
Proceedings Paper:
Liu, Y., Li, X., Xue, Y. et al. (2020) Outlier-robust Schmidt-Kalman filter using variational inference. In: Proceedings of 2020 IEEE 23rd International Conference on Information Fusion (FUSION). 2020 IEEE 23rd International Conference on Information Fusion (FUSION), 06-09 Jul 2020, Rustenburg, South Africa. IEEE, pp. 1-8. ISBN: 9781728168302.

https://doi.org/10.23919/FUSION45008.2020.9190507

Outlier-Robust Schmidt-Kalman Filter Using Variational Inference

Yi Liu[1], Xi Li[2], Yanbo Xue[3], Stephen Weddell[1], Le Yang[1*], Lyudmila Mihaylova[4]
1. Department of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand
2. School of Electronic Science and Engineering, National University of Defense Technology, Changsha, China
3. Career Science Lab, Kanzhun Technology, Beijing, China
4. Department of Automatic Control and System Engineering, University of Sheffield, Sheffield, UK
*: Corresponding author, le.yang@canterbury.ac.nz.

***Abstract*—The Schmidt-Kalman filter (SKF) achieves filtering consistency in the presence of biases in system dynamic and measurement models through accounting for their impacts when updating the state estimate and covariance. However, the performance of the SKF may break down when the measurements are subject to non-Gaussian and heavy-tail noise. To address this, we impose the Wishart prior distribution on the precision matrix of measurement noise, such that the measurement likelihood now has heavier tails than the Gaussian distribution to deal with the potential occurrence of outliers. Variational inference is invoked to establish analytically tractable methods for computing the posterior of the system state, system biases, and the measurement noise precision matrix. The principle ofn the SKF considers the effect of system biases but does not actively estimate them when two variants of outlier-robust SKFs are incorporated. We evaluate their performance in terms of estimation accuracy and filtering consistency using simulations and real-world data. Promising results are obtained.**

I. Introduction

For linear Gaussian state-space models, it is well known that the Kalman filter (KF) is the minimum mean square error (MMSE) state estimator [1]. Furthermore, the KF-calculated state covariance is equal to the true covariance of the state estimate when both the system dynamic and measurement models are accurate [2]. In practical state estimation problems, however, state dynamics and measurements may be subject to the presence of systematic errors such as sensor biases and imperfect knowledge on sensor locations and attitudes [3], [4]. Calibration can reduce the systematic errors but there may still be *residual* biases. The remaining system biases, if improperly handled, would make the KF inconsistent such that the state covariance computed by the KF would be smaller than, in the matrix positive-semidefinte sense, the true state covariance [5]. In other words, the KF becomes overly confident in this case, which degrades the tracking accuracy and increases the chance of filtering divergence [3], [4], [6].

One possible way to account for the presence of system biases and maintain filtering consistency is to inflate the process and measurement noise covariances [7]. This approach, despite its simplicity, may provide unsatisfactory estimation accuracy because it does not consider the cross correlation between system biases and state estimate. An alternative technique is commonly referred to as the consider analysis [8]. It first computes a delta covariance using the statistic information on the system biases and their cross correlation with the state estimate and then adds it to the state estimation covariance. In [9], [10], augmented-state KFs were adopted to estimate the system state and biases simultaneously in order to attain filtering consistency when the system biases are present. Nevertheless, in many cases, estimating the system biases is not necessary or they may have low observability, which could make jointly identifying system biases and state more prone to divergence.

To address the aforementioned drawback, the Schmidt-Kalman filter (SKF) that was originally proposed in [11] and derived in details in [2] can be employed. The SKF, also called the consider filter, still augments the system state with the bias terms but it propagates only the state estimate, its estimation covariance, and its cross correlation with the biases. The bias terms and their associated covariance are uncorrected during the filtering process. In this way, the SKF achieves filtering consistency by *considering* the contribution of system biases when updating the state estimate and its covariance.

The SKF has attracted a significant amount of attention due to its wide applications in e.g., orbit determination [12], Mars entry navigation [13] and source geolocation when satellite ephemeris errors are present [14]. In [15], it was established theoretically that the SKF is an unbiased state estimator under linear Gaussian state-space models. Furthermore, it has the smallest state estimation variance among all the unbiased state estimators that do not compensate system biases. In [16], an alternative derivation of the SKF was given and the difference between the SKF and consider analysis was discussed. The information-domain equivalent of the SKF, referred to as the inverse Schmidt estimator, was developed in [5]. In [17], the SKF was established from the Bayesian filtering perspective and the interacting multiple model-SKF (IMM-SKF) was proposed in [3] to track maneuvering objects. To handle the presence of nonlinearity in the system dynamics and measurement models, the Gauss-Hermite quadrature rule and unscented transformation were introduced into the SKF framework. As a result, the quadrature and unscented SKFs were established (see [18]–[20]). More recently, the polynomial chaos expansion method was integrated with the SKF to deal with system nonlinearity and non-Gaussian bias terms [21].

Most of the studies surveyed above on the SKF assumed that the measurements are subject to Gaussian noises. This renders existing SKF algorithms sensitive to measurement outliers, which could come from e.g., non-line-of-sight (NLOS) signal propagation [22], transient disturbance, and sensor failure [23]. In this paper, we develop new SKFs that are robust to the presence of outlying measurements. In particular, we shall consider the more general scenario where the measurement vector at every sampling instant can be partitioned into multiple subvectors with *independent* noises. For each measurement subvector, we impose a Wishart prior distribution on its noise precision matrix, which is equal to the inverse of the noise covariance, such that the measurement likelihood now has heavier tails than the Gaussian distribution. To find an approximation to the analytically intractable joint posterior of the system state, biases, and measurement noise covariances, variational inference similar to that used in [24], [25] is employed. The principle underlying the SKF that we consider, i.e., the effect of system biases that do not actively estimate them, is incorporated. Two variants of outlier-robust SKFs (ORSKFs) are established and their performance are evaluated using simulations and real-world data.

It is worthwhile to point out that the idea of imposing a prior distribution on the measurement noise covariance to achieve robustness to outliers in filtering techniques is not new (see e.g., [24]–[29] and references therein). This work focuses on robustifying the KF instead of the SKF, and as a result, the presence of system biases was not considered. An interesting topic for future research would be to investigate the integration of variational inference techniques in [26]–[29] into the SKF framework to develop other ORSKF algorithms.

The remainder of this paper is organized as follows. Section II formulates the linear state estimation problem with system biases and briefly presents the standard SKF. Section III develops two new ORSKFs using variational inference. Section IV provides the experimental results obtained using simulations and real-world data. The conclusions are given in Section V.

II. Problem Formulation

Consider the following linear state-space model

$$\begin{bmatrix}\mathbf{x}_k\\ \mathbf{b}\end{bmatrix} = \underbrace{\begin{bmatrix}\mathbf{F}_x & \mathbf{O}\\ \mathbf{O} & \mathbf{I}\end{bmatrix}}_{\mathbf{F}}\begin{bmatrix}\mathbf{x}_{k-1}\\ \mathbf{b}\end{bmatrix} + \begin{bmatrix}\mathbf{v}_k\\ \mathbf{0}\end{bmatrix} \tag{1a}$$

$$\mathbf{y}_k = \underbrace{\begin{bmatrix}\mathbf{H}_x & \mathbf{H}_b\end{bmatrix}}_{\mathbf{H}}\begin{bmatrix}\mathbf{x}_k\\ \mathbf{b}\end{bmatrix} + \mathbf{w}_k \tag{1b}$$

where $\mathbf{x}_k \in \mathcal{R}^{n_x\times 1}$ is the state of interest at the kth sampling instant, $\mathbf{b} \in \mathcal{R}^{n_b\times 1}$ is the system bias vector, $\mathbf{F}_x$ is the state transition matrix, and $\mathbf{H}_x$ and $\mathbf{H}_b$ are the measurement matrices for the state vector $\mathbf{x}_k$ and bias vector $\mathbf{b}$.

In (1), $\mathbf{v}_k$ is the process noise vector for $\mathbf{x}_k$, which is white Gaussian with zero mean and covariance $\mathbf{Q}_v$ (i.e., $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_\mathbf{v})$). Besides, we assume zero process noise for $\mathbf{b}$ because the biases are often strongly correlated over time and they will be modeled using a Gaussian random vector as in [3], [4], [15]. Mathematically, we have $\mathbf{b} \sim \mathcal{N}(0, \mathbf{B})$. To simplify the theoretical development, we consider in this work the scenario where system biases are present only in the measurement model (see (1b)). Incorporating non-zero process noise for the measurement bias terms in $\mathbf{b}$ and accounting for possible biases in the state dynamic model in (1a) as in [15], [17] would just require some straightforward modifications of the proposed ORSKFs, which will be omitted here for brevity.

We further assume that the measurement vector $\mathbf{y}_k \in \mathcal{R}^{n_y\times 1}$ can be divided into M subvectors as, with slight relaxation of notations,

$$\mathbf{y}_k = \left[\mathbf{y}_{1,k}^T, \mathbf{y}_{2,k}^T, ..., \mathbf{y}_{M,k}^T\right]^T. \tag{2}$$

As a result, the measurement noise vector $\mathbf{w}_k$, $\mathbf{H}_x$ and $\mathbf{H}_b$ can be correspondingly expressed in block form as

$$\mathbf{w}_k = \left[\mathbf{w}_{1,k}^T, \mathbf{w}_{2,k}^T, ..., \mathbf{w}_{M,k}^T\right]^T \tag{3a}$$

$$\mathbf{H}_x = \left[\mathbf{H}_{1,x}^T, \mathbf{H}_{2,x}^T, ..., \mathbf{H}_{M,x}^T\right]^T \tag{3b}$$

$$\mathbf{H}_b = \left[\mathbf{H}_{1,b}^T, \mathbf{H}_{2,b}^T, ..., \mathbf{H}_{M,b}^T\right]^T \tag{3c}$$

such that for $i = 1, 2, ..., M$ and,

$$\mathbf{y}_{i,k} = \mathbf{H}_{i,x}\mathbf{x}_k + \mathbf{H}_{i,b}\mathbf{b} + \mathbf{w}_{i,k}. \tag{4}$$

The noise subvectors $\mathbf{w}_{i,k}$ are white and *independent* to one another. A practical example for motivating the above formulation is the fusion of the time difference of arrivals (TDOAs) and their time derivative, frequency difference of arrivals (FDOAs) for localization applications. In this case, the measurements are TDOAs and FDOAs subject to independent random noises when they are estimated using the maximum likelihood (ML) method such as the one from [30]. Outlying measurements can occur due to e.g., temporary time and frequency offsets [31] but they are still independent, since they are from different sources of errors.

If the biases in $\mathbf{b}$ are ignored and the KF is applied directly to estimate $\mathbf{x}_k$ from the obtained measurements $\mathbf{y}_k$, the KF would be inconsistent and the state covariance calculated by the KF would be smaller than the true one. The standard SKF provides consistent state covariance by considering the impact of $\mathbf{b}$ without actually identifying it [3], [4], [18].

Specifically, suppose the measurement noise vector $\mathbf{w}_k$ has a Gaussian distribution with zero mean and covariance $\mathbf{Q}_w$ (i.e., $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_w)$). Let the joint posterior of $\mathbf{x}_k$ and $\mathbf{b}$ at the $(k-1)$th sampling instant be

$$\begin{bmatrix}\mathbf{x}_{k-1}\\ \mathbf{b}\end{bmatrix} \sim \mathcal{N}\left(\mathbf{m}_{k-1}, \mathbf{\Sigma}_{k-1}\right) \tag{5}$$

where $\mathbf{m}_{k-1}$ and $\mathbf{\Sigma}_{k-1}$ are defined as

$$\mathbf{m}_{k-1} = \begin{bmatrix}\boldsymbol{\mu}_{k-1}\\ \mathbf{0}\end{bmatrix},\ \mathbf{\Sigma}_{k-1} = \begin{bmatrix}\mathbf{P}_{k-1} & \mathbf{C}_{k-1}\\ \mathbf{C}_{k-1}^T & \mathbf{B}\end{bmatrix}. \tag{6}$$

$\boldsymbol{\mu}_{k-1}$ is the posterior mean of the estimate of the state vector $\mathbf{x}_{k-1}$ and $\mathbf{P}_{k-1}$ is its posterior covariance. $\mathbf{C}_{k-1}$ is the cross covariance between the state estimate and bias terms in $\mathbf{b}$.

At the kth sampling instant, the standard SKF carries out the following prediction and update steps in sequence:

1) Prediction: These steps are straightforward. From (1a) and (1b), the predicted mean and covariance for the vector combining the current state vector and biases, $[\mathbf{x}_k^T, \mathbf{b}^T]^T$, are

$$\mathbf{m}_{k|k-1} = \mathbf{F}\mathbf{m}_{k-1} \tag{7a}$$

$$\mathbf{\Sigma}_{k|k-1} = \mathbf{F}\mathbf{\Sigma}_{k-1}\mathbf{F}^T + \begin{bmatrix} \mathbf{Q}_v & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \tag{7b}$$

where the subscript $_{k|k-1}$ denotes prediction from sampling instant $k-1$ to the current instant k.

2) Update: The SKF updates the state estimate through evaluating

$$\mathbf{S}_k = \mathbf{H}\mathbf{\Sigma}_{k|k-1}\mathbf{H}^T + \mathbf{Q}_w \tag{8a}$$

$$\mathbf{K}_k = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{\Sigma}_{k|k-1}\mathbf{H}^T\mathbf{S}_k^{-1} \tag{8b}$$

$$\mathbf{m}_k = \begin{bmatrix} \boldsymbol{\mu}_k \\ \mathbf{0} \end{bmatrix} = \mathbf{m}_{k|k-1} + \mathbf{K}_k \cdot (\mathbf{y}_k - \mathbf{H}\mathbf{m}_{k|k-1}) \tag{8c}$$

$$\mathbf{\Sigma}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{\Sigma}_{k|k-1}(\mathbf{I} - \mathbf{K}_k\mathbf{H})^T + \mathbf{K}_k\mathbf{Q}_w\mathbf{K}_k^T. \tag{8d}$$

We can see from (6), (7b), (8a) and (8b) that the standard SKF *considers* the impact of biases via exploring their prior covariance $\mathbf{B}$ and the cross covariance $\mathbf{C}_{k-1}$ when computing the innovation covariance $\mathbf{S}_k$ and gain $\mathbf{K}_k$ for state update. Furthermore, according to (8b), the SKF sets the bias component of the gain matrix $\mathbf{K}_k$ to be zero. As a result, the bias terms are not actively estimated, they are still zero-mean (see (8c)) and their posterior covariance given in (8d) can be shown to be equal to $\mathbf{B}$, which is their prior covariance [18], [19], [32].

The performance of the standard SKF presented above depends heavily on the validity of the assumption that measurement noises are Gaussian. When outlying measurements are present, the SKF may no longer be the minimum–variance and consistent state estimator among all the unbiased state estimation techniques. We shall develop in the next section new ORSKFs that are robust to outliers.

III. Outlier-Robust Schmidt-Kalman Filtering

A. Heavy-tail measurement noise model

To cope with non-Gaussian and heavy-tail measurement noises within the SKF framework, we shall take an approach similar to the one proposed in [24], [25]. In particular, it is assumed that the noises $\mathbf{w}_{i,k}$ in the measurement subvectors $\mathbf{y}_{i,k}$ are zero-mean Gaussian distributed with covariance $\mathbf{Q}_{w_i}$, whose inverse, known as the precision matrix, is sampled from an independent and white Wishart prior distribution. Mathematically, let $\mathbf{\Lambda}_i = \mathbf{Q}_{w_i}^{-1}$ and we have that

$$\mathbf{\Lambda}_i \sim \mathcal{W}\left(\frac{\mathbf{R}_i^{-1}}{\nu_i}, \nu_i\right) \propto |\mathbf{\Lambda}_i|^{\frac{\nu_i - p_i - 1}{2}} \exp(-\frac{\nu_i}{2}\mathrm{tr}(\mathbf{R}_i\mathbf{\Lambda}_i)) \tag{9}$$

where ν_i and $\mathbf{R}_i^{-1}/\nu_i$ represent the degrees of freedom and scale matrix of this conjugate prior distribution [33], [34]. p_i is the dimensionality of the measurement subvector $\mathbf{y}_{i,k}$ such that $\sum_{i=1}^M p_i = n_y$.

It can be shown by carrying out the following integral and applying the definition of the Wishart distribution in (9) that the noise vector $\mathbf{w}_{i,k}$ would now have a distribution given by

$$\begin{aligned} p(\mathbf{w}_{i,k}) &= \int \mathcal{N}(\mathbf{0}, \mathbf{\Lambda}_i^{-1})\mathcal{W}\left(\frac{\mathbf{R}_i^{-1}}{\nu_i}, \nu_i\right) \mathrm{d}\mathbf{\Lambda}_i \\ &\propto \left(1 + \frac{1}{\nu_i}\mathbf{w}_{i,k}^T\mathbf{R}_i^{-1}\mathbf{w}_{i,k}\right)^{-\frac{\nu_i+1}{2}}. \end{aligned} \tag{10}$$

We have from (10) that if $\mathbf{y}_{i,k}$ has more than one element ($p_i > 1$), $p(\mathbf{w}_{i,k})$ would have longer tails than the Student's t-distribution with the degrees of freedom ν_i and shape matrix $\mathbf{R}_i$, which is proportional to $\left(1 + \frac{1}{\nu_i}\mathbf{w}_{i,k}^T\mathbf{R}_i^{-1}\mathbf{w}_{i,k}\right)^{-\frac{\nu_i+p_i}{2}}$. It is well-known that compared with the Gaussian distribution, the Student's t-distribution has heavier tails. Therefore, by imposing Wishart priors on the noise vectors $\mathbf{w}_{i,k}$, we obtain noise distributions with generally much heavier tails than the Gaussian to handle the presence of outlying measurements.

B. Variational Approximation

As the noise precision matrices $\mathbf{\Lambda}_i$ are not fixed and they are latent variables, estimating the state $\mathbf{x}_k$ from the measurements $\mathbf{y}_k$ using the state-space model in (1) requires computing the joint posterior of $\mathbf{x}_k$, the bias vector $\mathbf{b}$, and $\mathbf{\Lambda}_i$. This is analytically intractable despite that the state-space model is linear, because the noise distribution is no longer Gaussian (see (10)). We resort to the technique of variational inference and where appropriate, incorporating the idea of the SKF that the effect of biases is considered but they are not estimated to establish two new ORSKFs.

The algorithm development again starts with assuming that at the $(k-1)$th sampling instant, the joint posterior of the state vector $\mathbf{x}_{k-1}$, and bias terms $\mathbf{b}$ is Gaussian, which is given in (5). Let us denote the *approximate* posterior of $\mathbf{x}_k$, $\mathbf{b}$, and $\mathbf{\Lambda}_i$ at the current sampling instant k as $q(\mathbf{x}_k, \mathbf{b}, \mathbf{\Lambda}_1, \mathbf{\Lambda}_2, ..., \mathbf{\Lambda}_M)$. Applying the mean-field approximation [33] factorizes it into

$$q(\mathbf{x}_k, \mathbf{b}, \mathbf{\Lambda}_1, \mathbf{\Lambda}_2, ..., \mathbf{\Lambda}_M) = q(\mathbf{x}_k, \mathbf{b}) \prod_{i=1}^M q(\mathbf{\Lambda}_i). \tag{11}$$

To find the approximate posterior in (11), we first note that

$$\begin{aligned} &p(\mathbf{y}_k, \mathbf{x}_k, \mathbf{b}, \mathbf{\Lambda}_1, \mathbf{\Lambda}_2, ..., \mathbf{\Lambda}_M | \mathbf{y}_{1:k-1}) \\ &= p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{b}, \mathbf{\Lambda}_1, \mathbf{\Lambda}_2, ..., \mathbf{\Lambda}_M) \prod_{i=1}^M p(\mathbf{\Lambda}_i) p(\mathbf{x}_k, \mathbf{b} | \mathbf{y}_{1:k-1}) \end{aligned} \tag{12}$$

where the Markovian property of the state-space model in (1), and the fact that the noise precision matrices $\mathbf{\Lambda}_i$, have independent and white priors have been applied. Besides, $\mathbf{y}_{1:k-1} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{k-1}\}$ stands for the set of measurements collected up to sampling instant $k-1$. Utilizing that the measurement subvectors $\mathbf{y}_{i,k}$ are subject to independent noise $\mathbf{w}_{i,k}$ further transforms (12) into

$$\begin{aligned} &p(\mathbf{y}_k, \mathbf{x}_k, \mathbf{b}, \mathbf{\Lambda}_1, \mathbf{\Lambda}_2, ..., \mathbf{\Lambda}_M | \mathbf{y}_{1:k-1}) \\ &= \prod_{i=1}^M p(\mathbf{y}_{i,k} | \mathbf{x}_k, \mathbf{b}, \mathbf{\Lambda}_i) p(\mathbf{\Lambda}_i) \cdot p(\mathbf{x}_k, \mathbf{b} | \mathbf{y}_{1:k-1}). \end{aligned} \tag{13}$$

Marginalizing out the state vector $\mathbf{x}_k$, biases in $\mathbf{b}$ and precision matrices $\mathbf{\Lambda}_i$ on both sides of (13) yields the conditional measurement likelihood $p(\mathbf{y}_k|\mathbf{y}_{1:k-1})$. Taking the logarithm of the result, using (11) and applying the Jensen's inequality [33], we arrive at

$$\begin{aligned}\log p(\mathbf{y}_k|\mathbf{y}_{1:k-1}) \geq \int q(\mathbf{x}_k,\mathbf{b})\prod_{i=1}^{M} q(\mathbf{\Lambda}_i) \\ \times \log\frac{\prod_{i=1}^{M} p(\mathbf{y}_{i,k}|\mathbf{x}_k,\mathbf{b},\mathbf{\Lambda}_i)p(\mathbf{\Lambda}_i)p(\mathbf{x}_k,\mathbf{b}|\mathbf{y}_{1:k-1})}{q(\mathbf{x}_k,\mathbf{b})\prod_{i=1}^{M} q(\mathbf{\Lambda}_i)} \\ \mathrm{d}\mathbf{x}_k\mathrm{d}\mathbf{b}\ \mathrm{d}\mathbf{\Lambda}_1\mathrm{d}\mathbf{\Lambda}_2\cdots\mathrm{d}\mathbf{\Lambda}_M.\end{aligned} \tag{14}$$

The term on the right hand side of (14) is often referred to as the evidence lower bound (ELBO) [33], [34].

We shall attempt to maximize the ELBO to find the desired approximate posterior in (11) to develop the desired outlier-robust SKFs. For this purpose, the fixed-point optimization is adopted. Specifically, we shall first derive the approximate posterior of the precision matrices $q(\mathbf{\Lambda}_i)$ under the condition that $q(\mathbf{x}_k,\mathbf{b})$ is given and then proceed to find $q(\mathbf{x}_k,\mathbf{b})$ with $q(\mathbf{\Lambda}_i)$ being fixed.

C. Outlier-robust SKFs

Carrying out the integral in the ELBO and ignoring all the terms not related to the precision matrix $\mathbf{\Lambda}_i$ yields that to maximize the resulting ELBO, $q(\mathbf{\Lambda}_i)$ has the following functional form

$$q(\mathbf{\Lambda}_i) \propto \exp\left(\int q(\mathbf{x}_k,\mathbf{b})\log p(\mathbf{y}_{i,k}|\mathbf{x}_k,\mathbf{b},\mathbf{\Lambda}_i)p(\mathbf{\Lambda}_i)\mathrm{d}\mathbf{x}_k\mathrm{d}\mathbf{b}\right). \tag{15}$$

From (4) and the discussions above (9), we have that $p(\mathbf{y}_{i,k}|\mathbf{x}_k,\mathbf{b},\mathbf{\Lambda}_i)$ follows a Gaussian distribution given by

$$p(\mathbf{y}_{i,k}|\mathbf{x}_k,\mathbf{b},\mathbf{\Lambda}_i) \propto |\mathbf{\Lambda}_i|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{\delta}_i^T\mathbf{\Lambda}_i\boldsymbol{\delta}_i\right) \tag{16}$$

where

$$\boldsymbol{\delta}_i = \mathbf{y}_{i,k} - \mathbf{H}_{i,x}\mathbf{x}_k - \mathbf{H}_{i,b}\mathbf{b}. \tag{17}$$

Putting (16) and (9) into (15) gives

$$q(\mathbf{\Lambda}_i) \propto |\mathbf{\Lambda}_i|^{\frac{\nu_i-p_i}{2}}\exp\left(-\frac{1}{2}\mathrm{tr}\left((\mathbf{R}_i\nu_i + \langle\boldsymbol{\delta}_i\boldsymbol{\delta}_i^T\rangle)\mathbf{\Lambda}_i\right)\right) \tag{18}$$

where $\langle\boldsymbol{\delta}_i\boldsymbol{\delta}_i^T\rangle$ is equal to

$$\langle\boldsymbol{\delta}_i\boldsymbol{\delta}_i^T\rangle = \int q(\mathbf{x}_k,\mathbf{b})\boldsymbol{\delta}_i\boldsymbol{\delta}_i^T\mathrm{d}\mathbf{x}_k\mathrm{d}\mathbf{b}. \tag{19}$$

It can be seen from (18) and (9) that the approximate posterior of the precision matrix $\mathbf{\Lambda}_i$ is still a Wishart distribution but with the degrees of freedom ν_i+1 and scale matrix $(\mathbf{R}_i\nu_i + \langle\boldsymbol{\delta}_i\boldsymbol{\delta}_i^T\rangle)^{-1}$. That is, we have that

$$q(\mathbf{\Lambda}_i) = \mathcal{W}\left((\mathbf{R}_i\nu_i + \langle\boldsymbol{\delta}_i\boldsymbol{\delta}_i^T\rangle)^{-1},\ \nu_i+1\right). \tag{20}$$

We next find the approximate posterior of the state vector $\mathbf{x}_k$ and biases in $\mathbf{b}$, $q(\mathbf{x}_k,\mathbf{b})$, using the obtained posterior of the precision matrices $\mathbf{\Lambda}_i$, $q(\mathbf{\Lambda}_i)$, given in (20). Again, evaluating the integral in (14) and discarding the terms that are not dependent on $\mathbf{x}_k$ or $\mathbf{b}$, simplify the ELBO into

$$\int q(\mathbf{x}_k,\mathbf{b})\log\frac{\exp\left(\sum_{i=1}^{M}\gamma_{i,k} + \log p(\mathbf{x}_k,\mathbf{b}|\mathbf{y}_{1:k-1})\right)}{q(\mathbf{x}_k,\mathbf{b})}\mathrm{d}\mathbf{x}_k\mathrm{d}\mathbf{b} \tag{21}$$

where $\gamma_{i,k}$ is equal to, after applying (16),

$$\begin{aligned}\gamma_{i,k} &= \int q(\mathbf{\Lambda}_i)\log p(\mathbf{y}_{i,k}|\mathbf{x}_k,\mathbf{b},\mathbf{\Lambda}_i)\mathrm{d}\mathbf{\Lambda}_i \\ &\propto -\frac{1}{2}\boldsymbol{\delta}_i^T\langle\mathbf{\Lambda}_i\rangle\boldsymbol{\delta}_i\end{aligned} \tag{22}$$

and according to (20) and [33], [34],

$$\langle\mathbf{\Lambda}_i\rangle = \left(\frac{\mathbf{R}_i\nu_i + \langle\boldsymbol{\delta}_i\boldsymbol{\delta}_i^T\rangle}{\nu_i+1}\right)^{-1}. \tag{23}$$

Substituting (22) back into (21), using (16) and ignoring the constant terms, we can rewrite the ELBO in (21) as

$$\int q(\mathbf{x}_k,\mathbf{b})\log\frac{\prod_{i=1}^{M} p(\mathbf{y}_{i,k}|\mathbf{x}_k,\mathbf{b},\langle\mathbf{\Lambda}_i\rangle)p(\mathbf{x}_k,\mathbf{b}|\mathbf{y}_{1:k-1})}{q(\mathbf{x}_k,\mathbf{b})}\mathrm{d}\mathbf{x}_k\mathrm{d}\mathbf{b} \tag{24}$$

and it is proportional to

$$-\mathrm{KLD}(q(\mathbf{x}_k,\mathbf{b})||p(\mathbf{x}_k,\mathbf{b}|\mathbf{y}_{1:k},\langle\mathbf{\Lambda}_1\rangle,\langle\mathbf{\Lambda}_2\rangle,...,\langle\mathbf{\Lambda}_M\rangle)). \tag{25}$$

Here, $\mathrm{KLD}(q||p)$ denotes the Kullback-Leibler divergence (KLD) [35] between two distributions, $q(\cdot)$ and $p(\cdot)$. $p(\mathbf{x}_k,\mathbf{b}|\mathbf{y}_{1:k},\langle\mathbf{\Lambda}_1\rangle,\langle\mathbf{\Lambda}_2\rangle,...,\langle\mathbf{\Lambda}_M\rangle)$ is the posterior of the state vector $\mathbf{x}_k$ and biases $\mathbf{b}$ given that the measurement noises $\mathbf{w}_{i,k}$ are independent Gaussian vectors with zero mean and covariances $\langle\mathbf{\Lambda}_i\rangle^{-1}$, $i=1,2,...,M$. Under the linear state-space model (see (1)) and noting that the posterior of the state vector at the previous sampling instant $\mathbf{x}_{k-1}$ and bias terms $\mathbf{b}$ is Gaussian (see (5)), we have that

$$p(\mathbf{x}_k,\mathbf{b}|\mathbf{y}_{1:k},\langle\mathbf{\Lambda}_1\rangle,\langle\mathbf{\Lambda}_2\rangle,...,\langle\mathbf{\Lambda}_M\rangle) = \mathcal{N}(\bar{\mathbf{m}}_k,\bar{\mathbf{\Sigma}}_k) \tag{26}$$

and where the postorior mean and covariance, $\bar{\mathbf{m}}_k$ and $\bar{\mathbf{\Sigma}}_k$, are found via evaluating

$$\bar{\mathbf{S}}_k = \mathbf{H}\mathbf{\Sigma}_{k|k-1}\mathbf{H}^T + \bar{\mathbf{Q}}_w \tag{27a}$$

$$\bar{\mathbf{K}}_k = \mathbf{\Sigma}_{k|k-1}\mathbf{H}^T\bar{\mathbf{S}}_k^{-1} \tag{27b}$$

$$\bar{\mathbf{m}}_k = \begin{bmatrix}\bar{\boldsymbol{\mu}}_k \\ \bar{\boldsymbol{\mu}}_k^b\end{bmatrix} = \mathbf{m}_{k|k-1} + \bar{\mathbf{K}}_k\cdot(\mathbf{y}_k - \mathbf{H}\mathbf{m}_{k|k-1}) \tag{27c}$$

$$\bar{\mathbf{\Sigma}}_k = (\mathbf{I}-\bar{\mathbf{K}}_k\mathbf{H})\mathbf{\Sigma}_{k|k-1}(\mathbf{I}-\bar{\mathbf{K}}_k\mathbf{H})^T + \bar{\mathbf{K}}_k\bar{\mathbf{Q}}_w\bar{\mathbf{K}}_k^T. \tag{27d}$$

It can be seen that the computation in (27) is in fact the update stage in the standard KF [1]. The measurement noise covariance is equal to

$$\bar{\mathbf{Q}}_w = \mathrm{diag}(\langle\mathbf{\Lambda}_1\rangle^{-1},\langle\mathbf{\Lambda}_2\rangle^{-1},...,\langle\mathbf{\Lambda}_M\rangle^{-1}). \tag{28}$$

1) ORSKF-1: One way to derive $q(\mathbf{x}_k,\mathbf{b})$ and achieve robust Schmidt-Kalman filtering is to apply the idea of SKF such that

$$q(\mathbf{x}_k,\mathbf{b}) = \mathcal{N}(\mathbf{m}_k,\mathbf{\Sigma}_k) \tag{29}$$

where with slight relaxation of notations,

$$\mathbf{m}_k = \begin{bmatrix}\boldsymbol{\mu}_k \\ \mathbf{0}\end{bmatrix},\ \mathbf{\Sigma}_k = \begin{bmatrix}\mathbf{P}_k & \mathbf{C}_k \\ \mathbf{C}_k^T & \mathbf{B}\end{bmatrix}. \tag{30}$$

Because the posterior mean and covariance of the biases $\mathbf{b}$ are known priorly, we can find $q(\mathbf{x}_k, \mathbf{b})$ by determining $\boldsymbol{\mu}_k$, $\mathbf{P}_k$ and $\mathbf{C}_k$, which will be achieved by maximizing the ELBO in (24). Mathematically, this is equivalent to minimizing the KLD in (25) between $\mathcal{N}(\mathbf{m}_k, \boldsymbol{\Sigma}_k)$ in (29) and $\mathcal{N}(\bar{\mathbf{m}}_k, \bar{\boldsymbol{\Sigma}}_k)$ in (26). After applying the matrix determinant lemma [36], the associated KLD is proportional to [35]

$$\begin{aligned}\mathcal{L}_k = &-\log|\mathbf{P}_k - \mathbf{C}_k\mathbf{B}^{-1}\mathbf{C}_k^T| + \mathrm{tr}\left(\bar{\boldsymbol{\Sigma}}_k^{-1}\boldsymbol{\Sigma}_k\right) \\ &+ (\mathbf{m}_k - \bar{\mathbf{m}}_k)^T \bar{\boldsymbol{\Sigma}}_k^{-1} (\mathbf{m}_k - \bar{\mathbf{m}}_k).\end{aligned} \tag{31}$$

To facilitate the derivation, we express $\bar{\boldsymbol{\Sigma}}_k$ and its inverse in block form as

$$\bar{\boldsymbol{\Sigma}}_k = \begin{bmatrix}\bar{\mathbf{P}}_k & \bar{\mathbf{C}}_k \\ \bar{\mathbf{C}}_k^T & \bar{\mathbf{B}}_k\end{bmatrix}, \ \bar{\boldsymbol{\Sigma}}_k^{-1} = \begin{bmatrix}\mathbf{A}_k & \mathbf{D}_k \\ \mathbf{D}_k^T & \mathbf{S}_k\end{bmatrix}. \tag{32}$$

Putting (32) into (31) and neglecting the constant terms convert $\mathcal{L}_k$ into

$$\begin{aligned}\mathcal{L}_k = &-\log|\mathbf{P}_k - \mathbf{C}_k\mathbf{B}^{-1}\mathbf{C}_k^T| + \mathrm{tr}\left(\mathbf{A}_k\mathbf{P}_k + \mathbf{D}_k\mathbf{C}_k^T\right) \\ &+ \mathrm{tr}\left(\mathbf{D}_k^T\mathbf{C}_k + \mathbf{S}_k\mathbf{B}\right) - 2(\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}}_k)^T\mathbf{D}_k\bar{\boldsymbol{\mu}}_k^b \\ &+ (\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}}_k)^T\mathbf{A}_k(\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}}_k).\end{aligned} \tag{33}$$

To find $\boldsymbol{\mu}_k$, we calculate the partial derivative $\partial\mathcal{L}_k/\partial\boldsymbol{\mu}_k$ and set the result to zero to arrive at

$$\boldsymbol{\mu}_k = \bar{\boldsymbol{\mu}}_k + \mathbf{A}_k^{-1}\mathbf{D}_k\bar{\boldsymbol{\mu}}_k^b. \tag{34}$$

Similarly, the submatrix $\mathbf{P}_k$ can be shown, after using the results on the derivative of the matrix determinant [36], to be equal to

$$\mathbf{P}_k = \mathbf{A}_k^{-1} + \mathbf{C}_k\mathbf{B}^{-1}\mathbf{C}_k^T. \tag{35}$$

The cross covariance $\mathbf{C}_k$ can be found using $\partial\mathcal{L}_k/\partial\mathbf{C}_k = \mathbf{O}$, which is given by

$$\mathbf{C}_k = -\mathbf{A}_k^{-1}\mathbf{D}_k\mathbf{B}. \tag{36}$$

Using the block matrix inversion formula [36], we have that from (32), $\mathbf{A}_k^{-1}\mathbf{D}_k = -\bar{\mathbf{C}}_k\bar{\mathbf{B}}_k^{-1}$. Substituting this result into (34), (35) and (36) yields

$$\boldsymbol{\mu}_k = \bar{\boldsymbol{\mu}}_k - \bar{\mathbf{C}}_k\bar{\mathbf{B}}_k^{-1}\bar{\mathbf{b}}_k \tag{37a}$$

$$\mathbf{P}_k = \bar{\mathbf{P}}_k + \bar{\mathbf{C}}_k\bar{\mathbf{B}}_k^{-1}(\mathbf{B} - \bar{\mathbf{B}}_k)\bar{\mathbf{B}}_k^{-1}\bar{\mathbf{C}}_k^T \tag{37b}$$

$$\mathbf{C}_k = \bar{\mathbf{C}}_k\bar{\mathbf{B}}_k^{-1}\mathbf{B}. \tag{37c}$$

We shall provide a brief proof that the posterior covariance $\boldsymbol{\Sigma}_k$ with its blocks given in (37) is positive definite. Specifically, we know already that the prior covariance of the biases $\mathbf{B}$ is positive definite. Moreover, we have

$$\mathbf{P}_k - \mathbf{C}_k\mathbf{B}^{-1}\mathbf{C}_k^T = \bar{\mathbf{P}}_k - \bar{\mathbf{C}}_k\bar{\mathbf{B}}_k^{-1}\bar{\mathbf{C}}_k^T = \mathbf{A}_k^{-1} \tag{38}$$

which is positive definite. This completes verifying the positive definiteness of the obtained posterior covariance $\boldsymbol{\Sigma}_k$.

Putting (29) into (19), substituting the definition of $\boldsymbol{\delta}_i$ in (17) and carrying out the integral yield

$$\begin{aligned}\langle\boldsymbol{\delta}_i\boldsymbol{\delta}_i^T\rangle = &(\mathbf{y}_{i,k} - [\mathbf{H}_{i,x}, \mathbf{H}_{i,b}]\mathbf{m}_k)(\mathbf{y}_{i,k} - [\mathbf{H}_{i,x}, \mathbf{H}_{i,b}]\mathbf{m}_k)^T \\ &+ [\mathbf{H}_{i,x}, \mathbf{H}_{i,b}]\boldsymbol{\Sigma}_k[\mathbf{H}_{i,x}, \mathbf{H}_{i,b}]^T.\end{aligned} \tag{39}$$

Using the above result in (23) produces a closed-form expression for the posterior mean of the precision matrix $\boldsymbol{\Lambda}_i$.

We are ready to present the first proposed ORSKF, referred to as ORSKF-1. At the current sampling instant k, the ORSKF-1 takes the following processing steps:

1). Perform state prediction using (7);

2). Initialize $\langle\boldsymbol{\Lambda}_i\rangle$, $i = 1, 2, ..., M$, using their prior means such that $\langle\boldsymbol{\Lambda}_i\rangle = \mathbf{R}_i^{-1}$;

3). Generate $\bar{\mathbf{Q}}_w$ using (28), and find $\bar{\mathbf{m}}_k$ and $\bar{\boldsymbol{\Sigma}}_k$ using (27);

4). Compute the posterior mean and covariance of the state vector and biases, $\mathbf{m}_k$ and $\boldsymbol{\Sigma}_k$, using (37);

5). Update $\langle\boldsymbol{\Lambda}_i\rangle$ using (39) and (23);

6). Iterate steps 3)-5) until convergence.

It is important to note that at each sampling instant, the proposed ORSKF-1 attempts to approximate within the SKF framework (see (29)) the posterior of the state vector and bias terms when they are estimated *jointly* by the KF (see (27)). Therefore, with the ORSKF-1, the bias terms remain uncorrected over the whole state estimation process. This is the key difference between the ORSKF-1 and augmented-state approaches such as those in [9], [10] where the biases are always identified together with the state vector. As a result, the ORSKF-1 could be less prone to filtering divergence when the biases have low observability, compared with the augmented-state filters.

2) ORSKF-2: Alternatively, we can note from (23) and (39) that the precision matrix of the measurement noise $\boldsymbol{\Lambda}_i$ is updated in an adaptive manner such that the measurement residual $\boldsymbol{\delta}_i$ defined in (17) is taken into account. As such, if the current measurement subvector $\mathbf{y}_{i,k}$ is corrupted by outliers, the covariance of the measurement residual, $\langle\boldsymbol{\delta}_i\boldsymbol{\delta}_i^T\rangle$, would be large due to the increase in the residual (see (39))[1]. This motivates us to develop another ORSKF, which will be called ORSKF-2, that integrates the standard SKF presented in Section II and the adaptive updating of the measurement noise precision matrices.

At the current sampling instant k, the ORSKF-2 takes the following processing steps:

1). Perform state prediction using (7);

2). Initialize $\langle\boldsymbol{\Lambda}_i\rangle$, $i = 1, 2, ..., M$, using their prior means such that $\langle\boldsymbol{\Lambda}_i\rangle = \mathbf{R}_i^{-1}$;

3). Generate $\bar{\mathbf{Q}}_w$ using (28);

4). Find the posterior mean and covariance of the state vector and biases, $\mathbf{m}_k$ and $\boldsymbol{\Sigma}_k$, using the standard SKF given in (8) with $\mathbf{Q}_w$ replaced with $\bar{\mathbf{Q}}_w$;

5). Update $\langle\boldsymbol{\Lambda}_i\rangle$ using (39) and (23);

6). Iterate steps 3)-5) until convergence.

It can be seen that the difference between the two proposed ORSKFs lies in how they calculate the posterior mean and covariance of the state vector and biases. In particular, the ORSKF-2 applies the standard SKF with the updated noise covariances $\bar{\mathbf{Q}}_w$ while the ORSKF-1 resorts to the KLD minimization (see (25)).

[1]Note that the ORSKF-1 also uses this adaptive updating scheme to help achieve robustness to the outlying measurements (see (28), (23) and (39)).

IV. Experiment Results

This section evaluates the performance of the two proposed ORSKFs using synthetic and real-world data. The experiments are all concerned with estimating from raw TDOAs and FDOAs obtained over time by an array of N sensors the true TDOAs for the following localization applications [37]. At each sampling instant, the sensor array generates $N-1$ linearly independent TDOA and FDOA measurements, with respect to a reference sensor [38], [39]. To fuse these measurements, when realizing ORSKF-1 and ORSKF-2, we adopt the constant acceleration (CA) model [1] such that the state vector now consists of three $(N-1)\times 1$ subvectors. They contain respectively the true TDOAs, their first-order time derivative (i.e., the true FDOAs) and their second-order time derivative. Thus, the state transition matrix $\mathbf{F}_x$ in (1) is

$$\mathbf{F}_x = \begin{bmatrix} \mathbf{I}_{N-1} & T\cdot\mathbf{I}_{N-1} & \frac{T^2}{2}\cdot\mathbf{I}_{N-1} \\ \mathbf{O}_{N-1} & \mathbf{I}_{N-1} & T\cdot\mathbf{I}_{N-1} \\ \mathbf{O}_{N-1} & \mathbf{O}_{N-1} & \mathbf{I}_{N-1} \end{bmatrix}. \tag{40}$$

Besides, as the measurements are the obtained raw TDOAs and FDOAs, the measurement matrix $\mathbf{H}_x$ would be $\mathbf{H}_x = \mathbf{I}_{2(N-1)}$. Here, T denotes the sampling interval and $\mathbf{I}_{N-1}$ represents the $(N-1)\times(N-1)$ identity matrix. The state process noise $\mathbf{v}_k$ in (1a) is white Gaussian. It has zero mean and covariance $\mathbf{Q}_v = \sigma_v^2\mathbf{G}\mathbf{G}^T$, where $\mathbf{G} = [\frac{T^2}{2}\cdot\mathbf{I}_{N-1}, T\cdot\mathbf{I}_{N-1}, \mathbf{I}_{N-1}]^T$.

A. Synthetic Data

The scenario considered in [37] is used here. There are $N=5$ stationary sensors located at $\mathbf{s}_1=[10,0]^T$, $\mathbf{s}_2=[30,0]^T$, $\mathbf{s}_3=[50,0]^T$, $\mathbf{s}_4=[20,30]^T$, $\mathbf{s}_5=[40,30]^T$. Starting at $[25,15]^T$, the object to be localized moves along an 8-shaped trajectory with an average speed of 1.5/s for 84s. The object signal TDOAs and FDOAs are measuremed at the sensor array every $T=0.1$s. The obtained TDOAs are multiplied with the signal propagation speed and the measured FDOAs are scaled by the signal wavelength before they are fused. Figs. 1(a) and 1(b) plot the true TDOAs and FDOAs between sensor pair 2 and 1, while in Figs. 1(c) and 1(d), the true TDOAs and FDOAs between sensor pair 4 and 1 are shown. It is clear that the FDOAs are the changing rates of the corresponding TDOAs and they are more sensitive to the object movement.

We carry out two experiments and both have $L=1000$ ensemble runs. In each ensemble run, outlier-corrupted TDOA measurements are generated by adding to the true values multivariate Student's t-distributed noises with zero mean, degrees of freedom $\lambda=3$ and shape matrix $\mathbf{R}_t=\sigma_t^2(\mathbf{I}_{N-1}+\mathbf{1}\mathbf{1}^T)/2$. Here, $\sigma_t=0.5$ and $\mathbf{1}$ is an $(N-1)\times 1$ column vector with all the elements being equal to 1. The FDOAs with measurement noises independent to those in the TDOAs are generated in the same way, except that the shape matrix adopted is $\mathbf{R}_f=\sigma_f^2(\mathbf{I}_{N-1}+\mathbf{1}\mathbf{1}^T)/2$, where $\sigma_f=0.3$. The state process noise has a standard deviation $\sigma_v=1$.

For the purpose of comparison, we simulate the standard KF and SKF as well. The KF employs the CA model as in the ORSKFs but it *ignores* the presence of system biases. The SKF uses the same state-space model as the ORSKFs but it assumes Gaussian measurement noises. By moment matching [40], both the KF and SKF use a measurement covariance equal to $\mathbf{Q}_w = \text{diag}(\frac{\lambda}{\lambda-2}\mathbf{R}_t, \frac{\lambda}{\lambda-2}\mathbf{R}_f)$. The ORSKFs set that the Wishart priors for the precision matrices of the TDOA and FDOA measurement noises have scale matrices $(\lambda-2)\mathbf{R}_t^{-1}/(\lambda\nu)$ and $(\lambda-2)\mathbf{R}_f^{-1}/(\lambda\nu)$, where $\nu=4$. All the algorithms in consideration are initialized in the same way using the TDOA and FDOA measurements obtained at the first sampling instant.

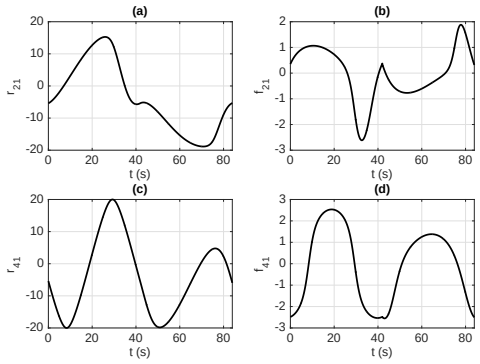

Fig. 1. True TDOAs and FDOAs collected by the sensor array over time. (a). True TDOAs between sensor pair 2 and 1. (b). True FDOAs between sensor pair 2 and 1. (c). True TDOAs between sensor pair 4 and 1. (d). True FDOAs between sensor pair 4 and 1.

We are interested in the estimation accuracy and consistency for the true TDOAs. They are quantified using the *averaged* root mean square error (RMSE), which is the TDOA estimation RMSE averaged over $L=1000$ ensemble runs and $N-1$ true TDOA estimates, and normalized estimation error squared (NEES) [1]. Ideally, the NEES should be close to $N-1=4$.

In the first synthetic data-based experiment, we consider the scenario where only the raw TDOAs are subject to additive measurement biases that are modeled as an $(N-1)\times 1$ Gaussian random vector with zero mean and covariance $\sigma_b^2\mathbf{I}_{N-1}$. σ_b is set to be 0.3. Figs. 2 and 3 depict the obtained averaged TDOA estimation RMSE and NEES results as a function of time. It can be seen that in this simulation, the proposed ORSKF-1 provides the best TDOA estimation accuracy while maintaining a NEES slightly smaller than 4, indicating that filtering consistency is achieved. The standard SKF suffers from significant degradation in its TDOA estimation performance due to the non-Gaussian noises in both the TDOA and FDOA measurements. The two developed ORSKFs performs better than the SKF, as they both can adaptively estimate the measurement noise covariance. Besides, according to Fig. 3, the standard KF has a NEES close to 40 (not shown in the figure), which is much greater than 4. Thus, it is highly overconfident because it simply neglects the presence of TDOA measurement biases.

We repeat the first experiment but this time, only the raw FDOAs are subject to additive zero-mean Gaussian-distributed measurement biases with covariance $\sigma_b^2\mathbf{I}_{N-1}$. The simulation results are summarized in Figs. 4 and 5. We can see that the

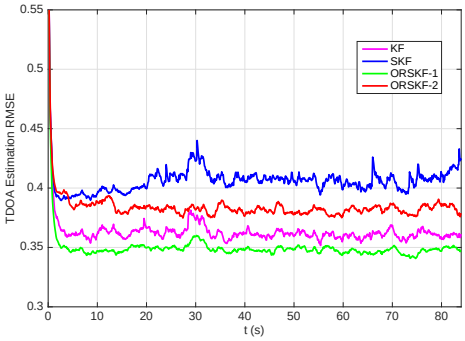

Fig. 2. Comparison of averaged TDOA estimation RMSEs over time under TDOA measurement biases.

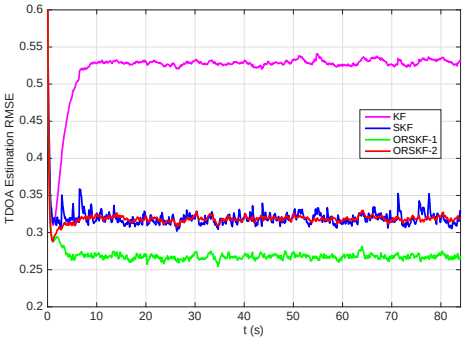

Fig. 4. Comparison of averaged TDOA estimation RMSEs over time under FDOA measurement biases.

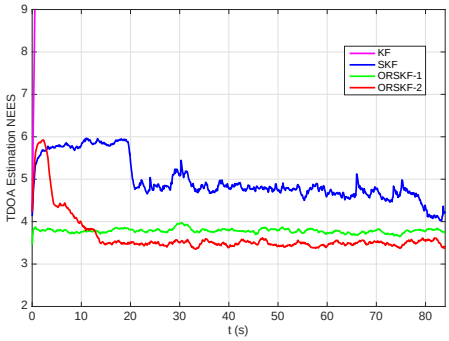

Fig. 3. Comparison of TDOA estimation NEESs over time under TDOA measurement biases.

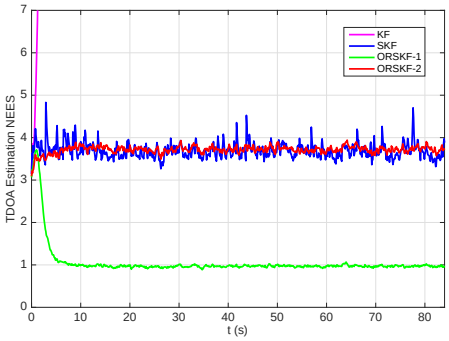

Fig. 5. Comparison of TDOA estimation NEESs over time under FDOA measurement biases.

ORSKF-1 continues to provide the smallest averaged TDOA estimation RMSE. But in this experiment, it over-estimates the true TDOA covariance and becomes a conservative filter, as the NEES is evidently lower than 4. This change in the filtering consistency of the ORSKF-1 might be explained by examining (37). It can be observed that the state estimate covariance is generated in a way somewhat *independent* of how the state estimate is modified (see (37a) and (37b)). The ORSKF-2 and standard SKF have similar performance but the ORSKF-2 is less sensitive to large measurement noises, again thanks to it adaptively estimating the measurement noise covariance. The KF remains overly confident in this simulation and it offers the worst TDOA estimation accuracy.

B. Real-World Data

A measurement compaign was conducted in November 2019, where three ground sensors were used to collect the signal TDOAs and FDOAs from a flying object. Some of the obtained measurements are shown in Fig. 6, where r_{i1} and f_{i1}, $i = 2, 3$, represent the TDOA and FDOA between sensor pair i and 1. Clearly, the TDOAs between sensors 3 and 1 are subject to a lot of outliers. The FDOA measurements have low noise level. But f_{21} has a negative bias from 5500s to 7900s; f_{31} has a negative bias as well from 8100s to 10000s.

The proposed ORSKFs are applied to fuse the TDOAs and FDOAs. Their implementations are the same as the ones used in the second synthetic data-based experiment, except that now, $N = 3$, $\sigma_t = 50$, $\sigma_f = 0.1$ and $\sigma_b = 0.05$. The estimated true TDOAs from ORSKFs are shown in Fig. (6a) and (6c). It can be seen that both ORSKFs show robustness to the outlying measurements but the TDOA estimates from the ORSKF-1 better follows the temporal evolution of the measurements.

V. Conclusions

The standard SKF attains filtering consistency by *considering* the impact of system biases when estimating the state and its covariance. Its optimality as a minimum variance state estimator without actively identifying the system biases would break down when the measurement noise is no longer Gaussian. To handle outlying measurements in the SKF framework, we assumed that the precision matrix of the measurement noise has a Wishart prior distribution. In this way, the measurement noise distribution now has longer tails than Gaussian, which could mitigate the effects of outliers on the filtering accuracy. Variational inference was utilized and two computationally tractable ORSKFs, namely ORSKF-1 and ORSKF-2, were established. They differ in how the SKF principle is integrated and they both estimate the measurement noise covariance adaptively to achieve robustness. Experiments using both synthetic and real-world data were conducted. It was found that both ORSKFs can provide improved state estimation

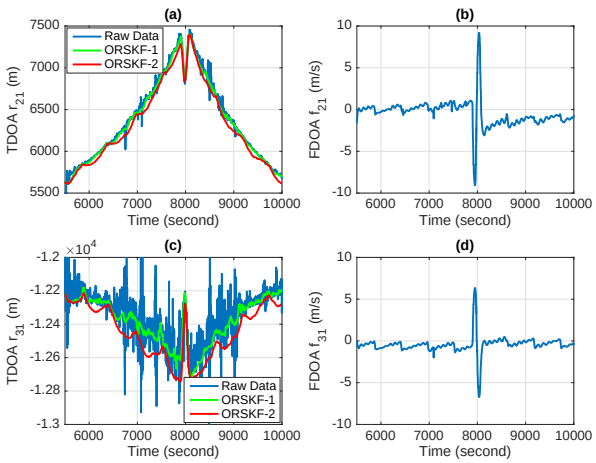

Fig. 6. (a). Measured and ORSKF-filtered TDOAs between sensor pair 2 and 1. (b). Measured FDOAs between sensor pair 2 and 1. (c). Measured and ORSKF-filtered TDOAs between sensor pair 3 and 1. (d). Measured FDOAs between sensor pair 3 and 1.

accuracy over the standard SKF when outliers are present. The estimated state covariance was either close to or greater than the true value, which corroborates that the proposed ORSKFs are consistent/conservative state estimators.

REFERENCES

[1] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. New York: Wiley, 2001.

[2] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York, NY: Academic Press, 1970.

[3] R. Y. Novoselov, S. M. Herman, S. M. Gadaleta, and A. B. Poore, "Mitigating the effects of residual biases with Schmidt-Kalman filtering," in *Proc. Int. Conf. Inf. Fusion (FUSION)*, Jul. 2005.

[4] R. Paffenroth, R. Novoselov, S. Danford, M. Teixeira, S. Chan, and A. B. Poore, "Mitigation of biases using the Schmidt-Kalman filter," in *Proc. SPIE 6699, Signal and Data Processing of Small Targets*, Sept. 2007.

[5] K. Wu and S. I. Roumeliotis, "Inverse Schmidt Estimators," Multiple Autonomous Robotic System Laboratory, Department of Computer Science & Engineering, University of Minnesota, Tech. Rep. Number-2016-003, Sept. 2016.

[6] K. M. Brink, "Partial-update Schmidt-Kalman filter," *Journal of Guidance, Control and Dynamics*, vol. 40, pp. 2214–2228, Sept. 2017.

[7] G. A. Watson, D. H. McCabe, and T. R. Rice, "Multisensor-multisite composite tracking in the presence of sensor residual bias," in *Proc. SPIE 3809, Signal and Data Processing of Small Targets*, Oct. 1999.

[8] B. D. Tapley, B. E. Schutz, and G. H. Born, *Statistical Orbit Determination*. New York: Elsevier, 2004.

[9] M. B. Ignagni, "Separate-bias Kalman estimator with bias state noise," *IEEE Trans. Autom. Control*, vol. 35, pp. 338–341, March 1990.

[10] A. T. Alouani, P. Xia, T. R. Rice, and W. D. Blair, "On the optimality of two-stage state estimationin the presence of random bias," *IEEE Trans. Autom. Control*, vol. 38, pp. 1279–1282, Aug. 1993.

[11] S. F. Schmidt, "Application of state space methods to navigation problems," *Advances in Control Systems*, vol. 3, pp. 293–340, 1966.

[12] M. E. Hough, "Orbit determination with improved covariance fidelity, including sensor measurement biases," *Journal of Guidance, Control and Dynamics*, vol. 34, pp. 903–911, May 2011.

[13] R. Zanetti and R. H. Bishop, "Precision entry navigation dead-reckoning error analysis: Theoretical foundations of discrete-time case," in *Proc. ASS/AIAA Astrodynamics Specialist Conf.*, 2007.

[14] J. L. Geeraert and J. W. McMahon, "Dual-satellite geolocation with ephemeris correction and uncertainty mapping," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, pp. 723–735, Feb. 2020.

[15] R. Zanetti and R. H. Bishop, "Kalman filters with uncompensated biases," *Journal of Guidance, Control and Dynamics*, vol. 35, pp. 327–330, Jan. 2012.

[16] D. P. Woodbury and J. L. Junkins, "On the consider Kalman filter," in *Proc. AIAA Guidance, Navigation, and Control Conf.*, Aug. 2010.

[17] J. S. McCabe and K. J. DeMars, "The Gaussian mixture consider Kalman filter," in *Proc. AAS/AIAA Space Flight Mechanics Meeting*, 2010.

[18] R. Zanetti and K. J. DeMars, "Joseph formulation of unscented and quadrature filters with application to consider states," *Journal of Guidance, Control and Dynamics*, vol. 36, pp. 1860–1864, Jan. 2013.

[19] J. Stauch and M. Jah, "Unscented Schmidt-Kalman filter algorithm," *Journal of Guidance, Control and Dynamics*, vol. 38, pp. 117–123, Jan. 2014.

[20] J. L. Geeraert and J. W. McMahon, "Square-root unscented Schmidt-Kalman filter," *Journal of Guidance, Control and Dynamics*, vol. 41, pp. 278–284, Jan. 2018.

[21] Y. Yang, H. Cui, R. Norman, and B. Gong, "Schmidt-Kalman filter with polynomial chaos expansion for state estimation," in *Proc. Int. Conf. Inf. Fusion (FUSION)*, Jul. 2019.

[22] H. Wymeersch, S. Marano, W. M. Gifford, and M. Z. Win, "A machine learning approach to ranging error mitigation for UWB localization," *IEEE Trans. Commun.*, vol. 60, pp. 1719–1728, June 2012.

[23] M. Roth, E. Özkan, and F. Gustafsson, "A Student's filter for heavy-tailed process and measurement noise," in *Proc. IEEE Intl. Conf. Acoust., Speech Signal Process., (ICASSP)*, May 2013.

[24] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "An outlier-robust Kalman filter," in *Proc. IEEE Intl. Conf. Robotics and Automation (ICRA)*, May 2011.

[25] G. Agamennoini, J. I. Nieto, and E. M. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Trans. Signal Process.*, vol. 60, pp. 5024–5037, Oct. 2012.

[26] H. Zhu, H. Leung, and Z. He, "A variational Bayesian approach to robust sensor fusion based on Student-t distribution," *Information Sciences*, vol. 221, pp. 201–214, 2013.

[27] Y. Huang, Y. Zhang, N. Li, Z. Wu, and J. Chambers, "A novel robust Student's t-based Kalman filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, pp. 1545–1554, June 2017.

[28] Y. Huang, Y. Zhang, Z. Wu, N. Li, and J. Chambers, "A novel adaptive Kalman filter with inaccurate process and measurement covariance matrices," *IEEE Trans. Autom. Control*, vol. 63, pp. 594–601, Feb. 2018.

[29] Y. Huang, Y. Zhang, Y. Zhao, L. Mihaylova, and J. Chambers, "Robust Rauch-Tung-Striebel smoothing framework for heavy-tail and/or skew noises," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, pp. 415–441, Feb. 2020.

[30] E. Weinstein and D. Kletter, "Delay and Doppler estimation by time-space partition of the array data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, pp. 1523–1535, Dec. 1983.

[31] C. Liu, L. Yang, and L. S. Mihaylova, "Dual-satellite source geolocation with time and frequency offsets and satellite location errors," in *Proc. Intl. Conf. Information Fusion (FUSION)*, July 2017.

[32] R. Zanetti and C. D'Souza, "Recursive implementations of the consider filter," *The Journal of the Astronautical Sciences*, vol. 60, pp. 672–685, Dec. 2013.

[33] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[34] K. P. Murphy, *Machine Learning: A Probablistic Perspective*. The MIT Press, 2012.

[35] S. Kullback and R. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, pp. 79–86, Jan. 1951.

[36] K. B. Peterson and M. S. Pederson, *The Matrix Cookbook*. http://matrixcookbook.com, 2012.

[37] Y. Liu, L. Yang, and J. Li, "Robust UWB indoor position tracking using TDOA measurements," in *Proc. IEEE Intl. Conf. Computer and Communications (ICCC)*, Dec. 2018.

[38] K. C. Ho and W. Xu, "An accurate algebraic solution for moving source location using TDOA and FDOA measurements," *IEEE Trans. Signal Process.*, vol. 52, pp. 2453–2463, Sept. 2004.

[39] L. Xi, F. Guo, L. Yang, and M. Zhang, "Improved solution for geolocating a known altitude source using TDOA and FDOA under random sensor location errors," *Electronics Letters*, vol. 54, pp. 597–599, Sept. 2018.

[40] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.