Robust Gaussian Kalman Filter With Outlier Detection

Hongwei Wang, *Student Member, IEEE*, Hongbin Li, *Senior Member, IEEE*, Jun Fang, and Heping Wang

***Abstract*—We consider the nonlinear robust filtering problem where the measurements are partially disturbed by outliers. A new robust Kalman filter based on a detect-and-reject idea is developed. To identify and exclude outliers automatically, each measurement is assigned an indicator variable, which is modeled by a beta-Bernoulli prior. The mean-field variational Bayesian method is then utilized to estimate the state of interest as well as the indicator in an iterative manner at each time instant. Simulation results reveal that the proposed algorithm outperforms several recent robust solutions with higher computational efficiency and better accuracy.**

***Index Terms*—beta-Bernoulli distribution, outlier detection, Robust Kalman filtering, state-space modeling.**

I. Introduction

DYNAMIC state space models (SSMs) have been widely used for target tracking, navigation, adaptive control, and many other applications. State estimation for SSMs is a fundamental problem. One seminal state estimation solution is the Kalman filter (KF) [1], which offers optimal estimation for linear SSMs with Gaussian process and measurement noises. Several suboptimal extensions were also presented for nonlinear SSMs under the Gaussian assumptions [2].

The performance of the KF and its suboptimal extensions may potentially degrade in the presence of outliers, which are frequently encountered in practical applications. The main reason for such degradation is that these methods were derived from the minimum mean square error criterion, which is sensitive to outliers [3]. A considerable amount of efforts have been devoted to improving the robustness of the filtering algorithms. One common strategy is robust statistics, especially *M*-estimation [4], [5]. In this approach, robust filters were constructed by recasting the KF as a linear regression. This type of robust filters were also extended via the nonlinear regression method [6]. Another popular family of solutions was based on heavy-tailed noise models. In [7]–[9], Student's *t*-distribution or Laplace distribution was utilized to model outlier-contaminated measurement noises. Other approaches for robust filtering, e.g., the H_∞ filter [10] and the sequential Monte-Carlo sampling/particle filter [11], were also studied.

The aforementioned robust solutions are all compensation-based strategies, that is, the outlier is considered as an inaccurate measurement which may contain some information of the latent state. In real applications, however, an outlier may directly come from clutter and should be discarded since it carries useless information. In such a case, the compensation-based strategies may degrade, especially for scenarios where the process of the states is well modeled.

In this letter, we derive a new robust KF which automatically identifies and excludes outliers in measurements. Such a detect-and-reject approach was recently introduced for principal component analysis and compressive sensing [12]–[14]. Specifically, we employ a probabilistic framework involving a binary indicator variable for each measurement to identify whether it is a nominal observation or an outlier. The indicator variable is modeled by a beta-Bernoulli hierarchical prior. A mean-field variational Bayesian (VB) inference method is employed to find the approximate posterior distributions of the indicator and the state of interest. During the review process of this letter, it was brought to our attention an extended KF based on detect-and-reject that was presented in [15]. The method employs a probability of outlier computed from the residual (difference between the measurement and one-step prediction), which is suitable for long-term estimation and low occurrence rate of outliers.

II. Problem Formulation

Consider a nonlinear discrete-time system described by the following SSM:

$$\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_{t-1}) + \boldsymbol{u}_{t-1} \tag{1}$$

$$\boldsymbol{y}_t = \boldsymbol{h}(\boldsymbol{x}_t) + \boldsymbol{v}_t \tag{2}$$

where $\boldsymbol{x}_t \in \Re^n$ is the state of interest at time instant t; $\boldsymbol{y}_t \in \Re^m$ is a measurement with respect to $\boldsymbol{x}_t$; $\boldsymbol{f}(\cdot)$ and $\boldsymbol{h}(\cdot)$ are some known nonlinear functions, denoting the state evolution and observation procedure, respectively; $\boldsymbol{u}_{t-1} \in \Re^n$ and $\boldsymbol{v}_t \in \Re^m$ are zero-mean Gaussian vectors with covariance $\boldsymbol{Q}_{t-1}$ and $\boldsymbol{R}_t$, representing the process noise and measurement noise. $\boldsymbol{u}_{t-1}$ and $\boldsymbol{v}_t$ are assumed to be mutually independent and both are also uncorrelated to the initial state $\boldsymbol{x}_0$, which has a known Gaussian prior distribution $p(\boldsymbol{x}_0) = \mathcal{N}(\hat{\boldsymbol{x}}_0, \boldsymbol{P}_0)$.

Manuscript received May 20, 2018; revised June 19, 2018; accepted June 22, 2018. Date of publication June 27, 2018; date of current version July 11, 2018. This work was supported by the China Scholarship Council and the National Natural Science Foundation of China under Grant 61473227, Grant 11472222, and Grant U1530154. The work of H. Li was supported in part by the National Science Foundation under Grant ECCS-1609393 and in part by the Air Force Office of Scientific Research under Grant FA9550-16-1-0243. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Woon-Seng Gan. *(Corresponding author: Hongbin Li.)*

H. Wang is with the School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China, and also with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: tianhangxinxiang@163.com).

H. Li is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: hli@stevens.edu).

J. Fang is with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: JunFang@uestc.edu.cn).

H. Wang is with the School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: wangheping@nwpu.edu.cn).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2018.2851156

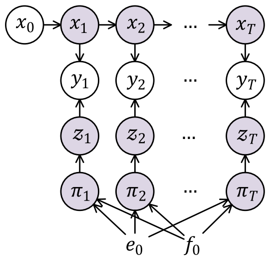

Fig. 1. Graphical model for the outlier-detecting SSM.

In real applications, the measurement may be disturbed by some outliers. In order to detect the outliers, we employ the beta-Bernoulli distribution which has been successfully used to identify outliers for various applications [12]–[14]. Specifically, we introduce a binary indicator variable z_t for the measurement $\boldsymbol{y}_t$ to indicate whether it is an outlier or not. More precisely, $z_t = 1$ means that $\boldsymbol{y}_t$ is a nominal measurement (i.e., regular outlier-free measurement), while $z_t = 0$ indicates $\boldsymbol{y}_t$ is an outlier. The measurement likelihood conditional on its indicator and the state can be formulated as

$$p(\boldsymbol{y}_t|\boldsymbol{x}_t, z_t) = \mathcal{N}(\boldsymbol{h}(\boldsymbol{x}_t), \boldsymbol{R}_t)^{z_t}. \tag{3}$$

Clearly, for $z_t = 1$, (3) is a standard Gaussian distribution, whereas for $z_t = 0$, (3) reduces to a constant. In the latter case, $\boldsymbol{y}_t$ is effectively marked as an outlier and cannot contribute to the estimation process since its distribution is independent of observation. Furthermore, we impose a beta-Bernoulli hierarchical prior [16] to the indicator. Specifically, z_t is a Bernoulli variable controlled by π_t

$$p(z_t|\pi_t) = \pi_t^{z_t}(1-\pi_t)^{(1-z_t)} \tag{4}$$

while π_t is a beta distribution parameterized by e_0 and f_0

$$p(\pi_t) = \frac{\pi_t^{e_0-1}(1-\pi_t)^{f_0-1}}{\mathrm{B}(e_0, f_0)} \tag{5}$$

where $\mathrm{B}(\cdot,\cdot)$ is the beta function. Altogether, the resulting graphical model is illustrated as in Fig. 1.

The objective of this work is to present a new framework for robust Kalman filtering to provide effective state estimation as well as outlier detection. The proposed framework is integrated with the Gaussian approximation filter, resulting in a class of outlier-detecting KF solutions.

III. VB Inference

In this section, we develop a robust Gaussian approximation KF via statistical inference over the aforementioned graphical model. Let us denote all latent variables by $\boldsymbol{\Theta}_t = \{\boldsymbol{x}_t, \pi_t, z_t\}$ for brevity. According to Bayes' law, the posterior distribution conditioned on the observations $\boldsymbol{y}_{1:t}$ is

$$p(\boldsymbol{\Theta}_t|\boldsymbol{y}_{1:t}) = \frac{p(\boldsymbol{\Theta}_t, \boldsymbol{y}_{1:t})}{p(\boldsymbol{y}_{1:t})}. \tag{6}$$

However, the posterior distribution is intractable since $p(\boldsymbol{y}_{1:t})$ is in general hard to compute. Therefore, some approximation methods should be employed. Here, we utilize the VB approach to provide an approximation of the posterior density. A primary reason for resorting to the VB approach is its computational efficiency. A close examination of the proposed graphical model in (3)–(5) and Fig. 1 reveals the conjugacy between the priors and likelihoods, which facilitates the derivation of efficient VB inference algorithms.

Specifically, a tractable distribution $q(\boldsymbol{\Theta}_t)$ is carefully chosen to approximate the posterior distribution $p(\boldsymbol{\Theta}_t|\boldsymbol{y}_{1:t})$. This is realized by minimizing the Kullback–Leibler divergence (KLD) between $q(\boldsymbol{\Theta}_t)$ and $p(\boldsymbol{\Theta}_t|\boldsymbol{y}_{1:t})$

$$\mathcal{D}(p\|q) = \int q(\boldsymbol{\Theta}_t)\log\frac{q(\boldsymbol{\Theta}_t)}{p(\boldsymbol{\Theta}_t|\boldsymbol{y}_{1:t})}d\boldsymbol{\Theta}_t \tag{7}$$

$$\overset{(6)}{=} \int q(\boldsymbol{\Theta}_t)\log\frac{q(\boldsymbol{\Theta}_t)}{p(\boldsymbol{\Theta}_t, \boldsymbol{y}_{1:t})}d\boldsymbol{\Theta}_t + \log p(\boldsymbol{y}_{1:t}). \tag{8}$$

It is still hard to find $q(\boldsymbol{\Theta}_t)$ by directly minimizing (8). For simplicity, we apply the mean-field approximation, which impose a factorized form for $q(\boldsymbol{\Theta}_t)$ [17]

$$q(\boldsymbol{\Theta}_t) = q(\boldsymbol{x}_t)q(z_t)q(\pi_t). \tag{9}$$

According to the update rule of the VB method, each variational distribution can be updated in the following alternating fashion [17]:

$$q(\boldsymbol{x}_t) \propto \exp\left\langle \ln p(\boldsymbol{x}_t, z_t, \pi_t, \boldsymbol{y}_{1:t})\right\rangle_{q(z_t)q(\pi_t)} \tag{10}$$

$$q(z_t) \propto \exp\left\langle \ln p(\boldsymbol{x}_t, z_t, \pi_t, \boldsymbol{y}_{1:t})\right\rangle_{q(\boldsymbol{x}_t)q(\pi_t)} \tag{11}$$

$$q(\pi_t) \propto \exp\left\langle \ln p(\boldsymbol{x}_t, z_t, \pi_t, \boldsymbol{y}_{1:t})\right\rangle_{q(\boldsymbol{x}_t)q(z_t)} \tag{12}$$

where $\langle g(x)\rangle_{q(x)}$ denotes the expectation of $g(x)$ over the distribution of $q(x)$. $p(\boldsymbol{x}_t, z_t, \pi_t, \boldsymbol{y}_{1:t})$ is the full distribution of the SSM, given by

$$p(\boldsymbol{x}_t, z_t, \pi_t, \boldsymbol{y}_{1:t}) \propto p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})p(\boldsymbol{y}_t|\boldsymbol{x}_t, z_t)p(z_t|\pi_t)p(\pi_t). \tag{13}$$

In the KF framework, the first term $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})$ on the right-hand side of (13) is the predictive density, which can be approximated by a Gaussian density $p(\boldsymbol{x}_{t|t-1}) \sim \mathcal{N}(\hat{\boldsymbol{x}}_{t|t-1}, \boldsymbol{P}_{t|t-1})$ since the process noise is Gaussian [2]

$$\hat{\boldsymbol{x}}_{t|t-1} = \int f(\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{k-1|k-1})d\boldsymbol{x}_{t-1} \tag{14}$$

$$\boldsymbol{P}_{t|t-1} = \int \left(f(\boldsymbol{x}_{t-1}) - \hat{\boldsymbol{x}}_{t|t-1}\right)\left(f(\boldsymbol{x}_{t-1}) - \hat{\boldsymbol{x}}_{t|t-1}\right)^T \times p(\boldsymbol{x}_{k-1|k-1})d\boldsymbol{x}_{t-1} + \boldsymbol{Q}_{t-1} \tag{15}$$

where $p(\boldsymbol{x}_{k-1|k-1}) \sim \mathcal{N}(\hat{\boldsymbol{x}}_{t-1|t-1}, \boldsymbol{P}_{t-1|t-1})$ is the optimally approximated posterior distribution at the last time instant.

A. Update of $q(\boldsymbol{x}_t)$

Keeping the terms that only depend on $\boldsymbol{x}_t$, we have

$$q(\boldsymbol{x}_t) \propto \exp\left(\ln p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}) + \langle \ln p(\boldsymbol{y}_t|\boldsymbol{x}_t, z_t)\rangle_{q(z_t)}\right). \tag{16}$$

Furthermore, (16) can be expressed as

$$q(\boldsymbol{x}_t) \propto \exp\left(-\frac{1}{2}\|\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t-1}\|^2_{\boldsymbol{P}^{-1}_{t|t-1}} - \frac{\langle z_t\rangle}{2}\|\boldsymbol{y}_t - h(\boldsymbol{x}_t)\|^2_{\boldsymbol{R}_t^{-1}}\right) \tag{17}$$

where $\langle z_t\rangle$ is the expectation of z_t. It is easy to illustrate that $q(\boldsymbol{x}_t)$ is a Gaussian distribution $\mathcal{N}(\hat{\boldsymbol{x}}_{t|t}, \boldsymbol{P}_{t|t})$. The mean and covariance of this Gaussian distribution can be found in the Kalman-filtering framework with a modified measurement covariance $\bar{\boldsymbol{R}}_t = \boldsymbol{R}_t / \langle z_t \rangle$, namely

$$\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1}) \tag{18}$$

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{K}_t(\boldsymbol{S}_t + \bar{\boldsymbol{R}}_t)\boldsymbol{K}_t^T \tag{19}$$

$$\boldsymbol{K}_t = \boldsymbol{C}_t(\boldsymbol{S}_t + \bar{\boldsymbol{R}}_t)^{-1} \tag{20}$$

where

$$\hat{\boldsymbol{y}}_{t|t-1} = \int \boldsymbol{h}(\boldsymbol{x}_t) p(\hat{\boldsymbol{x}}_{t|t-1}) d\boldsymbol{x}_t \tag{21}$$

$$\boldsymbol{S}_t = \int \left(\boldsymbol{h}(\boldsymbol{x}_t) - \boldsymbol{y}_{t|t-1}\right)\left(\boldsymbol{h}(\boldsymbol{x}_t) - \boldsymbol{y}_{t|t-1}\right)^T p(\hat{\boldsymbol{x}}_{t|t-1}) d\boldsymbol{x}_t \tag{22}$$

$$\boldsymbol{C}_t = \int \left(\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t-1}\right)\left(\boldsymbol{h}(\boldsymbol{x}_t) - \boldsymbol{y}_{t|t-1}\right)^T p(\hat{\boldsymbol{x}}_{t|t-1}) d\boldsymbol{x}_t. \tag{23}$$

When $\langle z_t \rangle$ is close to 0, say, $\langle z_t \rangle \leq 1 \times 10^{-10}$, we can treat the measurement as an outlier, which does not contain any useful information. In such a case, we ignore the measurement and simply update $q(\boldsymbol{x}_t) \sim \mathcal{N}(\hat{\boldsymbol{x}}_{t|t}, \boldsymbol{P}_{t|t})$ by the predicted distribution $p(\hat{\boldsymbol{x}}_{t|t-1})$, i.e.,

$$\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t|t-1} \tag{24}$$

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1}. \tag{25}$$

B. Update of $q(z_t)$

The variational distribution of $q(z_t)$ can be obtained as

$$\begin{aligned} q(z_t) &\propto \exp \langle \ln p(\boldsymbol{y}_t|\boldsymbol{x}_t, z_t) + \ln p(z_t|\pi_t)\rangle_{q(\boldsymbol{x}_t)q(\pi_t)} \\ &\propto \exp\{-0.5 z_t \mathrm{tr}(\boldsymbol{B}_t \boldsymbol{R}_t^{-1}) \\ &\qquad + z_t \langle \ln \pi_t \rangle + (1 - z_t)\langle \ln(1-\pi_t)\rangle\} \end{aligned} \tag{26}$$

where $\boldsymbol{B}_t = \int (\boldsymbol{y}_t - h(\boldsymbol{x}_t))(\boldsymbol{y}_t - h(\boldsymbol{x}_t))^T q(\boldsymbol{x}_t) d\boldsymbol{x}_t$. Clearly, z_t is a Bernoulli parameter with

$$P(z_t = 1) = A e^{\langle \ln \pi_t \rangle - 0.5 \mathrm{tr}(\boldsymbol{B}_t \boldsymbol{R}_t^{-1})} \tag{27}$$

$$P(z_t = 0) = A e^{\langle \ln(1-\pi_t)\rangle} \tag{28}$$

where A is a normalizing constant to ensure $P(z_t = 1) + P(z_t = 0) = 1$, and

$$\langle \ln \pi_t \rangle = \Psi(e_t) - \Psi(e_t + f_t + 1) \tag{29}$$

$$\langle \ln(1 - \pi_t)\rangle = \Psi(f_t + 1) - \Psi(e_t + f_t + 1) \tag{30}$$

where $\Psi(\cdot)$ is the digamma function [16]. Hence the expectation of z_t can be updated as

$$\langle z_t \rangle = \frac{P(z_t = 1)}{P(z_t = 1) + P(z_t = 0)}. \tag{31}$$

C. Update of $q(\pi_t)$

Dropping off the terms in (12) that do not depend on π_t, we have

$$\begin{aligned} q(\pi_t) &\propto \exp \langle \ln p(z_t|\pi_t) + \ln p(\pi_t)\rangle_{q(\boldsymbol{x}_t)q(z_t)} \\ &\propto \exp\{e_t \ln \pi_t + f_t \ln(1-\pi_t)\} \end{aligned} \tag{32}$$

Algorithm 1: Robust KF with outlier detection.

Input: $\boldsymbol{y}_{1:T}, \hat{\boldsymbol{x}}_0, \boldsymbol{P}_0, \boldsymbol{Q}_{1:T}, \boldsymbol{R}_{1:T}$.
Output: $\hat{\boldsymbol{x}}_{t|t}$ and $\boldsymbol{P}_{t|t}$ for $t = 1:T$.
for $t = 1, \ldots, T$ **do**
  Update $\hat{\boldsymbol{x}}_{t|t-1}$ and $\boldsymbol{P}_{t|t-1}$ via (14) and (15);
  Initialize $i = 0$, e_0, f_0 and $z_t^i = 1$;
  repeat $i = 1, \ldots$
    Update $\{\hat{\boldsymbol{x}}_{t|t}^{i+1}, \boldsymbol{P}_{t|t}^{i+1}\}$ via $\{(18), (19)\}$ or $\{(24), (25)\}$;
    Update $z_t^{i+1} = \langle z_t \rangle$ via (31);
    Update e_t^{i+1} and f_t^{i+1} via (33) and (34);
    Compute $\tau = \|\hat{\boldsymbol{x}}_{t|t}^{i+1} - \hat{\boldsymbol{x}}_{t|t}^{i}\| / \|\hat{\boldsymbol{x}}_{t|t}^{i}\|$
    $i = i + 1$;
  until $\tau \leq 10^{-6}$
  $\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t|t}^{i-1}$ and $\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t}^{i-1}$;
end for

where

$$e_t = e_0 + \langle z_t \rangle \tag{33}$$

$$f_t = f_0 + 1 - \langle z_t \rangle. \tag{34}$$

It can be verified that $q(\pi_t)$ is a beta distribution, i.e.,

$$q(\pi_t) = \mathrm{Beta}(e_t, f_t) \tag{35}$$

For clarity, the resulting robust KF with outlier detection is summarized in Algorithm 1.

IV. Simulation Results

We consider the problem of tracking a maneuvering target [18] to illustrate the performance of the proposed algorithm in the presence of measurement outliers. In the simulation, the Gaussian integrals in Algorithm 1 are approximately evaluated by the cubature rule [19], and the resulting method is referred as outlier-detecting cubature KF (OD-CKF). The beta-Bernoulli parameters are set as $e_0 = 0.9$ and $f_0 = 0.1$. We found that the proposed algorithm is insensitive to these parameters as long as e_0 is larger than f_0 such that the prior mean of π_t, which is $e_0/(e_0 + f_0)$, is close to 1. The latter condition is due to the fact that outliers are rare and the probability of observing a nominal measurement is close to 1 in general.

For comparison, we also consider the conventional cubature KF (CKF) [19] and several robust solutions, including the Institute of Geodesy and Geophysics based cubature information filter (IGG-CIF) [5], [20], Huber-based cubature information filter (Huber-CIF) [5], Hampel-based cubature information filter (Hampel-CIF) [5], and Student's *t*-based robust cubature KF (STU-CKF) [18]. The design parameters of these filters are set as recommended in the original literature.

The dynamics of the maneuvering target with an unknown turning rate are modeled as [18]

$$\boldsymbol{x}_{t+1} = \begin{pmatrix} 1 & \frac{\sin(\omega_t \Delta t)}{\omega_t} & 0 & \frac{\cos(\omega_t \Delta t) - 1}{\omega_t} & 0 \\ 0 & \cos(\omega_t \Delta t) & 0 & -\sin(\omega_t \Delta t) & 0 \\ 0 & \frac{1 - \cos(\omega_t \Delta t)}{\omega_t} & 1 & \frac{\sin(\omega_t \Delta t)}{\omega_t} & 0 \\ 0 & \sin(\omega_t \Delta t) & 0 & \cos(\omega_t \Delta t) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \boldsymbol{x}_t + \boldsymbol{u}_t \tag{36}$$

where the state $\boldsymbol{x}_t$ is defined as $[a_t, \dot{a}_t, b_t, \dot{b}_t, \omega_t]^T$, containing the two-dimensional location (a_t, b_t) and the corresponding ve-

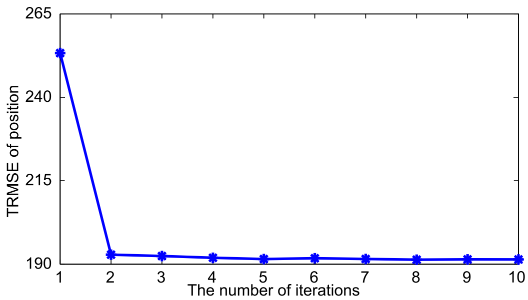

Fig. 2. TRMSE of position versus iterations.

locities $(\dot{a}_t, \dot{b}_t)$ as well as the turning rate ω_t, Δt is the sampling time, and $\boldsymbol{u}_t$ is a white Gaussian noise with covariance $\boldsymbol{Q}_t$ scaled by q_1 and q_2

$$\boldsymbol{Q}_t = \begin{pmatrix} q_1\boldsymbol{M} & 0 & 0 \\ 0 & q_1\boldsymbol{M} & 0 \\ 0 & 0 & q_2 \end{pmatrix}, \ \boldsymbol{M} = \begin{pmatrix} \Delta t^3/3 & \Delta t^2/2 \\ \Delta t^2/2 & \Delta t \end{pmatrix} \tag{37}$$

The measurement obtained by an active radar is

$$\boldsymbol{y}_{t+1} = \begin{pmatrix} r_{t+1} \\ \theta_{t+1} \end{pmatrix} = \begin{pmatrix} \sqrt{a_{t+1}^2 + b_{t+1}^2} \\ \text{atan2}(b_{t+1}, a_{t+1}) \end{pmatrix} + \boldsymbol{v}_{t+1} \tag{38}$$

where atan2 is the four-quadrant inverse tangent function and $\boldsymbol{v}_{t+1}$ is the measurement noise. In order to describe the heavy-tailed property of $\boldsymbol{v}_{t+1}$ caused by measurement outliers, we utilize the Gaussian mixture model [9]

$$\boldsymbol{v}_{t+1} = (1-\lambda)\mathcal{N}(0, \boldsymbol{R}_{t+1}) + \lambda\mathcal{N}(0, \alpha\boldsymbol{R}_{t+1}) \tag{39}$$

where λ is the contamination ratio to indicate the probability of outlier and α is a scale factor indicating the power of the contaminating noise compared with the nominal noise.

In the simulation, we set the initial conditions as follows: Sampling time $\Delta t = 1$; $q_1 = 0.1$ and $q_2 = 1.75 \times 10^{-4}$; the initial state $\boldsymbol{x}_0$ follows $\mathcal{N}(\hat{\boldsymbol{x}}_{0|0}, \boldsymbol{P}_{0|0})$ with $\hat{\boldsymbol{x}}_{0|0} = [-10\,000; 10; 5000; -5; -0.053]$ and $\boldsymbol{P}_{0|0} = \boldsymbol{Q}_t$; $\boldsymbol{R}_t = \text{diag}(50^2, 1.6 \times 10^{-3})$. For fair comparison, we implement $L = 1000$ independent Monte-Carlo runs and in each run $T = 100$ noisy measurements are collected. The time-averaged root mean square error (TRMSE) and implementation time (IT) are used as performance metrics. The TRMSE of the position estimate is defined as

$$\text{TRMSE}_{\text{pos}} = \frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{1}{L}\sum_{i=1}^{L}\left((a_t^i - \hat{a}_t^i)^2 + (b_t^i - \hat{b}_t^i)^2\right)} \tag{40}$$

where (a_t^i, b_t^i) and $(\hat{a}_t^i, \hat{b}_t^i)$, respectively, are the true and estimated state component at time t in the ith Monte-Carlo run. In this simulation, we omit the TRMSE of the velocity, which has a similar pattern as that of the position TRMSE.

Fig. 2 shows the position TRMSE of our algorithm versus the number of VB iterations when $\lambda = 0.2$ and $\alpha = 100$. It is seen that the proposed algorithm needs only two or three iterations to converge. In the following, we use three iterations as a default value for the proposed algorithm.

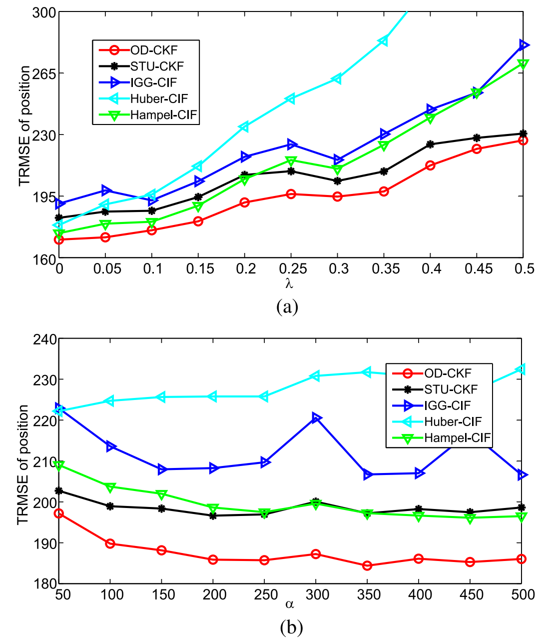

Fig. 3. TRMSEs of position in different scenarios. (a) Varying contamination ratio when $\alpha = 300$. (b) Varying contamination strength when $\lambda = 0.2$.

TABLE I
IMPLEMENTATION TIME (IT)

$\lambda = 0.2$, $\alpha = 300$			
Methods	Computational Time	Methods	Computational Time
CKF	45.04	OD-CKF	58.02
STU-CKF	88.77	Hampel-CIF	50.01
IGG-CIF	50.09	Huber-CIF	50.67

In Fig. 3(a) and (b), we show the position TRMSEs for the robust filters. We do not include the CKF, since its performance degrades considerably in these scenarios with outliers. The TRMSEs of position when $\alpha = 300$ and λ varies from 0 to 0.5 as shown in Fig. 3(a). Although robust filters have increasing TRMSEs, the Huber-CIF rises more significantly than the others. Among the robust filters, the proposed OD-CKF has the lowest TRMSE for all contamination ratios. The Hampel-CIF (on lower λ) and STU-CKF (on larger λ) are the closest, but their TRMSEs are still about 8%–10% larger. In Fig. 3(b), we show the position TRMSEs versus the contamination level α while λ is fixed at 0.2. It is seen that all robust filters are relatively insensitive to the contamination strength. Again, our algorithm achieves the lowest TRMSE.

The ITs of different algorithms are shown in Table I. It is seen that the ITs of the robust filters are larger than that of the CKF. All robust filters have similar computational complexity since they utilize the framework of CKF (or CIF) with modified measurement update strategies which lead to some differences in ITs. It is noted that our method is not the most computationally intensive one, but yields consistently the best estimation accuracy among all robust filters in consideration.

V. CONCLUSION

We proposed a new unified framework for robust Kalman filtering with outlier detection. The proposed approach employs an indicator variable with a beta-Bernoulli prior. A mean-field VB method was applied to find the approximate posterior distributions of the state of interest as well as the indicator. It is noteworthy that although in the simulation part, our algorithm was implemented in conjunction with the CKF, it is straightforward to extend the framework to use with other nonlinear Gaussian approximation filters as discussed in, e.g., [2].

References

[1] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960.

[2] Y. Wu, D. Hu, M. Wu, and X. Hu, “A numerical-integration perspective on Gaussian filters,” *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2910–2921, Aug. 2006.

[3] I. C. Schick and S. K. Mitter, “Robust recursive estimation in the presence of heavy-tailed observation noise,” *Ann. Stat.*, vol. 22, pp. 1045–1080, 1994.

[4] M. A. Gandhi and L. Mili, “Robust Kalman filter based on a generalized maximum-likelihood-type estimator,” *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2509–2520, May 2010.

[5] L. Chang and K. Li, “Unified form for the robust Gaussian information filtering based on M-estimate,” *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 412–416, Apr. 2017.

[6] C. D. Karlgaard, “Nonlinear regression Huber–Kalman filtering and fixed-interval smoothing,” *J. Guid, Cont. Dynam.*, vol. 38, no. 2, pp. 322–330, 2014.

[7] G. Agamennoni, J. I. Nieto, and E. M. Nebot, “Approximate inference in state-space models with heavy-tailed noise,” *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5024–5037, Oct. 2012.

[8] Y. Huang, Y. Zhang, N. Li, and J. Chambers, “Robust Student’s t based nonlinear filter and smoother,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 5, pp. 2586–2596, Oct. 2016.

[9] H. Wang, H. Li, W. Zhang, and H. Wang, “Laplace ℓ_1 robust Kalman filter based on majorization minimization,” in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, 2017, pp. 1–5.

[10] J.-S. Hu and C.-H. Yang, “Second-order extended H_∞ filter for nonlinear discrete-time systems using quadratic error matrix approximation,” *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3110–3119, Jul. 2011.

[11] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking,” *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[12] S. T. Garren, R. L. Smith, and W. W. Piegorsch, “Bootstrap goodness-of-fit test for the beta-binomial model,” *J. Appl. Stat.*, vol. 28, no. 5, pp. 561–571, 2001.

[13] X. Ding, L. He, and L. Carin, “Bayesian robust principal component analysis,” *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3419–3430, Dec. 2011.

[14] Q. Wan, H. Duan, J. Fang, H. Li, and Z. Xing, “Robust Bayesian compressed sensing with outliers,” *Signal Process.*, vol. 140, pp. 104–109, 2017.

[15] H.-Q. Mu and K.-V. Yuen, “Novel outlier-resistant extended Kalman filter for robust online structural identification,” *J. Eng. Mech.*, vol. 141, no. 1, Art. no. 04014100, 2014.

[16] J. Paisley and L. Carin, “Nonparametric factor analysis with beta process priors,” in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 777–784.

[17] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, “The variational approximation for Bayesian inference,” *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.

[18] R. Piché, S. Särkkä, and J. Hartikainen, “Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution,” in *IEEE Int. Workshop Mach. Learn. Signal Process.*, 2012, pp. 1–6.

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

[20] H. Wu, S. Chen, B. Yang, and K. Chen, “Robust derivative-free cubature Kalman filter for bearings-only tracking,” *J. Guid., Control Dynam.*, vol. 39, no. 8, pp. 1866–1871, 2016.