi-definiteness of the state covariances [21], [22]. MCSCKF combines the advantages of both square-root cubature Kalman filter (SCKF) [6] and MCKF. However, all the aforementioned MCC-based filters are susceptible to numerical instability problem when large measurement outliers occur [23]. Maximum correntropy criterion Kalman filter (MCCKF) and its square-root form were developed in [24], [25] to overcome the numerical problem. However, they are only applicable to linear systems. Inspired by [24]–[26], we propose a new square-root MCC-based CKF, denoted as NMCSCKF. The proposed algorithm is verified by two examples. Simulation results show that NMCSCKF is robust and stable when the measurement noise is heavy-tailed. Compared to MCSCKF, NMCSCKF not only retains the advantages of MCSCKF, but also avoids the numerical problem. Furthermore, it is shown that NMCSCKF has lower computational cost and higher estimation accuracy.

The rest of the paper is organized as below: Section II provides the preliminaries of MCC, SCKF, as well as the main structure of the existing MCSCKF. In Section III, we derive the NMCSCKF algorithm. Section IV uses two nonlinear models to demonstrate the effectiveness of the proposed filter. The final conclusions are drawn in Section V.

II. PRELIMINARIES

A. SQUARE-ROOT CUBATURE KALMAN FILTER

Considering a non-linear dynamic system with state and measurement equations expressed as follows:

$$\boldsymbol{x}_k = f_{k-1}(\boldsymbol{x}_{k-1}) + \boldsymbol{w}_{k-1} \tag{1}$$
$$z_k = h_k(\boldsymbol{x}_k) + \boldsymbol{v}_k \tag{2}$$

where $\boldsymbol{x}_k \in \mathbb{R}^n$ is the system state, $z_k \in \mathbb{R}^m$ is the measurement vector at discrete time k. $\boldsymbol{w}_{k-1}, \boldsymbol{v}_k$ represent the process and measurement noise with known covariance $\boldsymbol{Q}_{k-1}$ and $\boldsymbol{R}_k$ respectively. $f_{k-1}(\cdot)$ and $h_k(\cdot)$ denote the system and measurement functions. The filtering process is summarized as follows:

1) TIME UPDATE

Generate the cubature points according to the cubature rule:

$$\boldsymbol{x}_{i,k-1} = \sqrt{n}\boldsymbol{S}_{k-1|k-1}\,[1]_i + \hat{\boldsymbol{x}}_{k-1},\ i = 1, 2, \cdots, 2n \tag{3}$$

Calculate the propagated cubature points by:

$$\chi_{i,k|k-1} = f(\boldsymbol{x}_{i,k-1}) \tag{4}$$

where $\boldsymbol{S}_{k-1|k-1}$ is the square-root of the covariance matrix $\boldsymbol{P}_{k-1|k-1}$ at time $k-1$.

The predicted state is computed as:

$$\hat{\boldsymbol{x}}_{k|k-1} = \sum_{i=1}^{2n} \frac{1}{2n} \chi_{i,k|k-1} \tag{5}$$

The square-root of the predicted error covariance $\boldsymbol{P}_{k|k-1}$, denoted as $\boldsymbol{S}_{k|k-1}$, can be obtained by:

$$[\boldsymbol{S}_{k|k-1} 0] = [\chi^*_{k|k-1} \boldsymbol{S}_{Q,k-1}]\Theta \tag{6}$$

where $\boldsymbol{S}_{Q,k-1}$ represents the square-root of $\boldsymbol{Q}_{k-1}$. Θ is an orthogonal operator and the weighted, centered matrix

$$\chi^*_{k|k-1} = 1/\sqrt{2n}\cdot\left[\chi_{1,k|k-1} - \widehat{\boldsymbol{x}}_{k|k-1}, \ldots \chi_{2n,k|k-1} - \widehat{\boldsymbol{x}}_{k|k-1}\right] \tag{7}$$

2) MEASUREMENT UPDATE

Evaluate the cubature points and the propagated cubature points by:

$$\boldsymbol{x}_{i,k|k-1} = \sqrt{n}\boldsymbol{S}_{k|k-1}\,[1]_i + \hat{\boldsymbol{x}}_{k|k-1} \tag{8}$$
$$z_{i,k|k-1} = h_k(\boldsymbol{x}_{i,k-1|k}) \tag{9}$$

The predicted measurement $\hat{z}_k$ and the square-root of the innovation covariance $\boldsymbol{P}_{zz,k}$, denoted as $\boldsymbol{S}_{zz,k}$, are calculated by:

$$\hat{z}_k = \sum_{i=1}^{2n} \frac{1}{2n} z_{i,k|k-1} \tag{10}$$
$$[\boldsymbol{S}_{zz,k}\mathbf{0}] = [\mathcal{Z}_{k|k-1}\mathbf{S}_{R,k}]\Theta \tag{11}$$

where $\boldsymbol{S}_{R,k}$ represents the square-root of $\boldsymbol{R}_k$, $\mathcal{Z}_{k|k-1}$ is calculated by:

$$\mathcal{Z}_{k|k-1} = 1/\sqrt{2n}\cdot\left[z_{1,k|k-1} - \hat{z}_k, \ldots z_{2n,k|k-1} - \hat{z}_k\right] \tag{12}$$

Estimate the cross-covariance by:

$$\boldsymbol{P}_{xz,k} = \chi_{k|k-1}\mathcal{Z}^T_{k|k-1} \tag{13}$$

with

$$\chi_{k|k-1} = 1/\sqrt{2n}\cdot\left[\boldsymbol{x}_{1,k|k-1} - \widehat{\boldsymbol{x}}_{k|k-1}, \ldots \boldsymbol{x}_{2n,k|k-1} - \widehat{\boldsymbol{x}}_{k|k-1}\right] \tag{14}$$

Finally, the updated state $\hat{\boldsymbol{x}}_{k|k}$ and the square-root of the error covariance $\boldsymbol{P}_{k|k}$, denoted as $\boldsymbol{S}_{k|k}$, are obtained as follows:

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k(z_k - \hat{z}_k) \tag{15}$$

$$[\boldsymbol{S}_{k|k}\boldsymbol{0}] = [\boldsymbol{\chi}_{k|k-1} - \boldsymbol{K}_k \boldsymbol{\mathcal{Z}}_{k|k-1}\boldsymbol{K}_k \boldsymbol{S}_{R,k}]\Theta \tag{16}$$

with the Kalman gain computed by:

$$\boldsymbol{K}_k = \boldsymbol{P}_{xz,k}(\boldsymbol{S}_{zz,k}\boldsymbol{S}_{zz,k}^T)^{-1} \tag{17}$$

B. MAXIMUM CORRENTROPY CRITERION

Correntropy can be used to measure the similarity between two random variables. The correntropy between $X, Y \in \mathbb{R}$ with joint distribution $\mathrm{F}_{XY}(x, y)$ is defined by:

$$V(X, Y) = \mathrm{E}[\kappa(X, Y)] = \int \kappa(x, y) d\mathrm{F}_{XY}(x, y) \tag{18}$$

where E is the expectation operator, and $\kappa(\cdot, \cdot)$ denotes a shift-invariant Mercer kernel. In this paper, Gaussian kernel is chosen as the correntropy kernel function:

$$\kappa(x, y) = \mathrm{G}_\sigma(e) = \exp(\frac{-e^2}{2\sigma^2}) \tag{19}$$

where $e = x - y$ and $\sigma > 0$ represents the kernel bandwidth.

As for a finite number of data with unavailable joint density function, the correntropy can be estimated by:

$$\widehat{V}(X, Y) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{G}_\sigma(e(i)) \tag{20}$$

where $e(i) = x(i)-y(i)$, with $\{x(i)-y(i)\}_{i=1}^N$ being the samples drawn from the joint density function $\mathrm{F}_{XY}(x, y)$.

Expanding the Gaussian kernel in Taylor series yields:

$$\widehat{V}(X, Y) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n\sigma^{2n}n!}\mathrm{E}[(X - Y)^{2n}] \tag{21}$$

Clearly, the correntropy can capture all even order moments of $X - Y$ with an appropriate kernel bandwidth.

C. EXISTING MAXIMUM CORRENTROPY SQUARE-ROOT CUBATURE KALMAN FILTER

For the nonlinear model described in (1) and (2), we have:

$$\mathbf{B}_k^{-1}\begin{bmatrix}\hat{\boldsymbol{x}}_{k|k-1} \\ z_k\end{bmatrix} = \mathbf{B}_k^{-1}\begin{bmatrix}\boldsymbol{x}_k \\ h_k(\boldsymbol{x}_k)\end{bmatrix} + \boldsymbol{e}_k \tag{22}$$

with

$$\mathbf{B}_k = \begin{bmatrix}\boldsymbol{S}_{k|k-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_{R,k}\end{bmatrix}$$
$$\mathrm{E}[\boldsymbol{e}_k\boldsymbol{e}_k^T] = \boldsymbol{I}$$

where $\boldsymbol{I}$ is a unit matrix.

The MCC-based cost function is defined by:

$$J_{MCC}(\boldsymbol{x}_k) = \sum_{i=1}^{n+m} \mathrm{G}_\sigma(e_i(k)) \tag{23}$$

where $e_i(k)$ is the i-th element of $\boldsymbol{e}(k)$.

The optimal estimate of $\boldsymbol{x}_k$ can be found by setting the first-order derivative of the cost function equal to zero. MCSCKF can be regarded to improve the robustness by modifying the square-root measurement noise covariance matrix in SCKF with an adjusted matrix denoted as $\boldsymbol{C}_{y,k}$:

$$\boldsymbol{S}_{\widetilde{R},k} = \boldsymbol{S}_{R,k}\boldsymbol{C}_{y,k}^{-1/2} \tag{24}$$

with $\boldsymbol{C}_{y,k} = diag(G_\sigma(e_{n+1}(k)), \ldots, G_\sigma(e_{n+m}(k)))$.

Substituting $\boldsymbol{S}_{R,k}$ with $\boldsymbol{S}_{\widetilde{R},k}$ for operating the SCKF framework in the measurement update process yields the MCSCKF.

As can be seen from (24), there exists matrix inversion. Therefore, MCSCKF tends to face numerical problems since the adjusted matrix is probably singular in presence of large measurement outliers. The authors of MCSCKF noticed this problem and put forward to avoid the numerical problem by setting a preset threshold denoted as c to decide whether or not to conduct the measurement update step. However, how to choose an appropriate c was not discussed, which is of great importance to numerical stability and estimation accuracy.

III. NEW MAXIMUM CORRENTROPY SQUARE-ROOT CUBATURE KALMAN FILTER

To avoid the numerical problem in MCSCKF, we derive NMCSCKF in this section. Firstly, a linear measurement function is built through the statistical linearization technology [27]:

$$z_k = \boldsymbol{A}_k\boldsymbol{x}_k + \boldsymbol{b}_k + \zeta_k \tag{25}$$

with

$$\boldsymbol{A}_k = (\boldsymbol{P}_{xz,k}^T/\boldsymbol{S}_{k|k-1}^T)/\boldsymbol{S}_{k|k-1} \tag{26}$$
$$\boldsymbol{b}_k = \widehat{z}_k - \boldsymbol{A}_k\widehat{\boldsymbol{x}}_{k|k-1} \tag{27}$$
$$\zeta_k \sim \mathcal{N}(0, \boldsymbol{P}_{\zeta\zeta,k}), \boldsymbol{P}_{\zeta\zeta,k} = \boldsymbol{P}_{zz,k} - \boldsymbol{A}_k\boldsymbol{P}_{k|k-1}\boldsymbol{A}_k^T \tag{28}$$

where $\boldsymbol{A}_k$ is the statistical regression matrix, ζ_k is the statistical linearization error.

Secondly, unlike the cost function expressed in (23), the cost function is constructed by a combination of weighted least squares (WLS) and MCC [26]:

$$\begin{aligned} J_{MCC} &= \alpha||\widehat{\boldsymbol{x}}_{k|k} - \widehat{\boldsymbol{x}}_{k/k-1}||^2_{\boldsymbol{P}_{k|k-1}^{-1}} \\ &\quad + \beta\boldsymbol{G}_\sigma(||z_k - \widehat{z}_k - \boldsymbol{A}_k\widehat{\boldsymbol{x}}_{k|k} + \boldsymbol{A}_k\widehat{\boldsymbol{x}}_{k|k-1}||_{\boldsymbol{P}_{\zeta\zeta,k}^{-1}}) \end{aligned} \tag{29}$$

where α and β are adjusting weights, and $\|\boldsymbol{x}\|_{\boldsymbol{A}}^2 = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$. The weights should be properly selected to guarantee the convergence of the filter to CKF when the kernel bandwidth goes infinity. Therefore, we use $\alpha = 1, \beta = -2\sigma^2$ here. The optimal estimate of state $\widehat{\boldsymbol{x}}_{k|k}$ is computed by minimization of J_{MCC} with respect to $\widehat{\boldsymbol{x}}_{k|k}$:

$$\begin{aligned} \frac{\partial J_{MCC}}{\partial \widehat{\boldsymbol{x}}_{k|k}} &= 2\boldsymbol{P}_{k|k-1}^{-1}(\widehat{\boldsymbol{x}}_{k|k} - \widehat{\boldsymbol{x}}_{k|k-1}) \\ &\quad - 2\boldsymbol{L}_k\boldsymbol{A}_k\boldsymbol{P}_{\zeta\zeta,k}^{-1}(z_k - \widehat{z}_k - \boldsymbol{A}_k\widehat{\boldsymbol{x}}_{k|k} + \boldsymbol{A}_k\widehat{\boldsymbol{x}}_{k|k-1}) = 0 \end{aligned} \tag{30}$$

where

$$L_k = G_\sigma(||z_k - \widehat{z}_k - \boldsymbol{A}_k\widehat{\boldsymbol{x}}_{k|k} + \boldsymbol{A}_k\widehat{\boldsymbol{x}}_{k|k-1}||_{\boldsymbol{P}_{\zeta\zeta,k}^{-1}}) \tag{31}$$

We can establish the following equation from (30):

$$(\boldsymbol{P}_{k|k-1}^{-1} + L_k\boldsymbol{A}_k^T\boldsymbol{P}_{\zeta\zeta,k}^{-1}\boldsymbol{A}_k)\widehat{\boldsymbol{x}}_{k|k} = (L_k\boldsymbol{A}_k^T\boldsymbol{P}_{\zeta\zeta,k}^{-1}\boldsymbol{A}_k + \boldsymbol{P}_{k|k-1}^{-1})\widehat{\boldsymbol{x}}_{k|k-1} + L_k\boldsymbol{A}_k^T\boldsymbol{P}_{\zeta\zeta,k}^{-1}(z_k - \widehat{z}_k - \boldsymbol{A}_k\widehat{\boldsymbol{x}}_{k|k} + \boldsymbol{A}_k\widehat{\boldsymbol{x}}_{k|k-1}) \quad (32)$$

Equation (32) can be solved using the fixed point iteration algorithm by updating L_k and $\widehat{\boldsymbol{x}}_{k|k}$ alternately until $\widehat{\boldsymbol{x}}_{k|k}$ has been converged. To save computation time, we here carry out one iteration by replacing $\widehat{\boldsymbol{x}}_{k|k-1}$ with $\widehat{\boldsymbol{x}}_{k|k}$.

By simplifying (32), the state estimation can be calculated by:

$$\widehat{\boldsymbol{x}}_k = \widehat{\boldsymbol{x}}_{k|k-1} + \overline{\boldsymbol{K}}_k(z_k - \widehat{z}_k)^{\cdot} \quad (33)$$

with

$$\overline{\boldsymbol{K}}_k = L_k\boldsymbol{P}_{k|k-1}\boldsymbol{A}_k^T(\widetilde{\boldsymbol{R}}_{e,k})^{-1}, \widetilde{\boldsymbol{R}}_{e,k} = (\boldsymbol{P}_{\zeta\zeta,k} + \boldsymbol{A}_k\boldsymbol{P}_{k|k-1}L_k\boldsymbol{A}_k^T) \quad (34)$$

The error covariance is updated by:

$$\boldsymbol{P}_{k|k} = (\boldsymbol{I} - \overline{\boldsymbol{K}}_k\boldsymbol{A}_k)\boldsymbol{P}_{k|k-1}(\boldsymbol{I} - \overline{\mathbf{K}}_k\boldsymbol{A}_k)^T + \overline{\boldsymbol{K}}_k\boldsymbol{P}_{\zeta\zeta,k}\overline{\boldsymbol{K}}_k^T \quad (35)$$

The detailed derivations of (34) and (35) can be found in Appendix A.

Now we derive (33), (34) and (35) in square-root form. The predicted state estimation $\hat{\boldsymbol{x}}_{k|k-1}$, the square-root predicted error covariance $\boldsymbol{S}_{k|k-1}$, the predicted measurement $\widehat{z}_k$, the square-root innovation covariance matrix $\boldsymbol{S}_{zz,k}$ and the cross-covariance $\boldsymbol{P}_{xz,k}$ are computed in the same way as SCKF does.

To derive the revised Kalman gain $\overline{\boldsymbol{K}}_k$ and the square-root error covariance $\boldsymbol{S}_{k|k}$, we built a pre-array [28] (denoted as $\boldsymbol{V}$) as follows:

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{S}_{\zeta\zeta,k} & L_k^{1/2}\boldsymbol{A}_k\boldsymbol{S}_{k|k-1} \\ \boldsymbol{0} & \boldsymbol{S}_{k|k-1} \end{bmatrix} \quad (36)$$

where $\boldsymbol{S}_{\zeta\zeta,k}$ is the square-root of the statistical linearization error covariance $\boldsymbol{P}_{\zeta\zeta,k}$.

$\boldsymbol{S}_{\zeta\zeta,k}$ can be obtained by:

$$\begin{bmatrix} \boldsymbol{\mathcal{Z}}_{k|k-1} - \boldsymbol{A}_k\boldsymbol{\chi}_{k|k-1} & \boldsymbol{S}_{R,k} \end{bmatrix}\Theta = \begin{bmatrix} \boldsymbol{S}_{\zeta\zeta,k} & \boldsymbol{0} \end{bmatrix} \quad (37)$$

The detailed derivation of (37) can refer Appendix B.

Next, an orthogonal operator Θ is applied to $\boldsymbol{V}$ in order to get a lower triangular matrix (denoted as $\boldsymbol{W}$). We have the following equation:

$$\boldsymbol{V}\Theta = \begin{bmatrix} \boldsymbol{R}_{e,k}^{1/2} & \boldsymbol{0} \\ \overline{\boldsymbol{K}}_k^n & \boldsymbol{S}_{k|k} \end{bmatrix} = \boldsymbol{W} \quad (38)$$

where $\overline{\boldsymbol{K}}_k^n$ is the normalized revised Kalman gain:

$$\overline{\boldsymbol{K}}_k^n = L_k^{1/2}\boldsymbol{P}_{k|k-1}\boldsymbol{A}_k^T\boldsymbol{R}_{e,k}^{-T/2} \quad (39)$$

The detailed derivation of (38) and (39) can refer Appendix C.

The revised Kalman gain $\overline{\boldsymbol{K}}_k$ can be computed using $\overline{\boldsymbol{K}}_k^n$ by:

$$\overline{\boldsymbol{K}}_k = L_k^{1/2}\overline{\boldsymbol{K}}_k^n\boldsymbol{R}_{e,k}^{-1/2} \quad (40)$$

Finally, the estimated state is updated by:

$$\widehat{\boldsymbol{x}}_{k|k} = \widehat{\boldsymbol{x}}_{k|k-1} + L_k^{1/2}\overline{\boldsymbol{K}}_k^n\boldsymbol{R}_{e,k}^{-1/2}(z_k - \widehat{z}_k) \quad (41)$$

The square-root error covariance $\boldsymbol{S}_{k|k}$ is read off from $\boldsymbol{W}$ directly.

The NMCSCKF algorithm can be summarized as follows:

1) Assume an initial estimate state $\widehat{\boldsymbol{x}}_{0|0}$, a square-root error covariance $\boldsymbol{S}_{0|0}$; Select an appropriate kernel bandwidth σ;
2) Compute the predicted state $\widehat{\boldsymbol{x}}_{k|k-1}$ and the square-root predicted error covariance $\boldsymbol{S}_{k|k-1}$ using (3)−(7);
3) Compute the predicted measurement $\widehat{z}_k$, the square-root innovation covariance $\boldsymbol{S}_{zz,k}$ and the cross covariance $\boldsymbol{P}_{xz,k}$ using (8)−(14);
4) Compute the statistical regression matrix $\boldsymbol{A}_k$ using (26);
5) Compute the square-root statistical linearization error covariance $\boldsymbol{S}_{\zeta\zeta,k}$ using (37), and obtain the statistical linearization error covariance by:

$$\boldsymbol{P}_{\zeta\zeta,k} = \boldsymbol{S}_{\zeta\zeta,k}\boldsymbol{S}_{\zeta\zeta,k}^T$$

6) Compute the adjusting item L_k using (31);
7) Build the pre-array $\boldsymbol{V}$ through (36);
8) Apply an orthogonal operator to the pre-array $\boldsymbol{V}$ for computing the post-array $\boldsymbol{W}$; Read off $\boldsymbol{R}_{e,k}^{1/2}$, $\overline{\boldsymbol{K}}_k^n$ and $\boldsymbol{S}_{k|k}$ from $\boldsymbol{W}$;
9) The updated state is estimated by (41).

Theorem 1: NMCSCKF is equivalent to SCKF when the kernel bandwidth σ goes infinity.

Proof of Theorem 1. As σ goes infinity, L_k will approach 1, and NMCSCKF will reduce to SCKF.

As for NMCSCKF, the resulting adjusted item L_k will approach zero when extremely large measurement error occurs. In this case, $\widehat{\boldsymbol{x}}_{k|k}$ and $\boldsymbol{S}_{k|k}$ are equal to $\widehat{\boldsymbol{x}}_{k|k-1}$ and $\boldsymbol{S}_{k|k-1}$ respectively since the revised Kalman gain $\overline{\boldsymbol{K}}_k$ is close to zero. However, MCSCKF may fail to work properly since it needs to calculate the inversion of zero matrix when computing the adjusted square-root measurement covariance $\boldsymbol{S}_{\overline{R},k}$.

IV. SIMULATION EXAMPLES

We use two examples to illustrate the performance of NMCSCKF. The first example shows the influence of kernel bandwidth in NMCSCKF. In the second example, NMCSCKF is compared with SCKF [3] and other robust filters including MCSCKF [23], Huber-based cubature Kalman filter (HSCKF) [35], Variational Bayesian cubature Kalman filter (VB-CKF) [9] to prove its superiority. In this section, NMCSCKF with kernel bandwidth of x is denoted as NMCSCKF-x. MCSCKF with kernel bandwidth of x is denoted as MCSCKF-x. All filters are initialized with the same condition in each Monte Carlo run.

A. UNIVARIATE NON-STATIONARY GROWTH MODEL

First, we use a benchmark example called univariate nonstationary growth model [18,19]. The state and measurement equations are given by:

$$x_k = 0.5x_{k-1} + 25\frac{x_{k-1}}{1+x_{k-1}^2} + 8\cos(1.2(k-1)) + w_{k-1} \tag{42}$$

$$z_k = \frac{x_k^2}{20} + v_k \tag{43}$$

where $w_{k-1} \sim \mathcal{N}(0, 2)$, and $v_k \sim \begin{cases} \mathcal{N}(0,1) & w.p.0.8 \\ \mathcal{N}(0,1000) & w.p.0.2 \end{cases}$.

$\mathcal{N}(\mu, \sigma)$ denotes the Gaussian distribution with mean vector μ and covariance matrix σ, $w.p.$ denotes "with probability".

The total root mean square error (TRMSE) is used to evaluate the overall estimation performance in this example. The calculation formula is defined as follows:

$$\text{TRMSE} = \frac{1}{KM}\sum_{k=1}^{K}\sum_{m=1}^{M}\sqrt{(\boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{k|k})^2} \tag{44}$$

where M is the number of Monte Carlo runs and K is the total time steps in each Monte Carlo run.

We use $K = 100$ and $M = 100$ in this example. The TRMSE results of SCKF, NMCSCKF with different kernel bandwidths are listed in Table 1. As can be seen from the results, the estimation accuracy of NMCSCKF would decline if the kernel bandwidth is too large or too small. However, NMCSCKF still has great superiority over SCKF with a rough selection of the kernel bandwidth in presence of non-Gaussian measurement noise.

TABLE 1. TRMSEs under non-Gaussian noise.

Filters	TRMSE
SCKF	19.3102
NMCSCKF-1	8.3512
NMCSCKF-2	7.7084
NMCSCKF-3	7.5310
NMCSCKF-4	7.8770
NMCSCKF-5	8.2735
NMCSCKF-10	10.1047

B. VISION/IMU INTEGRATED ATTITUDE MEASUREMENT MODEL

The state propagation equation is expressed as [29]:

$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + \begin{bmatrix} 1 & \tan\theta_{k-1}\sin\phi_{k-1} & \tan\theta_{k-1}\cos\phi_{k-1} \\ 0 & \cos\phi_{k-1} & -\sin\phi_{k-1} \\ 0 & \sec\theta_{k-1}\sin\phi_{k-1} & \sec\theta_{k-1}\cos\phi_{k-1} \end{bmatrix} \times \vec{\boldsymbol{\omega}}_{k-1} \times \boldsymbol{\tau} + \boldsymbol{w}_{k-1} \tag{45}$$

where $\boldsymbol{x}_k = \begin{bmatrix} \phi_k & \theta_k & \psi_k \end{bmatrix}^T$ with ϕ_k, θ_k and ψ_k being the azimuth angle, pitch angle and rolling angle at time k respectively; $\vec{\boldsymbol{\omega}}_{k-1}$ refers to the angular rate obtained by IMU, τ is the sampling time interval, $\boldsymbol{w}_{k-1}$ is considered to be independent zero-mean Gaussian white process noise with:

$$\boldsymbol{w}_{k-1} \sim \mathcal{N}(0, 5\text{e-}9 \cdot \boldsymbol{I})$$

where $\boldsymbol{I} \in \mathbb{R}^{3\times 3}$ refers to identity matrix.

The angular rate obtained by IMU is simulated with a constant acceleration of $\boldsymbol{a} = [0.1, 0.1, 0.1]^T \text{rad}/s^2$. The angular rate model is constructed as follows:

$$\vec{\boldsymbol{\omega}}_k = \vec{\boldsymbol{\omega}}_{k-1} + \boldsymbol{a}\cdot\boldsymbol{\tau} + \boldsymbol{\varsigma} \tag{46}$$

where $\boldsymbol{\varsigma} \in \mathbb{R}^{3\times 1}$ is a random vector with mean of $[0.005, 0.005, 0.005]^T\, rad/s$.

Since the visual sensors can provide the angular information directly after processing captured visual images [30]–[32], we have the following measurement model:

$$\boldsymbol{z}_k = \boldsymbol{x}_k + \boldsymbol{v}_k \tag{47}$$

where $\boldsymbol{z}_k$ is the measurement vector, $\boldsymbol{v}_k$ represents the measurement noise.

IMU output is usually faster than visual output in reality. We assume that the rate of IMU output is 50 HZ, the rate of visual output is 16.7 HZ. When there is no visual output, only time update is carried out, as expressed in (48):

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{S}_{k|k} = \boldsymbol{S}_{k|k-1} \tag{48}$$

The RMSE and TRMSE in Euler angles are used to evaluate the performance of filters:

$$\text{RMSE}_{Angle}(k) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}((\phi_k - \widehat{\phi}_k)^2 + (\theta_k - \widehat{\theta}_k)^2 + (\psi_k - \widehat{\psi}_k)^2)} \tag{49}$$

$$\text{TRMSE}_{Angle} = \frac{1}{K}\sum_{k=1}^{K} RMSE_{Angle}(k) \tag{50}$$

We choose $K = 250$ and $M = 100$ in this example.

1) GAUSSIAN MEASUREMENT NOISES

Assuming the measurement noise is Gaussian:

$$\boldsymbol{v}_k \sim \mathcal{N}(0, 8\text{e-}5 \cdot \boldsymbol{I})$$

Since the simulation is performed under Gaussian measurement noise without large outliers, the preset threshold in MCSCKF is set to infinity. Table 2 illustrates the in Euler angles. It is shown that SCKF achieves the smallest TRMSE among all filters. VB-CKF and MCC-based robust filters with large kernel bandwidths have similar estimation performance as SCKF. A small kernel bandwidth in NMCSCKF leads to an unsatisfactory estimation; whereas, the accuracy of NMCSCKF increases and gets closer to that of SCKF as the kernel bandwidth becomes large.

TABLE 2. **TRMSEs under Gaussian noise.**

Filters	$TRMSE_{Angle}$ (rad)
SCKF	0.001869
HSCKF	0.003738
VB-CKF	0.001921
NMCSCKF-1	0.003090
NMCSCKF-2	0.001997
NMCSCKF-3	0.001890
NMCSCKF-4	0.001877
NMCSCKF-5	0.001877
NMCSCKF-10	0.001876
MCSCKF-10 (c=infinity)	0.001874

2) NON-GAUSSIAN MEASUREMENT NOISES

We further assume the measurement noise satisfies the following mixed-Gaussian distribution:

$$\boldsymbol{v}_k \sim \begin{cases} \mathcal{N}(0, 8e-5 \cdot \boldsymbol{I}), & w.p.0.8 \\ \mathcal{N}(0, 400 \cdot 8\text{e-}5 \cdot \boldsymbol{I}), & w.p.0.2 \end{cases}$$

Since the measurement contains outliers, the numerical problem of MCSCKF needs to be considered. MCSCKF uses a preset threshold for avoiding the dysfunction caused by matrix inversion. However, how to choose c is not discussed. We perform our simulations with five different c and σ respectively for MCSCKF. The TRMSE in Euler angles are shown in Table 3, where 'NAN' indicates that numerical problem occurs in 100 Monte Carlo runs. As can be seen from Table 3, c should be chosen cautiously to avoid numerical instability and ensure high accuracy. A small c help avoid the numerical problems, whereas a large c improves the accuracy.

TABLE 3. **TRMSEs of MCSCKF with different σ and c.**

Filters	$TRMSE_{Angle}$ (rad)				
	c=1e^{-3}	c=1	c=10	c=100	c=infinity
MCSCKF-1	NAN	NAN	NAN	NAN	NAN
MCSCKF-2	0.00612	NAN	NAN	NAN	NAN
MCSCKF-3	0.00594	0.00385	0.00330	NAN	NAN
MCSCKF-4	0.00591	0.00400	0.00336	0.00336	NAN
MCSCKF-5	0.00591	0.00407	0.00339	0.00339	0.00339
MCSCKF-10	0.00599	0.00425	0.00346	0.00344	0.00344

The TRMSEs in Euler angles of SCKF, VB-CKF, HSCKF, NMCSCKF with different kernel bandwidths and MSCKF with its optimal parameters in Table 3 are listed in Table 4. The corresponding RMSE in Euler angles is plotted in Fig. 1. Fig. 2 shows the details within the black bordered rectangle in Fig. 1. As we can see, SCKF is very sensitive to measurement outliers. Compared to SCKF, other robust filters show improvement in estimation accuracy to a great extent. NMCSCKF obtains the best estimation accuracy with an appropriate kernel bandwidth. The performance of NMCSCKF would decline if the kernel bandwidth is too large or too small. However, even with a rough selection of the kernel bandwidth (i.e., $\sigma = 2, 3, 4, 5$), NMCSCKF still gains

TABLE 4. **TRMSEs under non-Gaussian noise.**

Filters	$TRMSE_{Angle}$ (rad)
SCKF	0.02141
HSCKF	0.005263
VB-CKF	0.003311
NMCSCKF-1	0.004587
NMCSCKF-2	0.003462
NMCSCKF-3	0.003258
NMCSCKF-4	0.003221
NMCSCKF-5	0.003244
NMCSCKF-10	0.003446
MCSCKF-3 (c=10)	0.003304

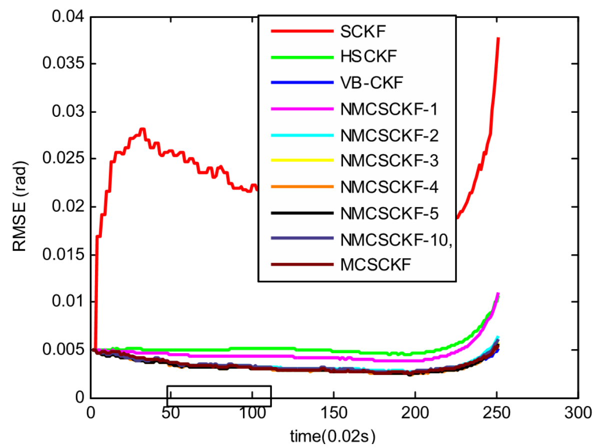


FIGURE 1. **RMSE in Euler angles of different filters.**

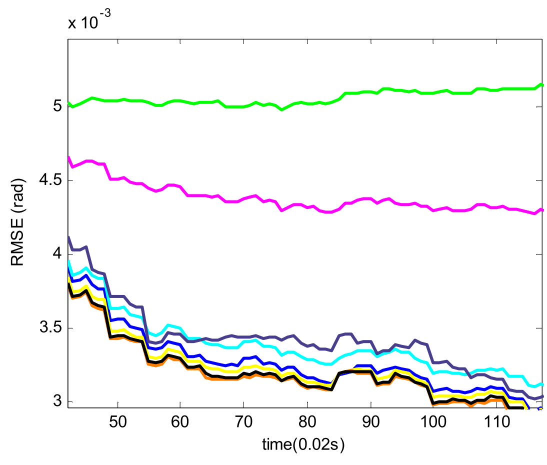


FIGURE 2. **Details within the black bordered rectangle in Fig. 1.**

satisfying results. We can also draw this conclusion from the first example.

Table 5 illustrates the computational times of filters. As we can see from the TRMSE results, both VB-CKF and MCC-based filters with appropriate parameters can obtain high estimation accuracy. However, the execution time is several times higher for VB-CKF than for MCC-based filters. NMCSCKF takes a bit more time than SCKF. Compared to MSCKF, NMCSCKF not only has lower computational cost

TABLE 5. Execution time comparisons.

Filters	Execution time
SCKF	0.09001
HSCKF	0.2388
VB-CKF	0.6165
NMCSCKF	0.1139
MCSCKF	0.1422

TABLE 6. TRMES and average iteration numbers.

Filters	Average iteration numbers	TRMSE_{Angle} (rad)
NMCSCKF-4	1.0000	0.0032961
NMCSCKF-4 (ε=10^{-2})	1.1488	0.0032242
NMCSCKF-4 (ε=10^{-4})	1.8112	0.0032197
NMCSCKF-4 (ε=10^{-6})	3.0864	0.0032129
NMCSCKF-4 (ε=10^{-8})	4.1812	0.0032129

but also obtain higher estimation accuracy. Most importantly, NMCSCKF has good numerical stability.

NMCSCKF uses only one iteration to save computational cost. However, the performance of NMCSCKF can be further improved by a few more iterations. We simply investigate the influence of iteration number here. The stop condition controls the number of iterations, which is defined by:

$$\frac{||\widehat{\boldsymbol{x}}_{k|k}^{t}-\widehat{\boldsymbol{x}}_{k|k}^{t-1}||}{||\widehat{\boldsymbol{x}}_{k|k}^{t-1}||}<\varepsilon \tag{51}$$

where ε is a small positive number, $\widehat{\boldsymbol{x}}_{k|k}^{t}$ represents the optimal estimate of state at the t-th fixed-point iteration.

Table 6 shows the TRMSEs and average iteration numbers for every step with different ε. The estimation accuracy is improved with a smaller ε. However, the improvement is slight and more computation is required.

V. CONCLUSIONS

In this paper, we derive a new maximum correntropy square-root cubature Kalman filter (NMCSCKF) to enhance the robustness in presence of heavy-tailed measurement noise. NMCSCKF can obtain much more accurate estimation than SCKF without much extra computation under impulsive noise. With a large kernel bandwidth, the estimation performance of NMCSCKF will be similar to that of SCKF under Gaussian noise. Simulation results demonstrate that NMCSCKF with a proper kernel bandwidth can outperform other robust filters in both speed and estimation accuracy. Compared to the existing maximum correntropy square-root cubature Kalman filter (MCSCKF), NMCSCKF is also superior in numerical stability.

APPENDIX

APPENDIX A. Derivation of (34) and (35)

By simplifying (32), the revised Kalman gain can be written as follows:

$$\overline{\boldsymbol{K}}_k=(\boldsymbol{P}_{k|k-1}^{-1}+L_k\boldsymbol{A}_k^T\boldsymbol{P}_{\zeta\zeta,k}^{-1}\boldsymbol{A}_k)^{-1}L_k\boldsymbol{A}_k^T\boldsymbol{P}_{\zeta\zeta,k}^{-1} \tag{A.1}$$

By using the matrix inversion lemma

$$(\mathbf{A}+\mathbf{B}\mathbf{C}^{-1}\mathbf{D})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{C}+\mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1} \tag{A.2}$$

with

$$\boldsymbol{P}_{k|k-1}^{-1}\rightarrow\mathbf{A},\ L_k\mathbf{A}_k^T\rightarrow\mathbf{B},\ \boldsymbol{P}_{\zeta\zeta,k}\rightarrow\mathbf{C},\ \mathbf{A}_k\rightarrow\mathbf{D} \tag{A.3}$$

We have:

$$\begin{aligned}\overline{\boldsymbol{K}}_k&=(\boldsymbol{P}_{k|k-1}-\boldsymbol{P}_{k|k-1}L_k\mathbf{A}_k^T(\boldsymbol{P}_{\zeta\zeta,k}+\mathbf{A}_k\boldsymbol{P}_{k|k-1}L_k\mathbf{A}_k^T)^{-1}\\&\quad\mathbf{A}_k\boldsymbol{P}_{k|k-1})\times L_k\boldsymbol{A}_k^T\boldsymbol{P}_{\zeta\zeta,k}^{-1}\\&=\boldsymbol{P}_{k|k-1}L_k\mathbf{A}_k^T(\boldsymbol{I}-(\boldsymbol{P}_{\zeta\zeta,k}+\mathbf{A}_k\boldsymbol{P}_{k|k-1}L_k\mathbf{A}_k^T)^{-1}\\&\quad\times\mathbf{A}_k\boldsymbol{P}_{k|k-1}L_k\mathbf{A}_k^T)\boldsymbol{P}_{\zeta\zeta,k}^{-1}\end{aligned} \tag{A.4}$$

By using the matrix inversion lemma again with:

$$\boldsymbol{I}\rightarrow\mathbf{A},\ \boldsymbol{I}\rightarrow\mathbf{B},\ \boldsymbol{P}_{\zeta\zeta,k}\rightarrow\mathbf{C},\ \mathbf{A}_k\boldsymbol{P}_{k|k-1}L_k\mathbf{A}_k^T\rightarrow\mathbf{D} \tag{A.5}$$

We have the following formula:

$$\begin{aligned}\overline{\boldsymbol{K}}_k&=\boldsymbol{P}_{k|k-1}L_k\mathbf{A}_k^T(\boldsymbol{I}+\boldsymbol{P}_{\zeta\zeta,k}^{-1}\mathbf{A}_k\boldsymbol{P}_{k|k-1}L_k\mathbf{A}_k^T)^{-1}\boldsymbol{P}_{\zeta\zeta,k}^{-1}\\&=\boldsymbol{P}_{k|k-1}L_k\mathbf{A}_k^T(\boldsymbol{P}_{\xi\xi,k}+\mathbf{A}_k\boldsymbol{P}_{k|k-1}L_k\mathbf{A}_k^T)^{-1}\end{aligned} \tag{A.6}$$

Now we give the derivation of (35).

We define $\boldsymbol{\varepsilon}_{x,k}=\boldsymbol{x}_k-\widehat{\boldsymbol{x}}_{k|k}$ and $\boldsymbol{\varepsilon}_{x,k-1}=\boldsymbol{x}_k-\widehat{\boldsymbol{x}}_{k|k-1}$, and the following formulas can be established:

$$\boldsymbol{P}_{k|k}=\mathrm{E}(\boldsymbol{\varepsilon}_{x,k}\boldsymbol{\varepsilon}_{x,k}^T),\ \boldsymbol{P}_{k|k-1}=\mathrm{E}(\boldsymbol{\varepsilon}_{x,k-1}\boldsymbol{\varepsilon}_{x,k-1}^T) \tag{A.7}$$

Combining (25) and (33), We have:

$$\begin{aligned}\boldsymbol{\varepsilon}_{x,k}&=\boldsymbol{x}_k-\widehat{\boldsymbol{x}}_{k|k-1}-\overline{\boldsymbol{K}}_k(z_k-\widehat{z}_k)\\&=\boldsymbol{x}_k-\widehat{\boldsymbol{x}}_{k|k-1}-\overline{\boldsymbol{K}}_k(\boldsymbol{A}_k\boldsymbol{x}_k\\&\quad+\boldsymbol{b}_k+\zeta_k-(\boldsymbol{A}_k\widehat{\boldsymbol{x}}_{k|k-1}+\boldsymbol{b}_k))\\&=(\boldsymbol{I}-\overline{\boldsymbol{K}}_k\boldsymbol{A}_k)\boldsymbol{\varepsilon}_{x,k-1}-\zeta_k\end{aligned} \tag{A.8}$$

Therefore, the error covariance is expressed as follows:

$$\begin{aligned}\boldsymbol{P}_{k|k}&=\mathrm{E}(\boldsymbol{\varepsilon}_{x,k}\boldsymbol{\varepsilon}_{x,k}^T)\\&=(\boldsymbol{I}-\overline{\boldsymbol{K}}_k\boldsymbol{A}_k)\mathrm{E}(\boldsymbol{\varepsilon}_{x,k-1}\boldsymbol{\varepsilon}_{x,k-1}^T)(\boldsymbol{I}-\overline{\boldsymbol{K}}_k\boldsymbol{A}_k)^T\\&\quad-\boldsymbol{K}_k\mathrm{E}(\zeta_k\boldsymbol{\varepsilon}_{x,k-1}^T)(\boldsymbol{I}-\overline{\boldsymbol{K}}_k\boldsymbol{A}_k)^T-\\&\quad(\boldsymbol{I}-\overline{\boldsymbol{K}}_k\boldsymbol{A}_k)\mathrm{E}(\boldsymbol{\varepsilon}_{x,k-1}\zeta_k^T)\boldsymbol{K}_k^T+\overline{\boldsymbol{K}}_k\mathrm{E}(\zeta_k\zeta_k^T)\overline{\boldsymbol{K}}_k^T\\&=(\boldsymbol{I}-\overline{\boldsymbol{K}}_k\boldsymbol{A}_k)\boldsymbol{P}_{k|k-1}(\boldsymbol{I}-\overline{\boldsymbol{K}}_k\boldsymbol{A}_k)^T+\overline{\boldsymbol{K}}_k\boldsymbol{P}_{\zeta\zeta,k}\overline{\boldsymbol{K}}_k^T\end{aligned} \tag{A.9}$$

APPENDIX B. Derivation of (37)

As expressed in (28), the following equalities can be established:

$$\begin{aligned}\boldsymbol{P}_{\zeta\zeta,k}&=\boldsymbol{P}_{zz,k}-\boldsymbol{A}_k\boldsymbol{P}_{k|k-1}\boldsymbol{A}_k^T\\&=\boldsymbol{P}_{zz}-(\boldsymbol{P}_{xz,k}^T\boldsymbol{P}_{k|k-1}^{-1})\boldsymbol{P}_{k|k-1}(\boldsymbol{P}_{xz,k}^T\boldsymbol{P}_{k|k-1}^{-1})^T\\&=\boldsymbol{P}_{zz,k}-\boldsymbol{A}_k\boldsymbol{P}_{xz,k}\end{aligned} \tag{B.1}$$

Some matrix manipulations in (26) lead to:

$$\boldsymbol{P}_{xz,k}^T\boldsymbol{A}_k^T=\boldsymbol{A}_k\boldsymbol{S}_{k|k-1}\boldsymbol{S}_{k|k-1}^T\boldsymbol{A}_k^T \tag{B.2}$$

Adding (B.1) and (B.2) together yields:

$$\begin{aligned}\boldsymbol{P}_{\zeta\zeta,k} &= \boldsymbol{S}_{zz,k}\boldsymbol{S}_{zz,k}^{T} - \boldsymbol{A}_k\boldsymbol{P}_{xz,k}\\ &\quad + \boldsymbol{A}_k\boldsymbol{S}_{k|k-1}\boldsymbol{S}_{k|k-1}^{T}\boldsymbol{A}_k^{T} - \boldsymbol{P}_{xz,k}^{T}\boldsymbol{A}_k^{T} \end{aligned} \tag{B.3}$$

Replacing $\boldsymbol{S}_{k|k-1}\boldsymbol{S}_{k|k-1}^{T}$, $\boldsymbol{S}_{zz,k}\boldsymbol{S}_{zz,k}^{T}$ and $\boldsymbol{P}_{xz,k}$ in (B.3) with:

$$\begin{aligned}\boldsymbol{S}_{k|k-1}\boldsymbol{S}_{k|k-1}^{T} &= \boldsymbol{\chi}_{k|k-1}^{*}\boldsymbol{\chi}_{k|k-1}^{*T}\\ &\quad + \boldsymbol{S}_{Q,k-1}\boldsymbol{S}_{Q,k-1}^{T} = \boldsymbol{\chi}_{k|k-1}\boldsymbol{\chi}_{k|k-1}^{T} \end{aligned} \tag{B.4}$$

$$\boldsymbol{S}_{zz,k}\boldsymbol{S}_{zz,k}^{T} = \boldsymbol{\mathcal{Z}}_{k|k-1}\boldsymbol{\mathcal{Z}}_{k|k-1}^{T} + \boldsymbol{S}_{R,k}\boldsymbol{S}_{R,k}^{T} \tag{B.5}$$

$$\boldsymbol{P}_{xz,k} = \boldsymbol{\chi}_{k|k-1}\boldsymbol{\mathcal{Z}}_{k|k-1}^{T} \tag{B.6}$$

Equation (B.3) can be rewritten in the following form:

$$\begin{aligned}\boldsymbol{P}_{\zeta\zeta,k} &= \boldsymbol{\mathcal{Z}}_{k|k-1}\boldsymbol{\mathcal{Z}}_{k|k-1}^{T} + \boldsymbol{S}_{R,k}\boldsymbol{S}_{R,k}^{T} - \boldsymbol{A}_k\boldsymbol{\chi}_{k|k-1}\boldsymbol{\mathcal{Z}}_{k|k-1}^{T}\\ &\quad + \boldsymbol{A}_k\boldsymbol{\chi}_{k|k-1}\boldsymbol{\chi}_{k|k-1}^{T}\boldsymbol{A}_k^{T} - \boldsymbol{\mathcal{Z}}_{k|k-1}\boldsymbol{\chi}_{k|k-1}^{T}\boldsymbol{A}_k^{T} \end{aligned} \tag{B.7}$$

Therefore, we can derive:

$$\begin{aligned}\boldsymbol{P}_{\zeta\zeta,k} &= \boldsymbol{S}_{\zeta\zeta,k}\boldsymbol{S}_{\zeta\zeta,k}^{T}\\ &= \left[\, \boldsymbol{\mathcal{Z}}_{k|k-1} - \boldsymbol{A}_k\boldsymbol{\chi}_{k|k-1} \;\; \boldsymbol{S}_{R,k} \,\right]\\ &\quad \times \left[\, \boldsymbol{\mathcal{Z}}_{k|k-1} - \boldsymbol{A}_k\boldsymbol{\chi}_{k|k-1} \;\; \boldsymbol{S}_{R,k} \,\right]^{T} \end{aligned} \tag{B.8}$$

APPENDIX C. Derivation of (38) and (39)

We denote the lower triangular matrix in the right side of (40) as $\boldsymbol{W}$. The relation between the pre-array $\boldsymbol{V}$ and the post-array $\boldsymbol{W}$ is [28]:

$$\boldsymbol{W}\boldsymbol{W}^{T} = (\boldsymbol{V}\Theta)\cdot(\boldsymbol{V}\Theta)^{T} = \boldsymbol{V}(\Theta\Theta^{T})\boldsymbol{V}^{T} = \boldsymbol{V}\boldsymbol{V}^{T} \tag{C.1}$$

The entries $\{\, X \;\; Y \;\; Z \,\}$ in the post-array $\boldsymbol{W}$ can be identified by squaring both sizes of (40):

$$\begin{aligned}&\begin{bmatrix} \boldsymbol{S}_{\zeta\zeta,k} & L_k^{1/2}\boldsymbol{A}_k\boldsymbol{S}_{k|k-1} \\ \boldsymbol{0} & \boldsymbol{S}_{k|k-1} \end{bmatrix}(\Theta\Theta^{T})\begin{bmatrix} \boldsymbol{S}_{\zeta\zeta,k} & L_k^{1/2}\boldsymbol{A}_k\boldsymbol{S}_{k|k-1} \\ \boldsymbol{0} & \boldsymbol{S}_{k|k-1} \end{bmatrix}^{T}\\ &\quad = \begin{bmatrix} X & \boldsymbol{0} \\ Y & Z \end{bmatrix}\begin{bmatrix} X & \boldsymbol{0} \\ Y & Z \end{bmatrix}^{T} \end{aligned} \tag{C.2}$$

Here we give an alternate form of error covariance [33]:

$$\boldsymbol{P}_{k|k} = \boldsymbol{P}_{k|k-1} - \boldsymbol{P}_{k|k-1}\boldsymbol{A}_k^{T}\overline{\boldsymbol{R}}_{e,k}^{-1}\boldsymbol{A}_k\boldsymbol{P}_{k|k-1} \tag{C.3}$$

The following equalities can be established based on (C.2):

$$\begin{aligned} XX^{T} &= \boldsymbol{S}_{\zeta\zeta,k}\boldsymbol{S}_{\zeta\zeta,k}^{T} + \boldsymbol{A}_k\boldsymbol{S}_{k|k-1}\boldsymbol{S}_{k|k-1}^{T}L_k\boldsymbol{A}_k^{T}\\ &= \overline{\boldsymbol{R}}_{e,k}\\ &= \overline{\boldsymbol{R}}_{e,k}^{1/2}(\overline{\boldsymbol{R}}_{e,k}^{1/2})^{T} \end{aligned} \tag{C.4}$$

$$YX^{T} = L_k^{1/2}\boldsymbol{S}_{k|k-1}\boldsymbol{S}_{k|k-1}^{T}\boldsymbol{A}_k^{T} = L_k^{1/2}\boldsymbol{P}_{k|k-1}\boldsymbol{A}_k^{T} \tag{C.5}$$

$$\begin{aligned} ZZ^{T} &= \boldsymbol{S}_{k|k-1}\boldsymbol{S}_{k|k-1}^{T} - YY^{T}\\ &= \boldsymbol{P}_{k|k-1} - L_k^{1/2}\boldsymbol{S}_{k|k-1}\boldsymbol{S}_{k|k-1}^{T}\boldsymbol{A}_k^{T}\cdot\\ &\quad (X^{-T}X^{-1})(L_k^{1/2}\boldsymbol{S}_{k|k-1}\boldsymbol{S}_{k|k-1}^{T}\boldsymbol{A}_k^{T})^{T}\\ &= \boldsymbol{P}_{k|k-1} - \boldsymbol{P}_{k|k-1}\boldsymbol{A}_k^{T}\overline{\boldsymbol{R}}_{e,k}^{-1}\boldsymbol{A}_k\boldsymbol{P}_{k|k-1}\\ &= \boldsymbol{P}_{k|k} = \boldsymbol{S}_{k|k}\boldsymbol{S}_{k|k}^{T} \end{aligned} \tag{C.6}$$

X can be obtained by (C.3) with:

$$X = \overline{\boldsymbol{R}}_{e,k}^{1/2} \tag{C.7}$$

Z can be obtained by (C.5) with:

$$Z = \boldsymbol{S}_{k|k} \tag{C.8}$$

By combing (C.6) and (C.4), we have:

$$\begin{aligned} Y &= YX^{T}\cdot(X^{T})^{-1}\\ &= L_k^{1/2}\boldsymbol{P}_{k|k-1}\boldsymbol{A}_k^{T}\cdot(\overline{\boldsymbol{R}}_{e,k}^{T/2})^{-1}\\ &= \overline{\boldsymbol{K}}_k^{n} \end{aligned} \tag{C.9}$$

REFERENCES

[1] A. Martinelli, "Vision and IMU data fusion: closed-form solutions for attitude, speed, absolute scale, and bias determination," *IEEE Trans. Robot.*, vol. 28, no. 1, pp. 44–60, Feb. 2012.

[2] X. Guo, C. Sun, P. Wang, and L. Huang, "Vision sensor and dual MEMS gyroscope integrated system for attitude determination on moving base," *Rev. Sci. Instrum.*, vol. 89, no. 1, Jan. 2018, Art. no. 015002.

[3] J. Bai, J. Gao, Y. Lin, Z. Liu, S. Lian, and D. Liu, "A novel feedback mechanism-based stereo visual-inertial SLAM," *IEEE Access*, vol. 7, pp. 147721–147731, 2019.

[4] M. Alatise and G. Hancke, "Pose estimation of a mobile robot based on fusion of IMU data and vision data using an extended Kalman filter," *Sensors*, vol. 17, no. 10, p. 2164, 2017.

[5] R. V. Garcia, P. C. P. M. Pardal, H. K. Kuga, and M. C. Zanardi, "Nonlinear filtering for sequential spacecraft attitude estimation with real data: Cubature Kalman filter, unscented Kalman filter and extended Kalman filter," *Adv. Space Res.*, vol. 63, no. 2, pp. 1038–1050, Jan. 2019.

[6] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1254–1269, Jun. 2009.

[7] C. Karlgaard and H. Schaub, "Comparison of several nonlinear filters for a benchmark tracking problem," in *Proc. AIAA Guid., Navigat., Control Conf. Exhibit*, Aug. 2006, pp. 6243–6259.

[8] F. Gustafsson, "Particle filter theory and practice with positioning applications," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 53–82, Jul. 2010.

[9] S. Sarkka and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *IEEE Trans. Autom. Control*, vol. 54, no. 3, pp. 596–600, Mar. 2009.

[10] R. Piche, S. Sarkka, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate student-t distribution," in *Proc. IEEE Int. Workshop Mach. Learn. for Signal Process.*, Sep. 2012, pp. 1–6.

[11] Y. Huang and Y. Zhang, "Robust student's t-based stochastic cubature filter for nonlinear systems with heavy-tailed process and measurement noises," *IEEE Access*, vol. 5, pp. 7964–7974, 2017.

[12] X. Wang, N. Cui, and J. Guo, "Huber-based unscented filtering and its application to vision-based relative navigation," *IET Radar, Sonar Navigat.*, vol. 4, no. 1, pp. 134–141, 2010.

[13] K. Li, B. Hu, L. Chang, and Y. Li, "Robust square-root cubature Kalman filter based on Huber's M-estimation methodology," *Proceedings Inst. Mech. Eng., G, J. Aerosp. Eng.*, vol. 229, no. 7, pp. 1236–1245, Aug. 2014.

[14] Y. Liu, H. Wang, and C. Hou, "UKF based nonlinear filtering using minimum entropy criterion," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 4988–4999, Oct. 2013.

[15] X. Liu, H. Qu, J. Zhao, P. Yue, and M. Wang, "Maximum correntropy unscented Kalman filter for spacecraft relative state estimation," *Sensors*, vol. 16, no. 9, p. 1530, 2016.

[16] G. Wang, Y. Zhang, and X. Wang, "Iterated maximum correntropy unscented Kalman filters for non-Gaussian systems," *Signal Process.*, vol. 163, pp. 87–94, Oct. 2019.

[17] X. Liu, B. Chen, H. Zhao, J. Qin, and J. Cao, "Maximum correntropy Kalman filter with state constraints," *IEEE Access*, vol. 5, pp. 25846–25853, 2017.

[18] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, Feb. 2017.

[19] X. Liu, B. Chen, B. Xu, Z. Wu, and P. Honeine, "Maximum correntropy unscented filter," *Int. J. Syst. Sci.*, vol. 48, no. 8, pp. 1607–1615, Jun. 2017.

[20] X. Liu, H. Qu, J. Zhao, and P. Yue, "Maximum correntropy square-root cubature Kalman filter with application to SINS/GPS integrated systems," *ISA Trans.*, vol. 80, pp. 195–202, Sep. 2018.

[21] X. Tang, X. Zhao, and X. Zhang, "The square-root spherical simplex unscented Kalman filter for state and parameter estimation," in *Proc. 9th Int. Conf. Signal Process.*, Salt Lake City, UT, USA, Oct. 2008, pp. 3461–3464.

[22] I. Arasaratnam and S. Haykin, "Square-root quadrature Kalman filtering," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2589–2593, Jun. 2008.

[23] B. Hou, Z. He, X. Zhou, and D. Li, "Maximum correntropy criterion Kalman filter for a-Jerk tracking model with non-Gaussian noise," *Entropy*, vol. 19, no. 12, p. 648, Nov. 2017.

[24] R. Izanloo, S. A. Fakoorian, H. S. Yazdi, and D. Simon, "Kalman filtering based on the maximum correntropy criterion in the presence of non-Gaussian noise," in *Proc. Annu. Conf. Inf. Sci. Syst. (CISS)*, Princeton, NJ, USA, Mar. 2016, pp. 500–505.

[25] M. V. Kulikova, "Square-root algorithms for maximum correntropy estimation of linear discrete-time systems in presence of non-Gaussian noise," *Syst. Control Lett.*, vol. 108, pp. 8–15, Oct. 2017.

[26] G. Wang, N. Li, and Y. Zhang, "Maximum correntropy unscented Kalman and information filters for non-Gaussian measurement noise," *J. Franklin Inst.*, vol. 354, no. 18, pp. 8659–8677, Dec. 2017.

[27] T. Lefebvre, H. Bruyninckx, and J. De Schuller, "Comment on 'A new method for the nonlinear transformation of means and covariances in filters and estimators' [with authors' reply]," *IEEE Trans. Autom. Control*, vol. 47, no. 8, pp. 1406–1409, Aug. 2002.

[28] T. Kailath, A. H. Sayed, and B. Hassibi, "Array algorithms," in *Linear Estimation*, 1nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2000, pp. 427–437.

[29] E. Foxlin, "Inertial head-tracker sensor fusion by a complementary separate-bias Kalman filter," in *Proc. IEEE Virtual Reality Annu. Int. Symp.*, Mar. 1996, pp. 185–194.

[30] X. Guo, J. Tang, J. Li, C. Shen, and J. Liu, "Attitude measurement based on imaging ray tracking model and orthographic projection with iteration algorithm," *ISA Trans.*, vol. 95, pp. 379–391, Dec. 2019.

[31] C. Sun, H. Dong, B. Zhang, and P. Wang, "An orthogonal iteration pose estimation algorithm based on an incident ray tracking model," *Meas. Sci. Technol.*, vol. 29, no. 9, Sep. 2018, Art. no. 095402.

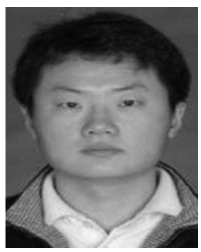
[32] H. Dong, C. Sun, B. Zhang, and P. Wang, "Simultaneous pose and correspondence determination combining softassign and orthogonal iteration," *IEEE Access*, vol. 7, pp. 137720–137730, Sep. 2019.

[33] D. Simon, "Least Squares estimation," in *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches*, 1nd ed., Hoboken, NJ, USA: Wiley, 2006, pp. 86–88.

JINGJING HE received the B.S. degree from the School of Precision Instrument and Opto-Electronics Engineering, Tianjin University, in 2017. She is currently pursuing the Ph.D. degree with the State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University. Her current research interests include inertial sensors and visual and inertial fusion pose measurement.

CHANGKU SUN received the M.S. degree from the Harbin Institute of Technology, in 1990, and the Ph.D. degree from the Saint Petersburg Precision Mechanics and Optics Institute, Russia, in 1994. He is currently a Professor with the State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University. His research interests include machine vision, optical measurement technology, inertial measurement technology, and visual and inertial fusion pose measurement.

BAOSHANG ZHANG received the B.S. and M.S. degrees from Northwestern Polytechnical University, in 2006 and 2009, respectively. He is currently an Engineer with the Science and Technology on Electro-Optic Control Laboratory, Luoyang Institute of Electro-Optical Equipment. His current research interests include visual localization and pose estimation.



PENG WANG received the B.S. and Ph.D. degrees from the Department of Precision Instruments Engineering, Tianjin University, China, in 2004 and 2008, respectively. He is currently an Associate Professor with the State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University. His research interests include machine vision, optical measurement technology, inertial measurement technology, and visual and inertial fusion pose measurement.