Generalized Iterated Kalman Filter and its Performance Evaluation

Xiaoqing Hu, Ming Bao, Xiao-Ping Zhang, *Senior Member, IEEE*, Luyang Guan, and Yu-Hen Hu, *Fellow, IEEE*

***Abstract*—In this paper, we present a generalized iterated Kalman filter (GIKF) algorithm for state estimation of a nonlinear stochastic discrete-time system with state-dependent multiplicative observation noise. The GIKF algorithm adopts the Newton–Raphson iterative optimization steps to yield an approximate maximum *a posteriori* estimate of the states. The mean-square estimation error (MSE) and the Cramér–Rao lower bound (CRLB) of the state estimates are also derived. In particular, the local convergence of MSE of GIKF is rigorously established. It is also proved that the GIKF yields a smaller MSE than those of the generalized extended Kalman filter and the traditional extended Kalman filter. The performance advantages and convergence of GIKF are demonstrated using Monte Carlo simulations on a target tracking application in a range measuring sensor network.**

***Index Terms*—Convergence, iterated Kalman filter, multiplicative noise, nonlinear systems.**

I. Introduction

DISCRETE-TIME filtering for nonlinear stochastic systems has been the subject of considerable research during the past few decades. Scientific and engineering applications include target tracking, infrastructure monitoring, habitat sensing, and battlefield surveillance [1]–[3]. The objective of nonlinear filtering is to estimate the state of a dynamic process based on noisy observations. To infer the states of the system (position, velocity, attitude, and heading, etc.), certain modalities of measurements such as time of arrival [4], signal intensity [5], phase lags [6] or images [7] must be incorporated.

In practical applications, the measurement model is often signal dependent. For instance, in a bearings-only sensor application [8], the measurement noise is a function of the signal to noise ratio (SNR) and the incident angle of the signal. In a ranging sensor application [9], the measurement noise increases as the relative distance to the object increases. In the stereo camera application [10], the absolute noise increases with the magnitude of the signal. The measurement signals contaminated by the multiplicative noise are common in many systems such as image processing systems [11], control systems [12], communication systems [13] and tracking systems [14].

However, the widely used Kalman filter (KF) [15] and its variants such as extended Kalman filter (EKF) [16], iterated Kalman filter (IKF) [17], interactive multiple model Kalman filter (IMM) [18], unscented Kalman filter (UKF) [19], iterated unscented Kalman filter [20] and Gaussian filter [21], often assume that the observation noise is additive and an i.i.d Gaussian process. The assumption would become invalid when the observation noise is signal-dependent, since the statistics of the multiplicative noise depends on the state of the system. This situation motivates our focus on an estimation problem of the state of a nonlinear stochastic discrete-time system with state-dependent multiplicative observation noise.

Previously, using a maximum likelihood approach coupled with Gauss-Newton algorithm, a recursive algorithm is proposed to estimate the state of a stochastic process from measurements with additive state-dependent observation noise [22]. This method, however, only applies to observations made on a single sensor, and does not readily generalize to observations made by multiple sensors. In [10], a robust estimation algorithm is presented for non-linear state-space models driven by state-dependent noise. The algorithm is derived from first principles as an iterative solver for a quadratic composite minimization problem. This method mainly deals with an additive measurement noise processes for a common type outlier non-Gaussian phenomenon. A KF+ML combination algorithm is proposed in [14], where the maximum likelihood (ML) estimator is used for pre-localization of the target and measurement conversion to remove the measurement nonlinearity. The converted measurement and its associated noise statistics are then used in the standard Kalman filter for recursive update of the state. With this approach, the conversion of sensor measurements to Cartesian coordinates requires knowledge of range of the targets. The accuracy of converted measurements degrades tremendously when an inaccurate range is used. As such, this method is applicable mainly to low noise situations. The work presented in this article differs from these earlier works in the following ways: First, this work extends the filtering problem to situations with multiplicative observation noise rather than only additive noise. As such, the earlier

Manuscript received September 02, 2014; revised January 14, 2015; accepted March 29, 2015. Date of publication April 15, 2015; date of current version May 14, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Tareq Al-Naffouri. This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA06020201) and the National Natural Science Foundation of China (Grant No. 11174316, 11304345), and the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant RGPIN239031. (*Corresponding author: Ming Bao.*)

X. Hu, M. Bao and L. Guan are with the Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190 China (e-mail: auxqhu@gmail.com; baoming@mail.ioa.ac.cn; guanluyang@mail.ioa.ac.cn).

X.-P. Zhang is with the Department of Electrical and Computer Engineering, Ryerson University, ON M5B2K3 Canada (e-mail: xzhang@ee.ryerson.ca).

Y.-H. Hu is with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706-1691 USA (e-mail: hu@engr.wisc.edu).

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

Digital Object Identifier 10.1109/TSP.2015.2423266

works cannot be applicable to the multiplicative observation noise cases. Second, the multiple sensors in this work would collaborate to measure information on the object of interest for the state filtering. In addition, an iterative method is adopted to linearize the measurement function. Such a filter can reduce the effects of linearization errors and therefore obtain better estimation performance even in high noise situations.

Filtering with multiplicative measurement noise has also been addressed in several earlier works. In [23]–[25], minimal variance filters and polynomial filters are employed for linear systems with multiplicative, state dependent observation noise. In [26], a recursive unscented filtering algorithm is derived for state estimation in a class of nonlinear discrete-time stochastic systems with uncertain observation described by a multiplicative noise. The multiplicative noise is assumed to be a sequence of independent Bernoulli random variables to indicate the presence or absence of the signal in the observations. A recursive EKF is designed to minimize the filtering error covariance bound [27] for a discrete time-varying nonlinear system with stochastic nonlinearities and multiple missing measurements that reflect the multiplicative stochastic disturbances. The feasibility and convergence of the algorithm greatly depend on the scalar parameter in difference Riccati equations. A new generalized extended Kalman filtering algorithm using a multiplicative measurement noise model is developed in [28]. A Taylor series expansion linearization procedure conditioned on the predicted state is applied to convert nonlinear function into a linear formulation. The linearization process may introduce large errors in the true posterior mean and covariance of the state [29]. To reduce the errors due to linearization, a modified gain extended Kalman filter is proposed [30]. This method assumes that the nonlinear function is modifiable, which may not always be true in many realistic applications.

In this paper, a type of multiplicative measurement noise model is considered to facilitate more accurate characterisation of measurement errors of sensors. Specifically, it is assumed that the measurements of sensors are contaminated by both additive Gaussian noise and multiplicative Gaussian noise. With this nonlinear state-dependent noise model, the filter update requires the conditional statistics of the observation. That is, the correlation between the state and observation noise must be taken into account. In this work, we adopt a maximum *a posteriori* (MAP) estimation method to compute the updated state. An approximate MAP estimate can be obtained by an iteration that amounts to re-linearization of the measurement equation, and then an iterated Kalman filter is developed based on Gaussian approximation of the posterior distribution.

The main contribution of this paper is to present a generalized iterated Kalman filter (GIKF) for nonlinear stochastic discrete-time system with state-dependent multiplicative observation noise. Compared with the work in [22] that mainly deals with an additive measurement noise of state-dependent covariance, our work elaborates the theoretical relation between the generalized iterated Kalman filter and the generalized extended Kalman filter as well as the traditional extended Kalman filter. It is found that the generalized IKF yields a higher estimation accuracy than the generalized EKF and traditional EKF in the multiplicative observation noise model. The error performance of the GIKF including the mean square estimation error (MSE) and the Cramér-Rao lower bound (CRLB) is analyzed as well.

The remainder of the paper is organized as follows. In Section II, we review the filtering problem models of nonlinear systems. A generalized iterated Kalman filter with state-dependent multiplicative measurement noise is derived in Section III. In Section IV, the error performance of the GIKF is analyzed in terms of the MSE and the CRLB. In Section V, we conduct performance evaluations of the proposed algorithm by simulation comparisons in a tracking application involving rang measuring sensors. Section VI concludes the paper. For the ease of reading, some proofs are included in the Appendix.

II. Problem Models

We discuss the nonlinear filtering problem for the discrete-time stochastic signal system [28], [31], [32]:

$$\boldsymbol{x}_k = \boldsymbol{F}_k \boldsymbol{x}_{k-1} + \boldsymbol{G}_k \boldsymbol{w}_{k-1}, \tag{1}$$

$$z_n(k) = [1 + u_n(k)]\, h_n(\boldsymbol{x}_k) + v_n(k) \tag{2}$$

where $\boldsymbol{x}_k \in \mathbb{R}^{\aleph}$ is the state vector of the target at the k^{th} time step. The input noise $\boldsymbol{w}_{k-1}$ is a Gaussian random vector with zero mean and a covariance matrix $\boldsymbol{Q}_{k-1}$; $\boldsymbol{F}_k$ and $\boldsymbol{G}_k$ are the state transition matrix and input matrix respectively. We assume that N identical sensor nodes with the same noise statistics are deployed over a sensing field. The observation measurement obtained by sensor n is denoted by $z_n(k)$; $h_n(\boldsymbol{x}_k)$ is the measuring function of the target state; $u_n(k)$ and $v_n(k)$ are the multiplicative and additive Gaussian noise respectively. Let $\boldsymbol{\mathcal{L}}_k$ to be the set of indices of sensor nodes that have detected the target at time k. It is normally assumed that $\{\boldsymbol{w}_k, u_n(k), v_n(k), \boldsymbol{x}_0\}$ are mutually independent, and for $i, j \in \boldsymbol{\mathcal{L}}_k$

$$\mathbb{E}\left\{\boldsymbol{w}_k \boldsymbol{w}_l^{\mathrm{T}}\right\} = \boldsymbol{Q}_k \delta_{k,l}, \tag{3}$$

$$\mathbb{E}\left\{u_i(k)\right\} = \mu_u, \mathbb{E}\left\{v_i(k)\right\} = \mu_v, \tag{4}$$

$$\mathbb{E}\left\{(u_i(k) - \mu_u)\,(u_j(l) - \mu_u)\right\} = \sigma_u^2 \delta_{i,j} \delta_{k,l}, \tag{5}$$

$$\mathbb{E}\left\{(v_i(k) - \mu_v)\,(v_j(l) - \mu_v)\right\} = \sigma_v^2 \delta_{i,j} \delta_{k,l}, \tag{6}$$

where $\delta_{i,j} = 1$ if $i = j$, $\delta_{k,l} = 1$ if $k = l$; otherwise $\delta_{i,j} = 0$, $\delta_{k,l} = 0$.

Denote by ℓ_k the cardinal number of $\boldsymbol{\mathcal{L}}_k$ (i.e. the number of elements in $\boldsymbol{\mathcal{L}}_k$), the sensor measurements at the k^{th} time step may be represented in a matrix form:

$$\boldsymbol{z}_k = \boldsymbol{A}_k \cdot \boldsymbol{h}_k(\boldsymbol{x}_k) + \boldsymbol{v}_k \tag{7}$$

where $\boldsymbol{A}_k = \mathrm{diag}\{1 + u_n(k); 1 \le n \le \ell_k\}$, $\boldsymbol{h}_k(\boldsymbol{x}_k) = [h_1(\boldsymbol{x}_k), \ldots, h_{\ell_k}(\boldsymbol{x}_k)]^{\mathrm{T}}$, and $\boldsymbol{v}_k = [v_1(k), \ldots, v_{\ell_k}(k)]^{\mathrm{T}}$. Moreover, $\mathbb{E}\{\boldsymbol{v}_k\} = \mu_v \mathbf{1} = \mu_v [1, 1, \ldots, 1]^{\mathrm{T}}_{1 \times \ell_k}$, and $\boldsymbol{R}_k = \mathbb{E}\{(\boldsymbol{v}_k - \mu_v \mathbf{1})(\boldsymbol{v}_k - \mu_v \mathbf{1})^{\mathrm{T}}\} = \sigma_v^2 \mathbf{I}$.

The filtering problem is to obtain a sequential Bayesian estimate of $\boldsymbol{x}_k$ given the noisy sensor observations $\boldsymbol{z}_k = [z_1(k), \ldots, z_{\ell_k}(k)]^{\mathrm{T}}$, denoted by $\hat{\boldsymbol{x}}^t_{k|k}$, using an iterated Kalman filter formulation.

III. Generalized Iterated Kalman Filter (GIKF)

In this section, we focus on the derivation of the generalized iterated Kalman filter for the nonlinear stochastic signal system with multiplicative observation noise.

The basic filtering solution to the state estimation problem can be described as a two-stage recursive process of prediction and update. Now, given the distribution $p(\boldsymbol{x}_{k-1}|\boldsymbol{z}_{1:k-1}) = \mathcal{N}(\hat{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{P}_{k-1|k-1})$, the joint distribution of $\boldsymbol{x}_k$ and $\boldsymbol{x}_{k-1}$ conditioned on the previous measurements $\boldsymbol{z}_{1:k-1} = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_{k-1}\}$ is

$$\begin{aligned} &p(\boldsymbol{x}_{k-1}, \boldsymbol{x}_k|\boldsymbol{z}_{1:k-1}) = p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p(\boldsymbol{x}_{k-1}|\boldsymbol{z}_{1:k-1}) \\ &= \mathcal{N}\left(\begin{bmatrix} \hat{\boldsymbol{x}}_{k-1|k-1} \\ \boldsymbol{F}_k\hat{\boldsymbol{x}}_{k-1|k-1} \end{bmatrix}, \right. \\ &\left. \begin{bmatrix} \boldsymbol{P}_{k-1|k-1} & \boldsymbol{P}_{k-1|k-1}\boldsymbol{F}_k^{\mathrm{T}} \\ \boldsymbol{F}_k\boldsymbol{P}_{k-1|k-1} & \boldsymbol{F}_k\boldsymbol{P}_{k-1|k-1}\boldsymbol{F}_k^{\mathrm{T}} + \boldsymbol{G}_k\boldsymbol{Q}_{k-1}\boldsymbol{G}_k^{\mathrm{T}} \end{bmatrix}\right). \end{aligned} \tag{8}$$

The predictive distribution of $\boldsymbol{x}_k$ given the measurement history up to time step $k-1$ can be calculated by Lemma A1 in Appendix A

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1}) &= \mathcal{N}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1}) \\ &= \mathcal{N}\left(\boldsymbol{F}_k\hat{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{F}_k\boldsymbol{P}_{k-1|k-1}\boldsymbol{F}_k^{\mathrm{T}} \right. \\ &\quad \left. + \boldsymbol{G}_k\boldsymbol{Q}_{k-1}\boldsymbol{G}_k^{\mathrm{T}}\right). \end{aligned} \tag{9}$$

According to (7) and (4)~(6), the conditional probability density function (pdf) of the measurement $\boldsymbol{z}_k$, given $\boldsymbol{x}_k$, is written as follows:

$$p(\boldsymbol{z}_k|\boldsymbol{x}_k) = \mathcal{N}(\Xi_k, \boldsymbol{\Sigma}_k) \tag{10}$$

where,

$$\Xi_k = \mathbb{E}\{\boldsymbol{z}_k|\boldsymbol{x}_k\} = (1+\mu_u)\boldsymbol{h}_k(\boldsymbol{x}_k) + \mu_v\boldsymbol{1}, \tag{11}$$

$$\begin{aligned} \boldsymbol{\Sigma}_k &= \mathbb{E}\left\{(\boldsymbol{z}_k - \Xi_k)(\boldsymbol{z}_k - \Xi_k)^{\mathrm{T}}|\boldsymbol{x}_k\right\} \\ &= \boldsymbol{\sigma}_u^2\mathrm{diag}\{h_n^2(\boldsymbol{x}_k); 1 \le n \le \ell_k\} + \boldsymbol{R}_k. \end{aligned} \tag{12}$$

The derivation of (12) can be seen in Appendix B.

Using this measurement likelihood $p(\boldsymbol{z}_k|\boldsymbol{x}_k)$, we get the parameters of the posterior distribution by the Bayes rule

$$p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k}) = \beta_k p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1})p(\boldsymbol{z}_k|\boldsymbol{x}_k) \tag{13}$$

where the proportionality factor is $\beta_k^{-1} = p(\boldsymbol{z}_k|\boldsymbol{z}_{1:k-1})$, and

$$\begin{aligned} p(\boldsymbol{z}_k|\boldsymbol{x}_k) &= \frac{1}{(2\pi)^{\ell_k/2}(\det\{\boldsymbol{\Sigma}_k\})^{1/2}} \\ &\quad \cdot \exp\left\{-\frac{1}{2}(\boldsymbol{z}_k - \Xi_k)^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{z}_k - \Xi_k)\right\}, \\ p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1}) &= \frac{1}{(2\pi)^{\aleph/2}\left(\det\{\boldsymbol{P}_{k|k-1}\}\right)^{\frac{1}{2}}} \\ &\quad \cdot \exp\left\{-\frac{1}{2}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}}\boldsymbol{P}_{k|k-1}^{-1}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1})\right\}. \end{aligned}$$

The MAP estimate associated to the posterior $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$ is identical to minimizing its negative logarithm:

$$\begin{aligned} L(\boldsymbol{x}_k) &= \frac{1}{2}\ln\{\det\{\boldsymbol{\Sigma}_k\}\} + \frac{1}{2}\mathfrak{F}_k^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1}\mathfrak{F}_k \\ &\quad + \frac{1}{2}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}}\boldsymbol{P}_{k|k-1}^{-1}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}). \end{aligned} \tag{14}$$

The above equation ignores the constant terms not dependent on $\boldsymbol{x}_k$, and $\mathfrak{F}_k = \boldsymbol{z}_k - \Xi_k$. The MAP estimate can be found by solving the optimization problem

$$\hat{\boldsymbol{x}}_{k|k} = \arg\min_{\boldsymbol{x}_k \in \mathbb{R}^{\aleph}} L(\boldsymbol{x}_k). \tag{15}$$

The minimization problem of (15) can be equivalently stated as a nonlinear unconstrained optimization problem, the common solution to the optimization problem is found through the following Newton-Raphson iterative method:

$$\hat{\boldsymbol{x}}_{k|k}^{t+1} = \hat{\boldsymbol{x}}_{k|k}^{t} - \left[\frac{\partial^2 L(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k^{\mathrm{T}}\partial \boldsymbol{x}_k}\bigg|_{\boldsymbol{x}_k=\hat{\boldsymbol{x}}_{k|k}^t}\right]^{-1} \frac{\partial L(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k}\bigg|_{\boldsymbol{x}_k=\hat{\boldsymbol{x}}_{k|k}^t} \tag{16}$$

where t is the iteration step, and the initial value $\hat{\boldsymbol{x}}_{k|k}^0 = \hat{\boldsymbol{x}}_{k|k-1}$. Equivalently, the MAP estimation problem is a nonlinear least square problem $L(\boldsymbol{x}_k) = \frac{1}{2}\boldsymbol{r}^{\mathrm{T}}(\boldsymbol{x}_k)\boldsymbol{r}(\boldsymbol{x}_k)$, where $\boldsymbol{r}(\boldsymbol{x}_k) = [(\ln\{\det\{\boldsymbol{\Sigma}_k\}\})^{1/2}\ \mathfrak{F}_k^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1/2}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}}\boldsymbol{P}_{k|k-1}^{-1/2}]^{\mathrm{T}}$. Unlike Newton-Raphson method, the Gauss-Newton algorithm can only be used to solve the nonlinear least square problem. It approximates the Hessian matrix of $L(\boldsymbol{x}_k)$ by ignoring the second-order derivative terms of the residuals $\boldsymbol{r}^{\mathrm{T}}\frac{\partial^2 \boldsymbol{r}}{\partial x^j \partial x^i}$. As a consequence, the convergence of the Gauss-Newton method depends on whether the omitted second-order derivative terms of the residuals are large parts of the Hessian. When $\left|\frac{\partial \boldsymbol{r}^{\mathrm{T}}}{\partial x^j}\frac{\partial \boldsymbol{r}^{\mathrm{T}}}{\partial x^i}\right| \gg \left|\boldsymbol{r}^{\mathrm{T}}\frac{\partial^2 \boldsymbol{r}}{\partial x^j \partial x^i}\right|$ cannot be satisfied, it is shown that the Gauss-Newton method may not be locally convergent at all [33].

Some computation facts about the derivatives and traces of a matrix are given to facilitate the calculation of derivatives of the function $L(\boldsymbol{x}_k)$. Let $\boldsymbol{P}$ be a nonsingular symmetric matrix dependent on a scalar x, and let $\boldsymbol{x}$ and $\boldsymbol{y}$ be vectors. All vectors will be regarded as column vectors in the context. Then [34]

$$\frac{\partial \det\{\boldsymbol{P}\}}{\partial x} = \det\{\boldsymbol{P}\}\mathrm{tr}\left\{\boldsymbol{P}^{-1}\frac{\partial \boldsymbol{P}}{\partial x}\right\}, \frac{\partial \boldsymbol{P}^{-1}}{\partial x} = -\boldsymbol{P}^{-1}\frac{\partial \boldsymbol{P}}{\partial x}\boldsymbol{P}^{-1}, \tag{17}$$

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{y} = \mathrm{tr}\{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{y}\} = \mathrm{tr}\{\boldsymbol{P}\boldsymbol{y}\boldsymbol{x}^{\mathrm{T}}\} = \mathrm{tr}\{\boldsymbol{P}\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}\}. \tag{18}$$

Using (17) and (18), the ith component of the gradient of the logarithm function $L(\boldsymbol{x}_k)$ with respect to $x^i \in \boldsymbol{x}_k$ is given as

$$\begin{aligned} \frac{\partial L(\boldsymbol{x}_k)}{\partial x^i} &= \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{P}_{k|k-1}^{-1}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}) + \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Sigma}_k}{\partial x^i}\right\} \\ &\quad + \frac{1}{2}\mathrm{tr}\left\{-\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Sigma}_k}{\partial x^i}\boldsymbol{\Sigma}_k^{-1}\mathfrak{F}_k\mathfrak{F}_k^{\mathrm{T}} + 2\frac{\partial \mathfrak{F}_k}{\partial x^i}\boldsymbol{\Sigma}_k^{-1}\mathfrak{F}_k\right\} \end{aligned} \tag{19}$$

where $\boldsymbol{e}_i$ is a basis vector with a 1 in the ith element and 0's elsewhere. Then, the element at the ith row and jth column of the Hessian matrix of the function $L(\boldsymbol{x}_k)$ can be computed as

$$\begin{aligned}\frac{\partial L(\boldsymbol{x}_k)}{\partial x^j \partial x^i} &= \boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{P}_{k|k-1}^{-1} \boldsymbol{e}_j \\ &+ \frac{1}{2}\mathrm{tr}\left\{-\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Sigma}_k}{\partial x^j}\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Sigma}_k}{\partial x^i} + \boldsymbol{\Sigma}_k^{-1}\frac{\partial^2 \boldsymbol{\Sigma}_k}{\partial x^j \partial x^i}\right\} \\ &+ \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Sigma}_k}{\partial x^j}\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Sigma}_k}{\partial x^i}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Im}_k\boldsymbol{\Im}_k^{\mathrm{T}}\right\} \\ &- \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}_k^{-1}\frac{\partial^2 \boldsymbol{\Sigma}_k}{\partial x^j \partial x^i}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Im}_k\boldsymbol{\Im}_k^{\mathrm{T}}\right\} \\ &+ \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Sigma}_k}{\partial x^i}\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Sigma}_k}{\partial x^j}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Im}_k\boldsymbol{\Im}_k^{\mathrm{T}}\right\} \\ &- \frac{1}{2}\mathrm{tr}\left\{2\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Sigma}_k}{\partial x^i}\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Im}_k}{\partial x^j}\boldsymbol{\Im}_k^{\mathrm{T}}\right\} + \frac{1}{2}\mathrm{tr}\left\{2\frac{\partial^2 \boldsymbol{\Im}_k^{\mathrm{T}}}{\partial x^j \partial x^j}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Im}_k^{\mathrm{T}}\right\} \\ &- \frac{1}{2}\mathrm{tr}\left\{2\frac{\partial \boldsymbol{\Im}_k^{\mathrm{T}}}{\partial x^i}\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Sigma}_k}{\partial x^j}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Im}_k\right\} + \frac{1}{2}\mathrm{tr}\left\{2\frac{\partial \boldsymbol{\Im}_k^{\mathrm{T}}}{\partial x^i}\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Im}_k}{\partial x^j}\right\}. \end{aligned} \tag{20}$$

The joint distribution of $\boldsymbol{x}_k$ and $\boldsymbol{z}_{1:k}$ is

$$\begin{aligned} p(\boldsymbol{x}_k, \boldsymbol{z}_{1:k}) &= p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})p(\boldsymbol{z}_{1:k}) \\ &= p(\boldsymbol{z}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1})p(\boldsymbol{z}_{1:k-1}). \end{aligned} \tag{21}$$

From (21) and (13), the derivative of negative logarithm function of $p(\boldsymbol{x}_k, \boldsymbol{z}_{1:k})$ with respect to $\boldsymbol{x}_k$ is also equal to $\boldsymbol{\Gamma}_k(\boldsymbol{x}_k) = \frac{\partial L(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k}$ which is called the score. Denote $\boldsymbol{K}_k(\boldsymbol{x}_k) = \frac{\partial^2 L(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k^{\mathrm{T}} \partial \boldsymbol{x}_k}$. The Fisher information matrix for the estimated parameter $\boldsymbol{x}_k$ is defined as [35]

$$\boldsymbol{J}_k = \mathbb{E}\left\{\boldsymbol{K}_k(\boldsymbol{x}_k)\right\}.$$

Note that the measurement function $\boldsymbol{h}_k(\boldsymbol{x}_k)$ in (7) is a nonlinear map. For most nonlinear models, the closed-form analytic expression of the posterior distributions $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$ does not exist in general [36]. A workaround is to adopt a Gaussian approximation of the filtering (posterior) distribution $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$ [21]. Moreover, the approximated posterior error covariance is equal to the inverse of the Fisher information matrix [37]. Therefore we have the following proposition.

Proposition 1: Given the discrete system of (1) and (7), let the error covariance of the state estimate $\hat{\boldsymbol{x}}_{k|k}^t$ be defined as $\boldsymbol{P}_{k|k}^t = \mathbb{E}\{(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}^t)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}^t)^{\mathrm{T}}\}$. If the posterior distribution is $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k}) = \mathcal{N}(\hat{\boldsymbol{x}}_{k|k}^t, \boldsymbol{P}_{k|k}^t)$, then the error covariance associated with $\hat{\boldsymbol{x}}_{k|k}^t$ can be obtained as

$$\begin{aligned} \boldsymbol{P}_{k|k}^t &= \left[\boldsymbol{J}_k|_{\boldsymbol{x}_k=\hat{\boldsymbol{x}}_{k|k}^t}\right]^{-1} \\ &= \left[\boldsymbol{H}_k^{\mathrm{T}}\left(\hat{\boldsymbol{x}}_{k|k}^t\right)\boldsymbol{B}_k\left(\hat{\boldsymbol{x}}_{k|k}^t\right)\boldsymbol{H}_k\left(\hat{\boldsymbol{x}}_{k|k}^t\right) + \boldsymbol{P}_{k|k-1}^{-1}\right]^{-1} \end{aligned} \tag{22}$$

where $\boldsymbol{H}_k(\hat{\boldsymbol{x}}_{k|k}^t) = \frac{\partial \boldsymbol{h}_k(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k^{\mathrm{T}}}|_{\boldsymbol{x}_k=\hat{\boldsymbol{x}}_{k|k}^t}$ is the Jacobian of the re-linearized measurement function, and $\boldsymbol{B}_k(\hat{\boldsymbol{x}}_{k|k}^t) = \mathrm{diag}\{b_n(k)|_{\boldsymbol{x}_k=\hat{\boldsymbol{x}}_{k|k}^t}; 1 \le n \le \ell_k\}$ is defined in (25).

Proof: Given $\boldsymbol{x}_k$, then $\mathbb{E}\{\boldsymbol{\Im}_k\} = \boldsymbol{0}$ and $\mathbb{E}\{\boldsymbol{\Im}_k\boldsymbol{\Im}_k^{\mathrm{T}}\} = \boldsymbol{\Sigma}_k$. From (20), the information matrix equals

$$\boldsymbol{J}_k|_{\boldsymbol{x}_k} = \boldsymbol{P}_{k|k-1}^{-1} + \boldsymbol{M}_k, \tag{23}$$

the element at the ith row and jth column of $\boldsymbol{M}_k$ is given by m_{ij},

$$\begin{aligned} m_{ij} &= \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Sigma}_k}{\partial x^i}\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Sigma}_k}{\partial x^j}\right\} + \mathrm{tr}\left\{\frac{\partial \boldsymbol{\Im}_k^{\mathrm{T}}}{\partial x^i}\boldsymbol{\Sigma}_k^{-1}\frac{\partial \boldsymbol{\Im}_k}{\partial x^j}\right\} \\ &= \sum_{n=1}^{n=\ell_k}\frac{2\sigma_u^4 h_n^2(\boldsymbol{x}_k)}{\left(\sigma_u^2 h_n^2(\boldsymbol{x}_k) + \sigma_v^2\right)^2}\frac{\partial h_n(\boldsymbol{x}_k)}{\partial x^i}\frac{\partial h_n(\boldsymbol{x}_k)}{\partial x^j} \\ &\quad + \sum_{n=1}^{n=\ell_k}\frac{(1+\mu_u)^2}{\sigma_u^2 h_n^2(\boldsymbol{x}_k) + \sigma_v^2}\frac{\partial h_n(\boldsymbol{x}_k)}{\partial x^i}\frac{\partial h_n(\boldsymbol{x}_k)}{\partial x^j}. \end{aligned} \tag{24}$$

Let

$$b_n(k) = \frac{2\sigma_u^4 h_n^2(\boldsymbol{x}_k)}{\left(\sigma_u^2 h_n^2(\boldsymbol{x}_k) + \sigma_v^2\right)^2} + \frac{(1+\mu_u)^2}{\sigma_u^2 h_n^2(\boldsymbol{x}_k) + \sigma_v^2}. \tag{25}$$

And then m_{ij} can be transformed as

$$m_{ij} = \left[\frac{\partial \boldsymbol{h}_k(\boldsymbol{x}_k)}{\partial x^i}\right]^{\mathrm{T}}\boldsymbol{B}_k(\boldsymbol{x}_k)\frac{\partial \boldsymbol{h}_k(\boldsymbol{x}_k)}{\partial x^j}.$$

Thus, we can get

$$\begin{aligned} \boldsymbol{M}_k &= \left[\frac{\partial \boldsymbol{h}_k(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k^{\mathrm{T}}}\right]^{\mathrm{T}}\boldsymbol{B}_k(\boldsymbol{x}_k)\frac{\partial \boldsymbol{h}_k(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k^{\mathrm{T}}} \\ &= \boldsymbol{H}_k^{\mathrm{T}}(\boldsymbol{x}_k)\boldsymbol{B}_k(\boldsymbol{x}_k)\boldsymbol{H}_k(\boldsymbol{x}_k) \end{aligned} \tag{26}$$

where $\boldsymbol{H}_k(\boldsymbol{x}_k) = \frac{\partial \boldsymbol{h}_k(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k^{\mathrm{T}}}$. Since $\boldsymbol{B}_k(\boldsymbol{x}_k)$ is a symmetric positive-definite diagonal matrix, $\boldsymbol{M}_k$ is positive semi-definite. Note that $\boldsymbol{P}_{k|k-1}^{-1}$ is positive, so $\boldsymbol{P}_{k|k-1}^{-1} + \boldsymbol{M}_k$ is nonsingular. Taking $\boldsymbol{x}_k = \hat{\boldsymbol{x}}_{k|k}^t$ at right side of (26), then we can easily conclude the theorem. ∎

Note that if the measurement noise is uncorrelated with the target state, that is the multiplicative noise u_n is ignored (in this case we can assume $\mu_u = 0$, $\sigma_u^2 = 0$). The observation function is approximated by the first order Taylor series in the standard extended (or iterated) Kalman filter derivation. It is consistent with Gauss-Newton method truncating the second-order derivative terms of residuals. As such retaining only up to the first derivative of $\Im_k$, (16) simplifies to the following update formula in the traditional iterated KF for the case of additive state-independent observation covariance [38],

$$\begin{aligned} \hat{\boldsymbol{x}}_{k|k}^{t+1} &= \hat{\boldsymbol{x}}_{k|k}^t + \left[\boldsymbol{H}_k^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k|k}^t)\boldsymbol{R}_k^{-1}\boldsymbol{H}_k(\hat{\boldsymbol{x}}_{k|k}^t) + \boldsymbol{P}_{k|k-1}^{-1}\right]^{-1} \\ &\quad \cdot \left[\boldsymbol{H}_k^{\mathrm{T}}\left(\hat{\boldsymbol{x}}_{k|k}^t\right)\boldsymbol{R}_k^{-1}\left(\boldsymbol{z}_k - \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k}^t\right)\right)\right. \\ &\qquad \left. -\boldsymbol{P}_{k|k-1}^{-1}\left(\hat{\boldsymbol{x}}_{k|k}^t - \hat{\boldsymbol{x}}_{k|k-1}\right)\right] \\ &= \hat{\boldsymbol{x}}_{k|k-1} + \left[\boldsymbol{H}_k^{\mathrm{T}}\left(\hat{\boldsymbol{x}}_{k|k}^t\right)\boldsymbol{R}_k^{-1}\boldsymbol{H}_k\left(\hat{\boldsymbol{x}}_{k|k}^t\right) + \boldsymbol{P}_{k|k-1}^{-1}\right]^{-1} \\ &\quad \cdot \boldsymbol{H}_k^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k|k}^t)\boldsymbol{R}_k^{-1}\left(\boldsymbol{z}_k - \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k}^t\right)\right. \\ &\qquad \left. -\boldsymbol{H}_k\left(\hat{\boldsymbol{x}}_{k|k}^t\right)\left(\hat{\boldsymbol{x}}_{k|k-1} - \hat{\boldsymbol{x}}_{k|k}^t\right)\right). \end{aligned} \tag{27}$$

The corresponding error covariance in the traditional iterated KF is, from (22), given by

$$\boldsymbol{P}_{k|k}^{t}=\left[\boldsymbol{H}_{k}^{\mathrm{T}}\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)\boldsymbol{R}_{k}^{-1}\boldsymbol{H}_{k}\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)+\boldsymbol{P}_{k|k-1}^{-1}\right]^{-1}. \quad (28)$$

For a single step iteration, setting $t=0$ in (27) and (28) above, we obtain the standard EKF update formula in the case of additive state-independent observation covariance

$$\hat{\boldsymbol{x}}_{k|k}=\hat{\boldsymbol{x}}_{k|k-1}+\left[\boldsymbol{H}_{k}^{\mathrm{T}}\left(\hat{\boldsymbol{x}}_{k|k-1}\right)\boldsymbol{R}_{k}^{-1}\boldsymbol{H}_{k}(\hat{\boldsymbol{x}}_{k|k-1})+\boldsymbol{P}_{k|k-1}^{-1}\right]^{-1}\cdot\boldsymbol{H}_{k}^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k|k-1})\boldsymbol{R}_{k}^{-1}\left(\boldsymbol{z}_{k}-\boldsymbol{h}_{k}(\hat{\boldsymbol{x}}_{k|k-1})\right), \quad (29)$$

$$\boldsymbol{P}_{k|k}=\left[\boldsymbol{H}_{k}^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k|k-1})\boldsymbol{R}_{k}^{-1}\boldsymbol{H}_{k}(\hat{\boldsymbol{x}}_{k|k-1})+\boldsymbol{P}_{k|k-1}^{-1}\right]^{-1}. \quad (30)$$

IV. Error Performance of GIKF

In this section, the error performance of GIKF is analyzed in terms of the MSE and the CRLB to evaluate the effectiveness of the GIKF, and then the MSE of GIKF is compared with a generalized extended Kalman filter that uses Taylor series expansion to linearize the measurement function.

A. Cramér-Rao Lower Bound

The CRLB that provides a lower bound on the MSE of the state estimate is widely used to assess the performance of an estimator. The CRLB on the error covariance matrix of the state estimate has the form [35],

$$\boldsymbol{P}_{k|k}^{t}\geq[\boldsymbol{\mathcal{J}}_{k}]^{-1}, \quad (31)$$

where $\boldsymbol{\mathcal{J}}_{k}$ is the Fisher information matrix evaluated at the true state $\boldsymbol{x}_{k}$. It is different with $\boldsymbol{J}_{k}$ in (22) evaluated at the estimated state $\hat{\boldsymbol{x}}_{k|k}^{t}$.

An efficient method for computing $\boldsymbol{\mathcal{J}}_{k}$ recursively is given in [35].

Proposition 2: Given the discrete system of (1) and (7), the CRLB can be recursively calculated by

$$\boldsymbol{\mathcal{J}}_{k}=(\boldsymbol{G}_{k}\boldsymbol{Q}_{k-1}\boldsymbol{G}_{k}^{\mathrm{T}})^{-1}+\boldsymbol{M}_{k}-(\boldsymbol{G}_{k}\boldsymbol{Q}_{k-1}\boldsymbol{G}_{k}^{\mathrm{T}})^{-1}\boldsymbol{F}_{k}\cdot\left[\boldsymbol{\mathcal{J}}_{k-1}+\boldsymbol{F}_{k}^{\mathrm{T}}\left(\boldsymbol{G}_{k}\boldsymbol{Q}_{k-1}\boldsymbol{G}_{k}^{\mathrm{T}}\right)^{-1}\boldsymbol{F}_{k}\right]^{-1}\boldsymbol{F}_{k}^{\mathrm{T}}\left(\boldsymbol{G}_{k}\boldsymbol{Q}_{k-1}\boldsymbol{G}_{k}^{\mathrm{T}}\right)^{-1}. \quad (32)$$

The derivation of Proposition 2 can be found in Appendix C. According to the matrix inversion lemma, the CRLB in (32) can be rewritten

$$\boldsymbol{\mathcal{J}}_{k}=\boldsymbol{M}_{k}+\left[\boldsymbol{G}_{k}\boldsymbol{Q}_{k-1}\boldsymbol{G}_{k}^{\mathrm{T}}+\boldsymbol{F}_{k}\boldsymbol{\mathcal{J}}_{k-1}^{-1}\boldsymbol{F}_{k}^{\mathrm{T}}\right]^{-1}. \quad (33)$$

B. Convergence of MSE in GIKF

The trace (sum of diagonal elements) of the covariance matrix $\mathrm{tr}\{\boldsymbol{P}_{k|k}\}=\mathbb{E}\{\|\hat{\boldsymbol{x}}_{k|k}-\boldsymbol{x}_{k}\|^{2}\}$ corresponds to the MSE of the updated state $\hat{\boldsymbol{x}}_{k|k}$ and can be used to measure the estimation accuracy [28], [39]. In general, smaller $\mathrm{tr}\{\boldsymbol{P}_{k|k}\}$ value implies more accurate state estimate.

From Proposition 1, we can find that there always exists an extremum $\hat{\boldsymbol{x}}_{k|k}^{t}$ for which $\boldsymbol{K}_{k}(\hat{\boldsymbol{x}}_{k|k}^{t})$ is positive definite. Since $\boldsymbol{K}_{k}(\boldsymbol{x}_{k})$ is a continuously differentiable function, we give the following lemma:

Lemma 1: $\boldsymbol{K}_{k}(\boldsymbol{x}_{k})$ is locally Lipschitz-continuous, that is there exists a scalar $\gamma>0$ with arbitrary $\varepsilon>0$ such that $\|\boldsymbol{K}_{k}(\boldsymbol{x})-\boldsymbol{K}_{k}(\boldsymbol{y})\|\leq\gamma\|\boldsymbol{x}-\boldsymbol{y}\|$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$ satisfying $\|\boldsymbol{x}-\hat{\boldsymbol{x}}_{k|k}^{t}\|\leq\varepsilon$ and $\|\boldsymbol{y}-\hat{\boldsymbol{x}}_{k|k}^{t}\|\leq\varepsilon$ [40].

Since $\boldsymbol{K}_{k}(\hat{\boldsymbol{x}}_{k|k}^{t})$ is positive, by the continuity of $\boldsymbol{K}_{k}(\boldsymbol{x}_{k})$, we can find a neighborhood of $\hat{\boldsymbol{x}}_{k|k}^{t}$ such that $\boldsymbol{K}_{k}(\boldsymbol{x}_{k})$, $\|\boldsymbol{x}_{k}-\hat{\boldsymbol{x}}_{k|k}^{t}\|\leq\varepsilon$, is positive. Therefore, we have another lemma:

Lemma 2: When $\|\boldsymbol{x}_{k}-\hat{\boldsymbol{x}}_{k|k}^{t}\|\leq\varepsilon$, there exists a scalar η for which $\|\boldsymbol{K}_{k}(\boldsymbol{x}_{k})^{-1}\|\leq1/\eta$, it satisfies the following inequality [41]

$$\left\|\boldsymbol{K}_{k}\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)^{-1}\right\|\leq\frac{1}{\eta-\gamma\left\|\boldsymbol{x}_{k}-\hat{\boldsymbol{x}}_{k|k}^{t}\right\|}. \quad (34)$$

Theorem 1: Suppose $\hat{\boldsymbol{x}}_{k|k}^{t}$ satisfies $\|\boldsymbol{x}_{k}-\hat{\boldsymbol{x}}_{k|k}^{t}\|\leq\rho$, where $\rho=\min\{\alpha\varepsilon,2\alpha\eta/3\gamma\}$ and $0<\alpha<1$, then $\mathbb{E}\{\|\hat{\boldsymbol{x}}_{k|k}^{t+1}-\boldsymbol{x}_{k}\|^{2}\}<\mathbb{E}\{\|\hat{\boldsymbol{x}}_{k|k}^{t}-\boldsymbol{x}_{k}\|^{2}\}$ so that $\mathrm{tr}\{\boldsymbol{P}_{k|k}^{t+1}\}<\mathrm{tr}\{\boldsymbol{P}_{k|k}^{t}\}$ in GIKF.

The detailed proof of above theorem is shown in Appendix D.

Corollary 1: Suppose $\hat{\boldsymbol{x}}_{k|k}^{t}$ satisfies $\|\boldsymbol{x}_{k}-\hat{\boldsymbol{x}}_{k|k}^{t}\|\leq\rho$, then the estimation accuracy in GIKF converges to a limit, that is

$$\lim_{t\to\infty}\mathrm{tr}\left\{\boldsymbol{P}_{k|k}^{t}\right\}=\lim_{t\to\infty}\mathrm{tr}\left\{\boldsymbol{P}_{k|k}^{t+1}\right\}.$$

Proof: From Theorem 1, we know that $\{\mathrm{tr}\{\boldsymbol{P}_{k|k}^{t}\}\}$ for $t=0,1,2,\ldots$ is a monotonous decreasing sequence and it is also bounded by $\mathrm{tr}\{[\boldsymbol{\mathcal{J}}_{k}]^{-1}\}\leq\mathrm{tr}\{\boldsymbol{P}_{k|k}^{t}\}\leq\mathrm{tr}\{\boldsymbol{P}_{k|k-1}\}$ from (31), so according to the monotone convergence theorem, the limit of the sequence exists, i.e., $\lim_{t\to\infty}\mathrm{tr}\{\boldsymbol{P}_{k|k}^{t}\}=\lim_{t\to\infty}\mathrm{tr}\{\boldsymbol{P}_{k|k}^{t+1}\}=\kappa_{k}$, where $\kappa_{k}>0$ is constant at time k. ■

From the above corollary, it can be found that $\lim_{t\to\infty}\mathbb{E}\{\|\hat{\boldsymbol{x}}_{k|k}^{t}-\boldsymbol{x}_{k}\|^{2}\}\neq0$ since $\lim_{t\to\infty}\mathrm{tr}\{\boldsymbol{P}_{k|k}^{t}\}>0$, thereby $\hat{\boldsymbol{x}}_{k|k}^{t}$ can not converge to $\boldsymbol{x}_{k}$ at each time step and the only claim is that $\lim_{t\to\infty}\hat{\boldsymbol{x}}_{k|k}^{t}=\lim_{t\to\infty}\hat{\boldsymbol{x}}_{k|k}^{t+1}$, which means that convergence is guaranteed as the iterate proceeds.

Note that the results regarding the convergence in the above theorems are local, they apply only if the starting point is sufficiently close to the desired limit. In the GIKF, the initial estimate of the iterative process comes from the Kalman state prediction which can assist nice convergence, and thus the sampling time interval is not too long for convergence. Typically, only a few iterations are sufficient to get the minimum, because the longer iteration steps cannot promise more accurate state estimate.

C. Comparison to Generalized Extended Kalman Filter

For the filtering problem with state-dependent multiplicative observation noise, our early work presented a generalized extended Kalman filter (GEKF) that adopts a first order Taylor series expansion linearization method for forming a Gaussian approximation of the posterior probability. The measurement model can be then approximated around $\hat{\boldsymbol{x}}_{k|k-1}$

$$\boldsymbol{z}_{k}\simeq\boldsymbol{A}_{k}\left[\boldsymbol{h}_{k}(\hat{\boldsymbol{x}}_{k|k-1})+\boldsymbol{H}_{k}(\hat{\boldsymbol{x}}_{k|k-1})(\boldsymbol{x}_{k}-\hat{\boldsymbol{x}}_{k|k-1})\right]+\boldsymbol{v}_{k}. \quad (35)$$

Thus, the expression of GEKF based on the discrete system of (1) and (35) is briefly recalled below [28]:

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{C}_E(k)\boldsymbol{S}_E(k)^{-1}\left[\boldsymbol{z}_k - \Xi_E(k)\right], \quad (36)$$

$$\boldsymbol{P}_{k|k} = \boldsymbol{P}_{k|k-1} - \boldsymbol{C}_E(k)\boldsymbol{S}_E(k)^{-1}\boldsymbol{C}_E^{\mathrm{T}}(k). \quad (37)$$

where,

$$\begin{aligned}\Xi_E(k) &= (1+\mu_u)\boldsymbol{h}_k(\hat{\boldsymbol{x}}_{k|k-1}) + \mu_v\mathbf{1},\\ \boldsymbol{C}_E(k) &= (1+\mu_u)\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k|k-1}),\\ \boldsymbol{S}_E(k) &= (1+\mu_u)^2\boldsymbol{H}_k(\hat{\boldsymbol{x}}_{k|k-1})\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k|k-1})\\ &\quad + \sigma_u^2\boldsymbol{\Theta}_k(\hat{\boldsymbol{x}}_{k|k-1}) + \sigma_v^2\mathbf{I}\end{aligned}$$

where $\boldsymbol{\Theta}_k(\hat{\boldsymbol{x}}_{k|k-1}) = \mathrm{diag}\{\boldsymbol{H}_k(\hat{\boldsymbol{x}}_{k|k-1})\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k|k-1}) + \boldsymbol{h}_k(\hat{\boldsymbol{x}}_{k|k-1})\boldsymbol{h}_k^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k|k-1})\}$ is a diagonal matrix consisting of the diagonal elements of the matrix inside the brackets.

With the Gaussian approximation of the posterior probability, the estimation accuracy comparison between GEKF and GIKF can be analytically summarized below.

Proposition 3: Given the discrete system of (1) and (7), the GIKF yields more accurate estimate than the GEKF, that is

$$\mathrm{tr}\{\boldsymbol{P}_{k|k}^t\} < \mathrm{tr}\{\boldsymbol{P}_{k|k}\}. \quad (38)$$

The proof of Proposition 3 is included in Appendix E.

Remark 1: It is known that when the observation equation is nonlinear, the traditional IKF with a single iteration obtains the same estimation performance with the traditional EKF in the case of additive state-independent observation noise [42], or in the case of additive state-dependent observation noise[22]. However, from Proposition 3, we can find that the generalized IKF with a single iteration yields a higher estimation accuracy than the generalized EKF in the multiplicative observation noise model.

Remark 2: The traditional EKF always assumes that the measurement noise is independent of the state, its covariance is unconditional. The multiplicative noise in (7) can be converted into an equivalent additive signal-dependent noise by a first-order linear approximation of the non-linear measurement model as follows

$$\boldsymbol{z}_k \simeq \boldsymbol{h}_k(\hat{\boldsymbol{x}}_{k|k-1}) + \boldsymbol{H}_k(\hat{\boldsymbol{x}}_{k|k-1})(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}) + \boldsymbol{v}_k'$$

where $\boldsymbol{v}_k' = (\boldsymbol{A}_k - \mathbf{I})[\boldsymbol{h}_k(\hat{\boldsymbol{x}}_{k|k-1}) + \boldsymbol{H}_k(\hat{\boldsymbol{x}}_{k|k-1})(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1})] + \boldsymbol{v}_k$ is additive noise with mean value $\Xi_E(k)$ and covariance $\boldsymbol{R}_k' = \mu_u^2\boldsymbol{H}_k(\hat{\boldsymbol{x}}_{k|k-1})\ \boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k|k-1}) + \sigma_u^2\boldsymbol{\Theta}_k(\hat{\boldsymbol{x}}_{k|k-1}) + \sigma_v^2\mathbf{I}$. This leads to the state update equations in the traditional EKF formulation [28]

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{C}'_E(k)\boldsymbol{S}'_E(k)^{-1}\left[\boldsymbol{z}_k - \Xi_E(k)\right], \quad (39)$$

$$\begin{aligned}\boldsymbol{P}_{k|k} &= \boldsymbol{P}_{k|k-1} - \boldsymbol{C}_E(k)\boldsymbol{S}'_E(k)^{-1}\boldsymbol{C}'^{\mathrm{T}}_E(k)\\ &\quad - \boldsymbol{C}'_E(k)\boldsymbol{S}'_E(k)^{-1}\boldsymbol{C}_E^{\mathrm{T}}(k)\\ &\quad + \boldsymbol{C}'_E(k)\boldsymbol{S}'_E(k)^{-1}\boldsymbol{S}_E(k)\boldsymbol{S}'_E(k)^{-1}\boldsymbol{C}'^{\mathrm{T}}_E(k). \quad (40)\end{aligned}$$

where,

$$\begin{aligned}\boldsymbol{C}'_E(k) &= \boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k|k-1}),\\ \boldsymbol{S}'_E(k) &= \boldsymbol{H}_k(\hat{\boldsymbol{x}}_{k|k-1})\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k|k-1}) + \boldsymbol{R}_k'.\end{aligned}$$

It is verified that the traditional EKF using the additive noise yields bigger posterior error covariance matrix than the GEKF [31]. From Proposition 3, the generalized IKF can also obtain higher accuracy than the traditional EKF under the multiplicative measurement model in (7). In addition, it is worth pointing out that, for the problem of state estimation problem, the implementations of the traditional Kalman filter requires $O(\aleph^3)$ operations. Note that computing of the inverses in the covariance in (50) for GEKF requires $O(\aleph^3)$ operations in general. Similarly, for each instant k, if the iteration is repeated t times, then the computational complexity of the GIKF is comparable to that of the traditional IKF and is in general $O(t\aleph^3)$ [43]. On average, the computational complexities of GIKF and GEKF are in the same order of magnitude if t is not too large.

V. Numerical Simulations

To show the efficiency of the proposed GIKF, it is applied to a target tracking application in comparison with the GEKF, and the traditional EKF (TEKF) as well as a two-phase KF+ML algorithm proposed in [14]. Error performance of the filters is evaluated with Monte Carlo simulations in this section.

For simplicity, we only consider a single target tracking scenario, but nevertheless our proposed GIKF algorithm still is applicable to multi-target tracking as the targets have been classified. The constant velocity model is adopted to represent the motion of the target in a 2-D space with $\boldsymbol{x}_k = [x(k)\ \dot{x}(k)\ y(k)\ \dot{y}(k)]^{\mathrm{T}}$, where $(x(k), y(k))$ and $(\dot{x}(k), \dot{y}(k))$ are the position coordinates and velocities of the target along $x-$and $y-$directions at k^{th} time step, respectively. The state transition matrix and input matrix are

$$F_k = \begin{bmatrix}1 & \Delta_k & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & \Delta_k\\ 0 & 0 & 0 & 1\end{bmatrix}, G_k = \begin{bmatrix}\frac{\Delta_k^2}{2} & 0\\ \Delta_k & 0\\ 0 & \frac{\Delta_k^2}{2}\\ 0 & \Delta_k\end{bmatrix}, \quad (41)$$

respectively, where Δ_k is the sampling time interval between two successive time step k and $k-1$. Denote by (x_n, y_n) the known location of the n^{th} sensor. The measuring function for a sensor is given by

$$h_n(\boldsymbol{x}_k) = \sqrt{(x(k)-x_n)^2 + (y(k)-y_n)^2}. \quad (42)$$

Suppose that four sensor nodes are deployed at the four corners of a 2 meters by 2 meters sensing field, each sensor is equipped with an ultrasonic range sensor with a given detection range $\mathrm{d} \geq 2\sqrt{2}$, when the sensor to target distance is less than d, the target will be detected and the distance will be estimated. These detecting sensor nodes dynamically form a cluster [44]. A cluster head is selected and all other sensor nodes within the cluster will transmit their distance observations to the cluster head for tracking computation.

A moving target travels at a constant angular velocity of 0.122 rad/s along a trajectory consisting of two adjacent circles of radius 0.35 meters. The sampling interval is $\Delta_k = 0.2$ second (5 Hz). The process noise $\boldsymbol{w}_{k-1}$ corresponds to the variable acceleration of the target and can be approximated by a white Gaussian sequence with zero mean and covariance

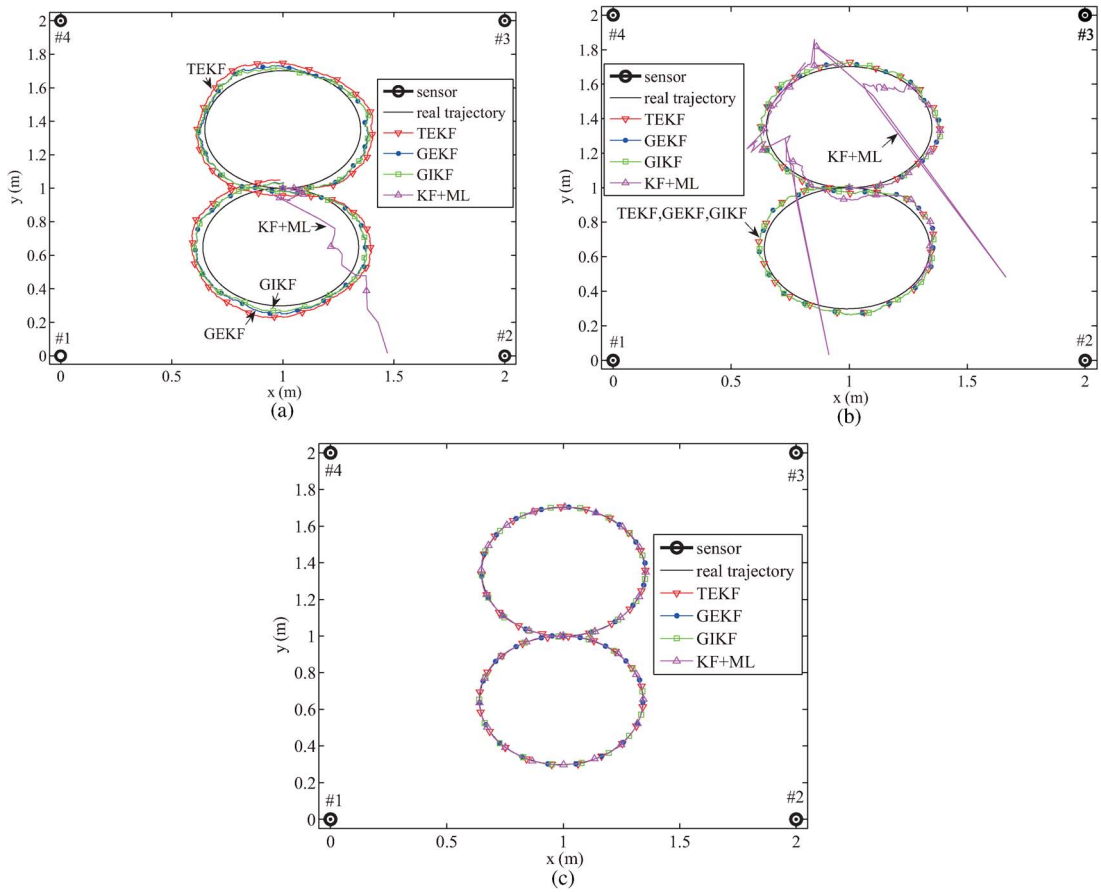


Fig. 1. Tracking trajectories under different noise conditions: The target travels along a trajectory consisting of two adjacent circles, the high noise condition imitates an outdoor environment, the additive noise condition supposes a traditional additive noise assumption while the low noise condition represents an actual indoor environment; (a) High noise; (b) Additive noise; (c) Low noise.

matrix $\boldsymbol{Q}_{k-1} = \text{diag}\{0.0027, 0.0027\}$ [14]. Three different measurement noise conditions are given to test the algorithms: (a) High noise condition with $\mu_u = 0.94$, $\sigma_u^2 = 0.3$, $\mu_v = 0.45$, $\sigma_v^2 = 0.2$, which imitates an outdoor environment, (b) Additive noise condition with $\mu_u = 0$, $\sigma_u^2 = 0$, $\mu_v = 0.3$, $\sigma_v^2 = 0.1$, which supposes a traditional additive noise assumption, and (c) Low noise condition with $\mu_u = 0.0174$, $\sigma_u^2 = 2.916 \times 10^{-4}$, $\mu_v = -0.0386$, $\sigma_v^2 = 7.97 \times 10^{-5}$, which represents an actual indoor environment.

The initial values of target state and error covariance are assumed to be:

$$\hat{\boldsymbol{x}}_{0|0} = [1.0\ m\ 0.0428\ m/s\ 1.0\ m\ 0.0\ m/s]^{\mathrm{T}},$$
$$\boldsymbol{P}_{0|0} = 0.001 \times \text{diag}\{1\ 1\ 1\ 1\}.$$

One hundred Monte-Carlo runs with the above noise conditions are performed. The results are summarized in the following. Fig. 1 shows the average tracking trajectories of 100 realizations for the four filtering algorithms. Fig. 2 gives the population and its mean of quadratic sums of the state estimation errors in each time step. Specially the population for the KF+ML is not depicted here to clearly distinguish from these filters. Fig. 3 describes the corresponding mean value of the population of target estimated location biases $\|(x(k), y(k)) - (\hat{x}^{\mathrm{t}}_{k|k}, \hat{y}^{\mathrm{t}}_{k|k})\|$ for each filter, where $(\hat{x}^{\mathrm{t}}_{k|k}, \hat{y}^{\mathrm{t}}_{k|k})$ comes from the estimated value $\hat{\boldsymbol{x}}^{\mathrm{t}}_{k|k}$ of the Kalman filters, in GIKF $t = 2$, in the other filters $\hat{\boldsymbol{x}}^{\mathrm{t}}_{k|k} = \hat{\boldsymbol{x}}_{k|k}$. In the low noise condition, the impact of the noise has negligible magnitude, then the observations are very accurate, thus all the filtering algorithms attain similar performance. However, under the high noise condition, the measurement errors become bigger, as shown in Fig. 1. The tracking trajectory in the KF+ML greatly diverges from the true trajectory, and the mean squared error(MSE) of the state estimation and the estimation bias are the worst among the four algorithms. From Fig. 2 we can see that, due to the inadequate noise model, the TEKF has bigger MSE and estimation bias than the GEKF and the GIKF, and the tracking MSE and estimation bias in the GEKF are more than those of the GIKF. In particular, the MSE of the GIKF is closest to the CRLB, and the GIKF has the best tracking performance. This result indicates that the GIKF can yield more accurate estimate than the GEKF in the high noise environments.

In contrast, in the additive noise condition, the GEKF simplifies to the traditional EKF, then the GEKF and TEKF have the same tracking accuracy. Table II lists the average root mean square of the target estimated position error (RMSE) $\sqrt{\mathbb{E}\{\|(x(k), y(k)) - (\hat{x}^{\mathrm{t}}_{k|k}, \hat{y}^{\mathrm{t}}_{k|k})\|^2\}}$ in all the time steps under the additive noise condition when the sampling interval is 0.2 second. In this case, the GIKF also simplifies to the traditional

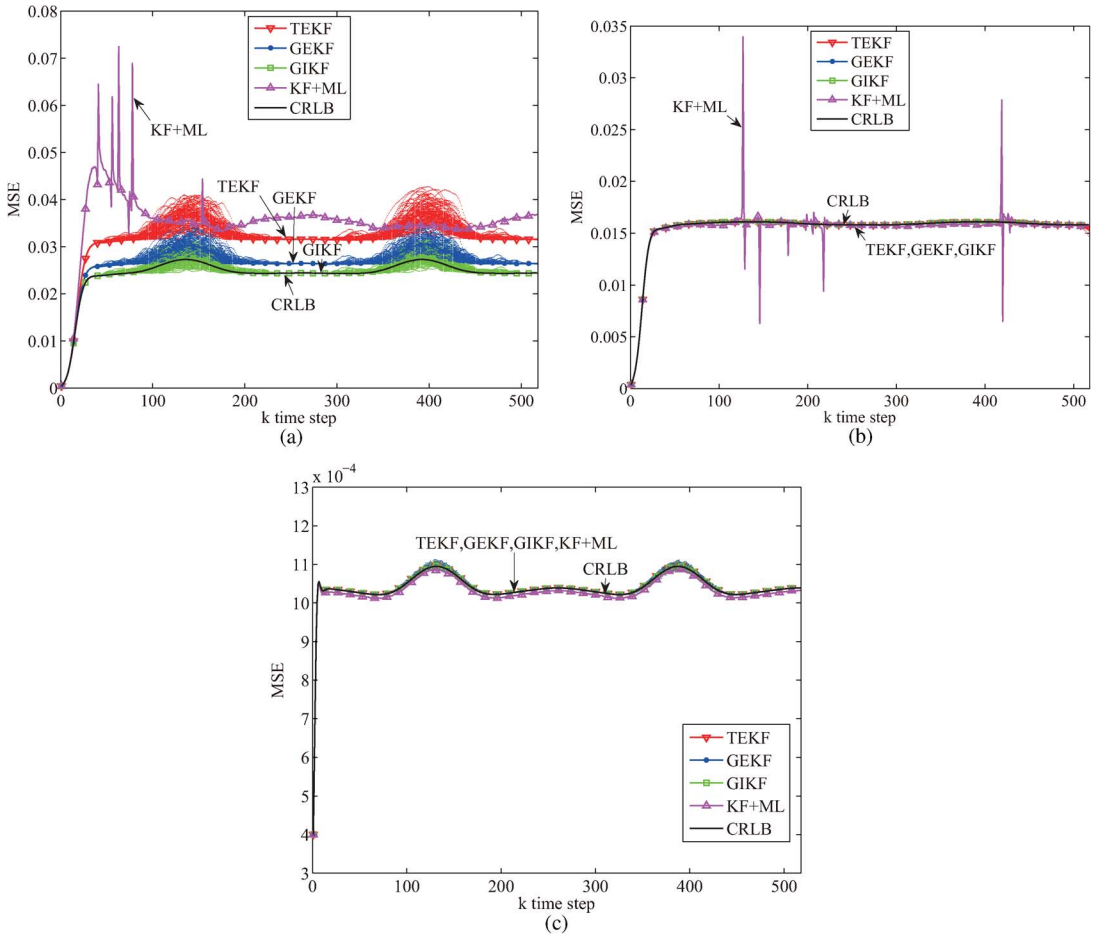


Fig. 2. The mean squared errors of the state estimation under different noise conditions: The sample data points are the MSE population of 100 realizations, the solid line with marker describes the mean value of population, the black solid line represents the real CRLB; (a)High noise; (b) Additive noise; (c) Low noise.

IKF. From Table II, it can be seen that the GIKF with one iteration can obtain the same RMSE with the GEKF and the TEKF.

Table I and III also show the average RMSEs of the TEKF, GEKF, KF+ML and the GIKF with different sampling intervals under the other two noise conditions. In the high noise condition, the RMSE of the GIKF with a single iteration is smaller than that of the GEKF, and the RMSE of the GIKF decreases as the number of iterations increases. In the low noise condition, the RMSEs of the four filtering algorithms are almost the same. Because the state prediction is very close to the true target state in the low noise condition, the GIKF has a higher rate of convergence than that in the other noise conditions. Moreover, the RMSEs of Kalman filters decrease as the sampling interval decreases, this is because the constant-velocity motion model becomes more precise in the shorter sampling interval, which results in better state prediction and convergence. Similarly, Fig. 4 shows the MSEs of the GIKF with different numbers of iterations at each time step under the high noise condition when the sampling interval is 0.2 second. The MSE of the GIKF at each time step decreases as the number of iterations increases, but the longer iterations can not promise much higher accurate state estimate, the MSE of the GIKF with a few iterations is very close to the CRLB. This result indicates that the GIKF with a few iterations can yield higher accurate estimate and nicer convergence.

To investigate the stability of the proposed GIKF, another simulation is conducted that uses the same setting as above except for much worse noise condition. In particular, we set $\mu_u = 0.94, \sigma_u^2 = 1.5, \mu_v = 0.45, \sigma_v^2 = 1.2$ and $\Delta_k = 0.3$. With such unfavorable noise conditions, the KF+ML method failed miserably (as shown in Fig. 5) and hence its results are excluded from reporting. In Fig. 5, the average tracking trajectories of the TEKF, GEKF and GIKF are plotted. Clearly, the GIKF achieves better tracking accuracy, and the GEKF has marginally better performance than the TEKF. Fig. 6 plots the distributions of tracking errors of the position states along $x-$ and $y-$directions for these three methods, together with the corresponding error covariance ellipses of position state $(x(k), y(k))$. The error covariance of the position state is defined as

$$\boldsymbol{P}^{\mathrm{t}}_{k|k,xy} = \begin{bmatrix} \boldsymbol{P}^{\mathrm{t}}_{k|k}[1,1] & \boldsymbol{P}^{\mathrm{t}}_{k|k}[1,3] \\ \boldsymbol{P}^{\mathrm{t}}_{k|k}[3,1] & \boldsymbol{P}^{\mathrm{t}}_{k|k}[3,3] \end{bmatrix},$$

where $\boldsymbol{P}^{\mathrm{t}}_{k|k}[i,i]$ denotes the diagonal element of the error covariance matrix $\boldsymbol{P}^{\mathrm{t}}_{k|k}$. The major and minor axes of each ellipse correspond to the square roots of eigenvalues of the error covariance matrix, and the angles between these axes and x-axis are determined by the orientation of corresponding eigenvectors. It

TABLE I
AVERAGE RMSES OF DIFFERENT SAMPLING INTERVALS UNDER THE HIGH NOISE CONDITION

Sampling Interval (s)	Average RMSE							
	TEKF	GEKF	KF+ML	GIKF				
				t=1	t=2	t=10	t=20	t=50
0.3	0.199631	0.181551	0.213282	0.173419	0.173140	0.173080	0.172940	0.172842
0.2	0.165686	0.150147	0.174012	0.143588	0.143583	0.143507	0.143437	0.143384
0.1	0.118821	0.107551	0.124506	0.103061	0.103057	0.103044	0.103035	0.103026
0.05	0.084889	0.076803	0.089231	0.073661	0.073638	0.073598	0.073568	0.073520

TABLE II
AVERAGE RMSES OF DIFFERENT SAMPLING INTERVALS UNDER THE ADDITIVE NOISE CONDITION

Sampling Interval (s)	Average RMSE							
	TEKF	GEKF	KF+ML	GIKF				
				t=1	t=2	t=10	t=20	t=50
0.3	0.133807	0.133807	0.136701	0.133807	0.133804	0.133799	0.133787	0.133783
0.2	0.111186	0.111186	0.111462	0.111186	0.111181	0.111178	0.111177	0.111176
0.1	0.080112	0.080112	0.080481	0.080112	0.080107	0.080105	0.080105	0.080104
0.05	0.057296	0.057296	0.057861	0.057296	0.057294	0.057294	0.057293	0.057293

TABLE III
AVERAGE RMSES OF DIFFERENT SAMPLING INTERVALS UNDER THE LOW NOISE CONDITION

Sampling Interval (s)	Average RMSE							
	TEKF	GEKF	KF+ML	GIKF				
				t=1	t=2	t=10	t=20	t=50
0.3	0.018279	0.018276	0.018277	0.018274	0.018270	0.018269	0.018268	0.018268
0.2	0.015769	0.015767	0.015768	0.015764	0.015762	0.015762	0.015762	0.015762
0.1	0.011814	0.011813	0.011811	0.011811	0.011810	0.011810	0.011810	0.011810
0.05	0.008612	0.008611	0.008611	0.008610	0.008610	0.008610	0.008610	0.008610

is quite clear that the GIKF achieves smaller error covariance compared to the TEKF or the GEKF, and the tracking errors of the position states in the GIKF for most realizations lie in the ellipse.

The average estimated location bias of the whole tracking trajectory in each realization is shown in Fig. 7. The corresponding average RMSE in the CRLB $\sigma_{xy}(k) = \sqrt{\boldsymbol{\mathcal{J}}_k^{-1}[1,1] + \boldsymbol{\mathcal{J}}_k^{-1}[3,3]}$ is also shown in Fig. 7. As can be seen in Fig. 7 the average estimated location biases in the GIKF for most realizations are below the average RMSE σ_{xy}, while the TEKF and the GEKF are above the average line in many realizations. A comparison of the computation times is also shown in Fig. 8. These simulations are performed on a 2.80-GHz PC equipped with an Intel Xeon processor using MATLAB scripts. As expected, the computation time for the GIKF is similar to those of the TEKF, GEKF or KF+ML when the number of iterations of the GIKF is limited to one. When the number of iterations increases, the computation time for the GIKF increases proportionally.

VI. CONCLUSION

In this paper, we derive a generalized iterated Kalman filter (GIKF) for the nonlinear stochastic discrete-time system with state-dependent multiplicative observation noise. The theoretical relation between the GIKF and the GEKF as well as the TEKF is demonstrated. It is found that the GIKF yields a smaller MSE than the GEKF and the TEKF in the multiplicative observation noise model. Furthermore, the CRLB is introduced as the performance measure of the error behavior and the local convergence of MSE of GIKF is rigorously established. Simulation results are also reported that the GIKF with a few iterations can yield higher accurate estimate and nicer convergence than existing methods. In the future work, we hope to apply the proposed algorithm to develop a target tracking platform in a mobile acoustic array network.

APPENDIX

A. Lemma A1

Lemma A1: For two random Gaussian column vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, if the joint density of $\boldsymbol{x}, \boldsymbol{y}$ has the Gaussian probability densities, it is given as [45]

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \Xi_{\boldsymbol{x}} \\ \Xi_{\boldsymbol{y}} \end{bmatrix}, \begin{bmatrix} \mathrm{Cov}\{\boldsymbol{x},\boldsymbol{x}\} & \mathrm{Cov}\{\boldsymbol{x},\boldsymbol{y}\} \\ \mathrm{Cov}^{\mathrm{T}}\{\boldsymbol{x},\boldsymbol{y}\} & \mathrm{Cov}\{\boldsymbol{y},\boldsymbol{y}\} \end{bmatrix} \right) \tag{43}$$

where $\Xi_{\boldsymbol{x}} = \mathbb{E}\{\boldsymbol{x}\}$, $\Xi_{\boldsymbol{y}} = \mathbb{E}\{\boldsymbol{y}\}$, then the marginal and conditional densities of $\boldsymbol{x}$ and $\boldsymbol{y}$ are given as follows:

$$p(\boldsymbol{x}) = \mathcal{N}\left(\Xi_{\boldsymbol{x}}, \mathrm{Cov}\{\boldsymbol{x},\boldsymbol{x}\}\right), \tag{44}$$

$$p(\boldsymbol{y}) = \mathcal{N}\left(\Xi_{\boldsymbol{y}}, \mathrm{Cov}\{\boldsymbol{y},\boldsymbol{y}\}\right), \tag{45}$$

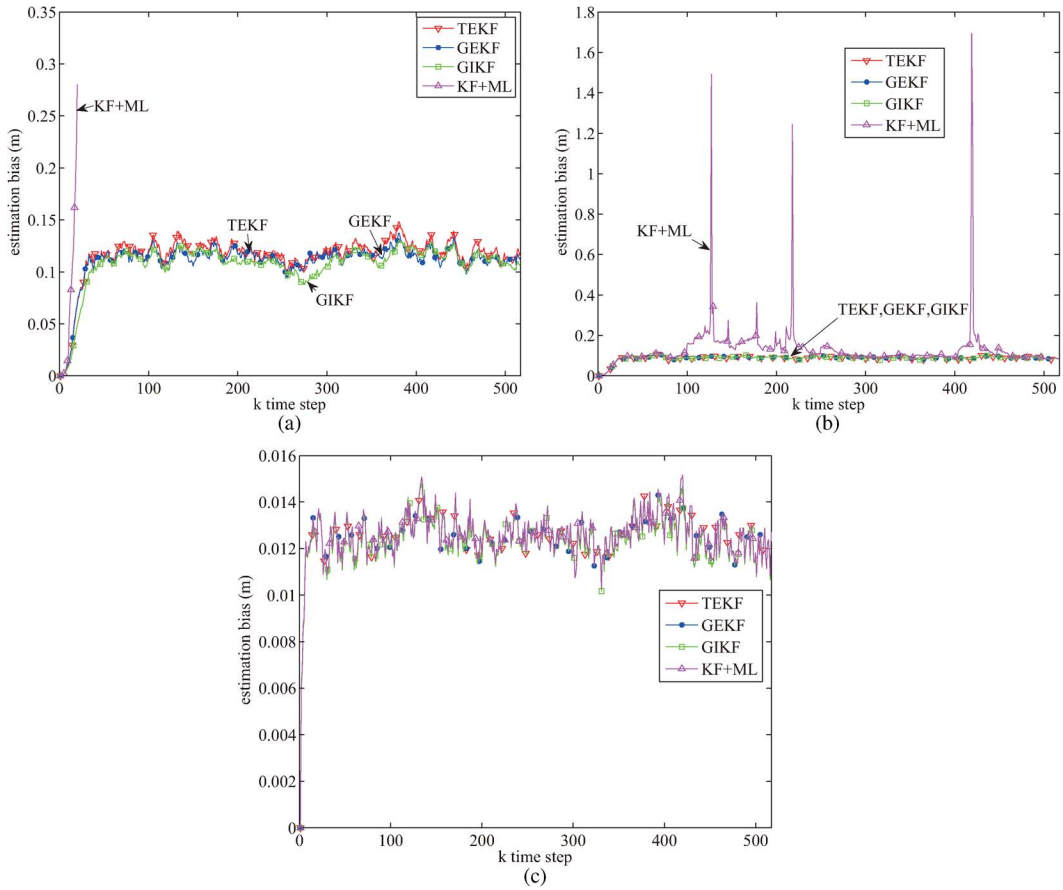


Fig. 3. The average estimated location biases of 100 realizations under different noise conditions: The solid line with marker represents the mean value of population of target estimated location biases for all the time steps, the KF+ML takes biases at the some previous time steps while the others go beyond the frame of axes in the high noise condition; (a) High noise; (b) Additive noise; (c) Low noise.

$$p(\boldsymbol{x}|\boldsymbol{y}) = \mathcal{N}\left(\Xi_{\boldsymbol{x}} + \mathrm{Cov}\{\boldsymbol{x},\boldsymbol{y}\}\mathrm{Cov}^{-1}\{\boldsymbol{y},\boldsymbol{y}\}(\boldsymbol{y}-\Xi_{\boldsymbol{y}}),\ \mathrm{Cov}\{\boldsymbol{x},\boldsymbol{x}\} - \mathrm{Cov}\{\boldsymbol{x},\boldsymbol{y}\}\mathrm{Cov}^{-1}\{\boldsymbol{y},\boldsymbol{y}\}\mathrm{Cov}^{\mathrm{T}}\{\boldsymbol{x},\boldsymbol{y}\}\right) \tag{46}$$

$$p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}\left(\Xi_{\boldsymbol{y}} + \mathrm{Cov}^{\mathrm{T}}\{\boldsymbol{x},\boldsymbol{y}\}\mathrm{Cov}^{-1}\{\boldsymbol{x},\boldsymbol{x}\}(\boldsymbol{x}-\Xi_{\boldsymbol{x}}),\ \mathrm{Cov}\{\boldsymbol{y},\boldsymbol{y}\} - \mathrm{Cov}^{\mathrm{T}}\{\boldsymbol{x},\boldsymbol{y}\}\mathrm{Cov}^{-1}\{\boldsymbol{x},\boldsymbol{x}\}\mathrm{Cov}\{\boldsymbol{x},\boldsymbol{y}\}\right). \tag{47}$$

B. Derivation of (12)

According to (7) and (4)~(6), the covariance of $\boldsymbol{z}_k$ given $\boldsymbol{x}_k$ is shown in the equation at the bottom of the page.

C. Proof of Proposition 2

Proof: Given the discrete system (1) and (7), the posterior information matrices for estimating state obeys the recursion

$$\boldsymbol{\mathcal{J}}_k = \boldsymbol{\theta}_k^{22} - \boldsymbol{\theta}_k^{21}\left(\boldsymbol{\mathcal{J}}_{k-1} + \boldsymbol{\theta}_k^{11}\right)^{-1}\boldsymbol{\theta}_k^{12} \tag{48}$$

$$\begin{aligned}
\boldsymbol{\Sigma}_k =& \mathbb{E}\left\{(\boldsymbol{z}_k - \Xi_k)(\boldsymbol{z}_k - \Xi_k)^{\mathrm{T}}|\boldsymbol{x}_k\right\} \\
=& \mathbb{E}\left\{\begin{bmatrix} h_1(\boldsymbol{x}_k)(u_1(k)-\mu_u) + (v_1(k)-\mu_v) \\ h_2(\boldsymbol{x}_k)(u_2(k)-\mu_u) + (v_2(k)-\mu_v) \\ \vdots \\ h_{\ell_k}(\boldsymbol{x}_k)(u_{\ell_k}(k)-\mu_u) + (v_{\ell_k}(k)-\mu_v) \end{bmatrix} \cdot \begin{bmatrix} h_1(\boldsymbol{x}_k)(u_1(k)-\mu_u) + (v_1(k)-\mu_v) \\ h_2(\boldsymbol{x}_k)(u_2(k)-\mu_u) + (v_2(k)-\mu_v) \\ \vdots \\ h_{\ell_k}(\boldsymbol{x}_k)(u_{\ell_k}(k)-\mu_u) + (v_{\ell_k}(k)-\mu_v) \end{bmatrix}^{\mathrm{T}}\right\} \\
=& \begin{bmatrix} \sigma_u^2 h_1^2(\boldsymbol{x}_k)+\sigma_v^2 & 0 & \dots & 0 \\ 0 & \sigma_u^2 h_2^2(\boldsymbol{x}_k)+\sigma_v^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_u^2 h_{\ell_k}^2(\boldsymbol{x}_k)+\sigma_v^2 \end{bmatrix} \\
=& \sigma_u^2 \mathrm{diag}\{h_n^2(\boldsymbol{x}_k); 1 \le n \le \ell_k\} + \boldsymbol{R}_k.
\end{aligned}$$

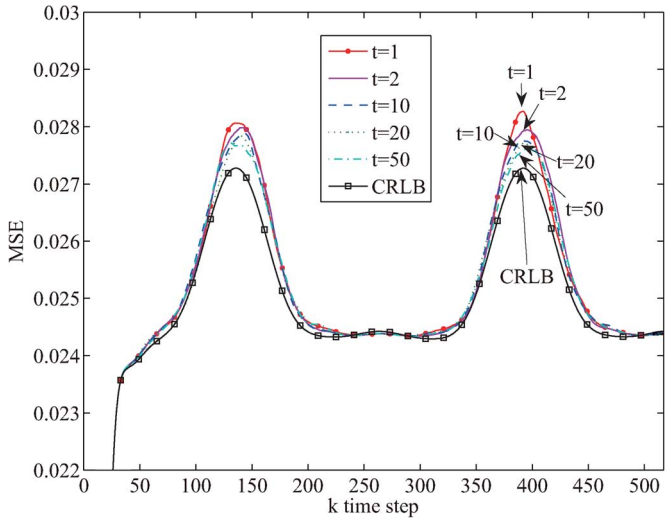


Fig. 4. The MSEs of GIKF with different iteration numbers under the high noise condition: The line represents the mean value of MSE population when the sampling interval is 0.2 second.

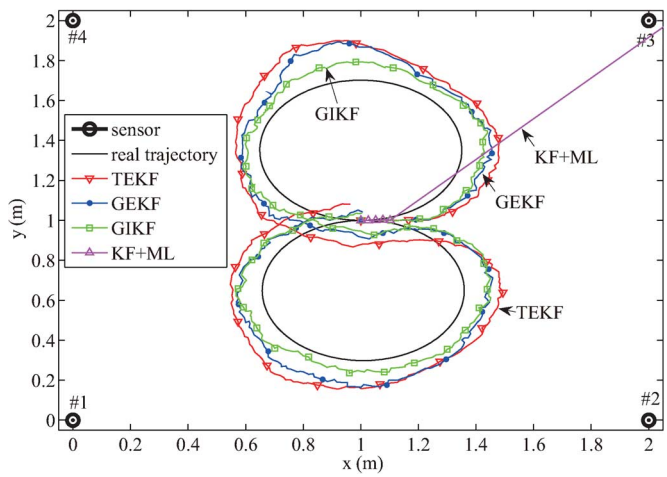


Fig. 5. Tracking trajectories under a heavily noisy condition.

where,

$$\boldsymbol{\theta}_k^{11} = \mathbb{E}\left\{-\frac{\partial^2 \ln p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})}{\partial \boldsymbol{x}_{k-1}^{\mathrm{T}} \partial \boldsymbol{x}_{k-1}}\right\},$$

$$\boldsymbol{\theta}_k^{12} = \mathbb{E}\left\{-\frac{\partial^2 \ln p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})}{\partial \boldsymbol{x}_k^{\mathrm{T}} \partial \boldsymbol{x}_{k-1}}\right\},$$

$$\boldsymbol{\theta}_k^{21} = \mathbb{E}\left\{-\frac{\partial^2 \ln p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})}{\partial \boldsymbol{x}_{k-1}^{\mathrm{T}} \partial \boldsymbol{x}_k}\right\},$$

$$\boldsymbol{\theta}_k^{22} = \mathbb{E}\left\{-\frac{\partial^2 \ln p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})}{\partial \boldsymbol{x}_k^{\mathrm{T}} \partial \boldsymbol{x}_k}\right\} + \mathbb{E}\left\{-\frac{\partial^2 \ln p(\boldsymbol{z}_k|\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k^{\mathrm{T}} \partial \boldsymbol{x}_k}\right\}.$$

According to (13) and $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = \mathcal{N}(\boldsymbol{F}_k\boldsymbol{x}_{k-1}, \boldsymbol{G}_k\boldsymbol{Q}_{k-1}\boldsymbol{G}_k^{\mathrm{T}})$, it gets

$$\boldsymbol{\theta}_k^{11} = \boldsymbol{F}_k^{\mathrm{T}}\left(\boldsymbol{G}_k\boldsymbol{Q}_{k-1}\boldsymbol{G}_k^{\mathrm{T}}\right)^{-1}\boldsymbol{F}_k,$$

$$\boldsymbol{\theta}_k^{12} = -\boldsymbol{F}_k^{\mathrm{T}}\left(\boldsymbol{G}_k\boldsymbol{Q}_{k-1}\boldsymbol{G}_k^{\mathrm{T}}\right)^{-1},$$

$$\boldsymbol{\theta}_k^{21} = -\left(\boldsymbol{G}_k\boldsymbol{Q}_{k-1}\boldsymbol{G}_k^{\mathrm{T}}\right)^{-1}\boldsymbol{F}_k,$$

$$\boldsymbol{\theta}_k^{22} = \left(\boldsymbol{G}_k\boldsymbol{Q}_{k-1}\boldsymbol{G}_k^{\mathrm{T}}\right)^{-1} + \boldsymbol{M}_k.$$

■

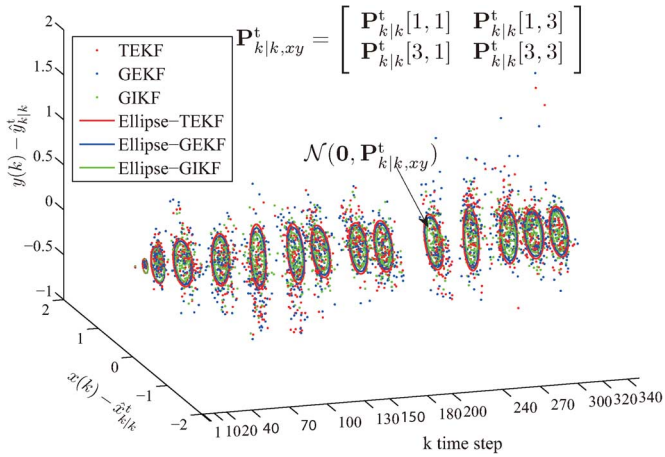


Fig. 6. The tracking errors along $x-$ and $y-$directions and the error covariance ellipses of the position state: The sample data points are the population of 100 realizations of $x(k) - \hat{x}^{\mathrm{t}}_{k|k}$ and $y(k) - \hat{y}^{\mathrm{t}}_{k|k}$ at some random time steps.

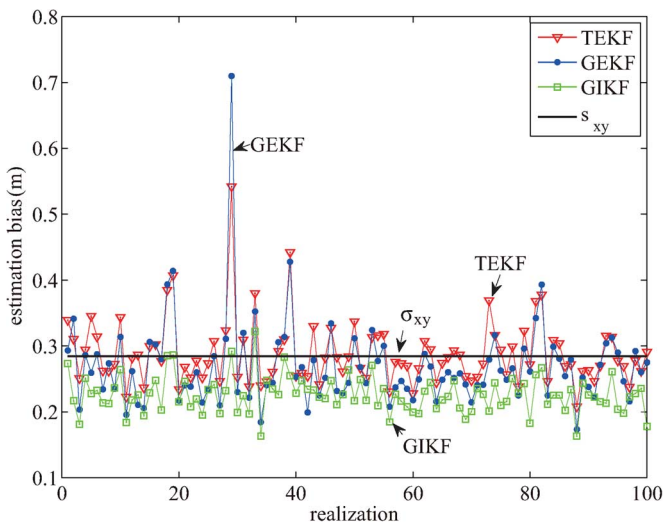


Fig. 7. The average estimated location bias of the whole tracking trajectory in each realization: The solid line with marker represents the average estimated location bias of all the time steps. The estimation biases in the KF+ML entirely go beyond the frame of axes, it is not shown in the figure.

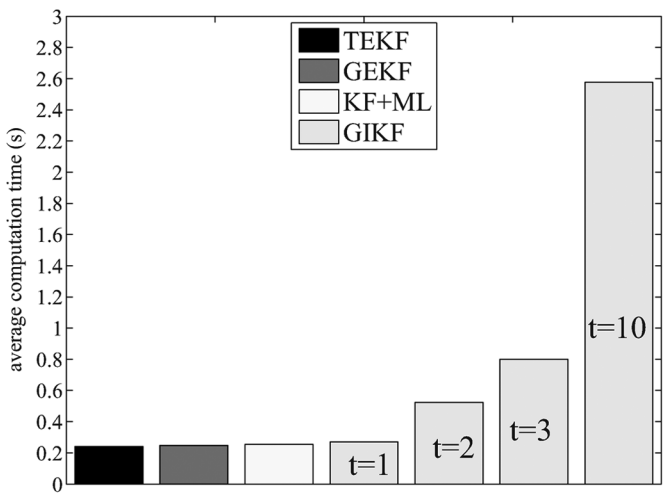


Fig. 8. The average computation time of the TEKF, GEKF, GIKF and KF+ML: The computation time is an average elapsed time for one realization.

D. Proof of Theorem 1

Proof: For normal probability density functions, the first moment of the score $\mathbb{E}\{\boldsymbol{\Gamma}_k(\boldsymbol{x}_k)\} = \mathbf{0}$ [46], then

$$\begin{aligned}
&\mathbb{E}\left\{\left\|\hat{\boldsymbol{x}}_{k|k}^{t+1}-\boldsymbol{x}_k\right\|^2\right\} = \mathbb{E}\left\{(\hat{\boldsymbol{x}}_{k|k}^{t+1}-\boldsymbol{x}_k)^{\mathrm{T}}\left(\hat{\boldsymbol{x}}_{k|k}^{t+1}-\boldsymbol{x}_k\right)\right\}\\
&=\mathbb{E}\left\{\left(\hat{\boldsymbol{x}}_{k|k}^{t}-\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)^{-1}\boldsymbol{\Gamma}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)-\boldsymbol{x}_k\right)^{\mathrm{T}}\right.\\
&\left.\cdot\left(\hat{\boldsymbol{x}}_{k|k}^{t}-\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)^{-1}\boldsymbol{\Gamma}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)-\boldsymbol{x}_k\right)\right\}\\
&=\mathbb{E}\left\{\left(\hat{\boldsymbol{x}}_{k|k}^{t}+\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)^{-1}\left(\boldsymbol{\Gamma}_k(\boldsymbol{x}_k)-\boldsymbol{\Gamma}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)\right)-\boldsymbol{x}_k\right)^{\mathrm{T}}\right.\\
&\cdot\left(\hat{\boldsymbol{x}}_{k|k}^{t}+\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)^{-1}\left(\boldsymbol{\Gamma}_k(\boldsymbol{x}_k)-\boldsymbol{\Gamma}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)\right)-\boldsymbol{x}_k\right)\\
&\left.-\boldsymbol{\Gamma}_k^{\mathrm{T}}(\boldsymbol{x}_k)\boldsymbol{K}_k^{\mathrm{T}}\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)^{-1}\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)^{-1}\boldsymbol{\Gamma}_k(\boldsymbol{x}_k)\right\}.
\end{aligned}$$

Due to $\boldsymbol{K}_k(\hat{\boldsymbol{x}}_{k|k}^t) > 0$, then $\boldsymbol{\Gamma}_k^{\mathrm{T}}(\boldsymbol{x}_k)\boldsymbol{K}_k^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k|k}^t)^{-1}\cdot\boldsymbol{K}_k(\hat{\boldsymbol{x}}_{k|k}^t)^{-1}\boldsymbol{\Gamma}_k(\boldsymbol{x}_k) \geq 0$. Using Lemma 1, it has

$$\begin{aligned}
&\mathbb{E}\left\{\left\|\hat{\boldsymbol{x}}_{k|k}^{t+1}-\boldsymbol{x}_k\right\|^2\right\}\\
&\leq\mathbb{E}\left\{\left\|\hat{\boldsymbol{x}}_{k|k}^{t}-\boldsymbol{x}_k+\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)^{-1}\left(\boldsymbol{\Gamma}_k(\boldsymbol{x}_k)-\boldsymbol{\Gamma}_k(\hat{\boldsymbol{x}}_{k|k}^{t})\right)\right\|^2\right\}\\
&=\mathbb{E}\left\{\left\|\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)^{-1}\int_0^1\left[\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}+\tau\left(\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^{t}\right)\right)\right.\right.\\
&\qquad\left.\left.-\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)\right]\left(\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^{t}\right)d\tau\right\|^2\right\}\\
&\leq\mathbb{E}\left\{\left[\left\|\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)^{-1}\right\|\right.\right.\\
&\qquad\cdot\int_0^1\left\|\left[\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}+\tau\left(\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^{t}\right)\right)\right.\right.\\
&\qquad\left.\left.\left.-\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)\right]\right\|\left\|\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^{t}\right\|d\tau\right]^2\right\}\\
&\leq\mathbb{E}\left\{\left[\left\|\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^{t}\right\|\left\|\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)^{-1}\right\|\int_0^1\gamma\tau\left\|\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^{t}\right\|d\tau\right]^2\right\}\\
&=\mathbb{E}\left\{\left[\left\|\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^{t}\right\|^2\left\|\boldsymbol{K}_k\left(\hat{\boldsymbol{x}}_{k|k}^{t}\right)^{-1}\right\|\frac{\gamma}{2}\right]^2\right\}
\end{aligned}$$

Because $\gamma\|\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^t\| \leq \frac{2\alpha\eta}{3} < \frac{2\eta}{3}$, and combining (34) with the above yields

$$\begin{aligned}
&\mathbb{E}\left\{\left\|\hat{\boldsymbol{x}}_{k|k}^{t+1}-\boldsymbol{x}_k\right\|^2\right\}\\
&\leq\mathbb{E}\left\{\left[\left\|\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^{t}\right\|^2\frac{\gamma}{2\left(\eta-\gamma\left\|\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^{t}\right\|\right)}\right]^2\right\}\\
&=\mathbb{E}\left\{\left[\left\|\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^{t}\right\|\frac{\gamma\left\|\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^{t}\right\|}{2\left(\eta-\gamma\left\|\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^{t}\right\|\right)}\right]^2\right\}\\
&<\mathbb{E}\left\{\left\|\boldsymbol{x}_k-\hat{\boldsymbol{x}}_{k|k}^{t}\right\|^2\right\}. \qquad (49)
\end{aligned}$$

■

E. Proof of Proposition 3

Proof: From (37), using the matrix inversion lemma, the covariance of GEKF can be rewritten as

$$\begin{aligned}
\boldsymbol{P}_{k|k} = \Big[&\boldsymbol{P}_{k|k-1}^{-1}+(1+\mu_u)^2\boldsymbol{H}_k^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k|k-1})\\
&\left(\sigma_u^2\boldsymbol{\Theta}_k(\hat{\boldsymbol{x}}_{k|k-1})+\sigma_v^2\mathbf{I}\right)^{-1}\boldsymbol{H}_k(\hat{\boldsymbol{x}}_{k|k-1})\Big]^{-1}. \qquad (50)
\end{aligned}$$

The ij entry of $\boldsymbol{\Omega}_k = (1+\mu_u)^2\boldsymbol{H}_k^{\mathrm{T}}(\boldsymbol{x}_k)(\sigma_u^2\boldsymbol{\Theta}_k(\boldsymbol{x}_k)+\sigma_v^2\mathbf{I})^{-1}\boldsymbol{H}_k(\boldsymbol{x}_k)$ is given by

$$\omega_{ij} = \sum_{n=1}^{n=\ell_k} a_n(k)\frac{\partial h_n(\boldsymbol{x}_k)}{\partial x_i}\frac{\partial h_n(\boldsymbol{x}_k)}{\partial x_j}$$

in the above, $a_n(k) = \frac{(1+\mu_u)^2}{\sigma_u^2\dot{h}_n(k)\boldsymbol{P}_{k|k-1}\dot{h}_n^{\mathrm{T}}(k)+\sigma_u^2h_n^2(\boldsymbol{x}_k)+\sigma_v^2}$ where $\dot{h}_n(k) = \frac{\partial h_n(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k^{\mathrm{T}}}$.

According to the definition of positive-definite matrix, due to $\boldsymbol{P}_{k|k-1} > 0$, then $\dot{h}_n(k)\boldsymbol{P}_{k|k-1}\dot{h}_n^{\mathrm{T}}(k) \geq 0$, then $b_n(k) > a_n(k)$. let $b_n(k) - a_n(k) = c_n(k) > 0$, it follows that from (24) the entry of $\boldsymbol{M}_k - \boldsymbol{\Omega}_k$ is

$$m_{ij}-\omega_{ij} = \sum_{n=1}^{n=\ell_k} c_n(k)\frac{\partial h_n(\boldsymbol{x}_k)}{\partial x_i}\frac{\partial h_n(\boldsymbol{x}_k)}{\partial x_j}.$$

Similarly with Proposition 1, we can rewrite $\boldsymbol{M}_k - \boldsymbol{\Omega}_k = \boldsymbol{H}_k^{\mathrm{T}}(\boldsymbol{x}_k)\boldsymbol{C}_k(\boldsymbol{x}_k)\boldsymbol{H}_k(\boldsymbol{x}_k)$ where $\boldsymbol{C}_k(\boldsymbol{x}_k) = \mathrm{diag}\{c_n(k); 1 \leq n \leq \ell_k\}$. Since $\boldsymbol{C}_k(\boldsymbol{x}_k)$ is a symmetric positive definite diagonal matrix, $\boldsymbol{M}_k - \boldsymbol{\Omega}_k$ is positive semi-definite. Thus in the initial iteration $t = 0$, $\hat{\boldsymbol{x}}_{k|k}^0 = \hat{\boldsymbol{x}}_{k|k-1}$, it has

$$\boldsymbol{P}_{k|k-1}^{-1}+\boldsymbol{\Omega}_k|_{\boldsymbol{x}_k} = \hat{\boldsymbol{x}}_{k|k-1} < \boldsymbol{P}_{k|k-1}^{-1}+\boldsymbol{M}_k|_{\boldsymbol{x}_k=\hat{\boldsymbol{x}}_{k|k}^0}. \qquad (51)$$

Then, $[\boldsymbol{P}_{k|k-1}^{-1}+\boldsymbol{M}_k|_{\boldsymbol{x}_k=\hat{\boldsymbol{x}}_{k|k}^0}]^{-1} < [\boldsymbol{P}_{k|k-1}^{-1}+\boldsymbol{\Omega}_k|_{\boldsymbol{x}_k=\hat{\boldsymbol{x}}_{k|k-1}}]^{-1}$, or equivalently $\mathrm{tr}\{\boldsymbol{P}_{k|k}^0\} < \mathrm{tr}\{\boldsymbol{P}_{k|k}\}$. By Theorem 1, for $t \geq 1$, $\mathrm{tr}\{\boldsymbol{P}_{k|k}^t\} < \mathrm{tr}\{\boldsymbol{P}_{k|k}^0\} < \mathrm{tr}\{\boldsymbol{P}_{k|k}\}$.

REFERENCES

[1] A. Farina, B. Ristic, and D. Benvenuti, "Tracking a ballistic target: Comparison of several nonlinear filters," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 854–867, 2002.

[2] D. Naso, A. Scalera, G. Aurisicchio, and B. Turchiano, "Removing spike noise from railway geometry measures with a fuzzy filter," *IEEE Trans. Syst., Man, Cybern. C*, vol. 36, no. 4, pp. 485–494, 2006.

[3] T. He, S. Krishnamurthy, L. Luo, T. Yan, L. Gu, R. Stoleru, G. Zhou, C. Q. , P. Vicaire, J. A. Stankovic, A. T. F. , J. Hui, and B. Krogh, "Vigilnet: An integrated sensor network system for energy-efficient surveillance," *ACM Trans. Sens. Netw.*, vol. 2, no. 1, pp. 1–38, 2006.

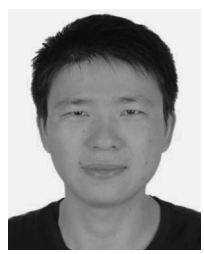
[4] K. Ho, "Bias reduction for an explicit solution of source localization using tdoa," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2101–2114, 2012.

[5] Y. H. Hu and X. Sheng, "Dynamic sensor self-organization for distributed moving target tracking," *J. Signal Process. Sys.*, vol. 51, no. 2, pp. 161–172, 2008.

[6] I. Osborne, R. W. , Y. Bar-Shalom, J. George, and L. Kaplan, "Statistical efficiency of simultaneous target and sensors localization with position dependent noise," in *Proc. SPIE Int. Soc. Opt. Eng.*, Baltimore, MD, USA, 2012, vol. 8392, pp. 03–14.

[7] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, pp. 1–45, 2006.

[8] A. Logothetis, A. Isaksson, and R. Evans, "An information theoretic approach to observer path design for bearings-only tracking," in *Proc. 36th Conf. Decision Control*, San Diego, CA, 1997, vol. 4, pp. 3132–3137.
[9] G. Benet, F. Blanes, J. Sim, and P. Prez, "Using infrared sensors for distance measurement in mobile robots," *Robot. Auton. Syst.*, vol. 40, no. 4, pp. 255–266, 2002.
[10] G. Agamennoni and E. M. Nebot, "Robust estimation in non-linear state-space models with state-dependent noise," Australian Centre for Field Robotics, Sydney, Australia, 2013.
[11] S. Aja-Fernández, G. Vegas-Sánchez-Ferrero, M. Martin-Fernández, and C. Alberola-López, "Automatic noise estimation in images using local statistics. additive and multiplicative cases," *Image Vision Comput.*, vol. 27, no. 6, pp. 756–770, 2009.
[12] E. Gershon and U. Shaked, "h_∞ output-feedback control of discrete-time systems with state-multiplicative noise," *Automatica*, vol. 44, no. 2, pp. 574–579, 2008.
[13] H. Dong, Z. Wang, and H. Gao, "Robust filtering for a class of nonlinear networked systems with multiple stochastic communication delays and packet dropouts," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 1957–1966, April 2010.
[14] X. Wang, M. Fu, and H. Zhang, "Target tracking in wireless sensor networks based on the combination of KF and MLE using distance measurements," *IEEE Trans. Mobile Comput.*, vol. 11, no. 4, pp. 567–576, 2012.
[15] R. E. Kalman, "A new approach to linear filtering and prediction problems," *ASME J. Basic Eng.*, vol. 82, no. Series D, pp. 35–45, 1960.
[16] S. J. Julier and J. K. Uhlmann, "New extension of the Kalman filter to nonlinear systems," in *Proc. SPIE Int. Symp. Aerosp./Defense Sens.*, 1997, vol. 3068, pp. 182–193.
[17] L. Li, H. Ji, and J. Luo, "The iterated extended kalman particle filter," in *Proc. Int. Symp. Commun. Inf. Technol.*, Beijing, China, 2005, vol. 2, pp. 1213–1216.
[18] T. Kirubarajan and Y. Bar-Shalom, "Kalman filter versus IMM estimator: When do we need the latter?," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1452–1457, 2003.
[19] S. Julier, J. Uhlmann, and H. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *Proc. Amer. Control Conf.*, 1995, vol. 3, pp. 1628–1632.
[20] R. Zhan and J. Wan, "Iterated unscented Kalman filter for passive target tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, pp. 1155–1163, 2007.
[21] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. Autom. Control*, vol. 45, no. 5, pp. 910–927, 2000.
[22] D. Spinello and D. Stilwell, "Nonlinear estimation with state-dependent Gaussian observation noise," *IEEE Trans. Autom. Control*, vol. 55, no. 6, pp. 1358–1366, 2010.
[23] F. Yang, Z. Wang, and Y. Hung, "Robust Kalman filtering for discrete time-varying uncertain systems with multiplicative noises," *IEEE Trans. Autom. Control*, vol. 47, no. 7, pp. 1179–1183, 2002.
[24] J. Jimenez and T. Ozaki, "Linear estimation of continuous-discrete linear state space models with multiplicative noise," *Syst. Control Lett.*, vol. 47, no. 2, pp. 91–101, 2002.
[25] F. Carravetta, A. Germani, and M. Raimondi, "Polynomial filtering of discrete-time stochastic linear systems with multiplicative state noise," *IEEE Trans. Autom. Control*, vol. 42, no. 8, pp. 1106–1126, 1997.
[26] R. Caballero-Aguila, A. Hermoso-Carazo, and J. Linares-Perez, *Unscented Filtering Algorithm for Discrete-Time Systems with Uncertain Observations and State-Dependent Noise*. New York, NY, USA: Academic, 2011, ch. 7.
[27] J. Hu, Z. Wang, H. Gao, and L. K. Stergioulas, "Extended Kalman filtering with stochastic nonlinearities and multiple missing measurements," *Automatica*, vol. 48, no. 9, pp. 2007–2015, 2012.
[28] X. Hu, Y. Hu, and B. Xu, "Generalized Kalman filter tracking with multiplicative measurement noise in a wireless sensor network," *IET Signal Process.*, vol. 8, no. 5, pp. 467–474, 2014.
[29] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. New York, NY, USA: Wiley, 2004.
[30] T. L. Song and J. Speyer, "A stochastic analysis of a modified gain extended Kalman filter with applications to estimation with bearings only measurements," *IEEE Trans. Autom. Control*, vol. 30, no. 10, pp. 940–949, 1985.
[31] X. Hu, B. Xu, and Y. H. Hu, "Target tracking with distance-dependent measurement noise in wireless sensor networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Vancouver, BC, Canada, 2013, pp. 5200–5203.
[32] X. Hu, M. Bao, Y. Hu, and B. Xu, "Energy balanced scheduling for target tracking with distance-dependent measurement noise in a wsn," *Int. J. Distrib. Sens. Netw.*, vol. 2013, 2013, Article ID 179623.
[33] *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, J. E. Dennis and R. B. Schnabel, Eds. Philadelphia, PA, USA: SIAM, 1987.
[34] D. S. Bernstein, *Matrix Mathematics*. Princeton, NJ, USA: Princeton Univ. Press, 2005.
[35] P. Tichavsky, C. Muravchik, and A. Nehorai, "Posterior cramer-rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1386–1396, 1998.
[36] J. H. Kotecha and P. Djuric, "Gaussian particle filtering," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2592–2601, Oct. 2003.
[37] A. G. O. Mutambara, *Decentralized Estimation and Control for Multisensor Systems*. Boca Raton, FL, USA: DCRC Press LLC, 1998.
[38] Y. Bar-Shalom and T. E. Fortman, *Tracking and Data Association*. New York, NY, USA: Academic, 1988.
[39] X. Hu, B. Xu, S. Wen, and Y. Liu, "An energy balanced optimal distributed clustering scheme," *J. South China Univ. Tech. (Natur. Sci. Ed.)*, vol. 40, no. 8, pp. 27–34, 2012.
[40] O. Lakkis, *Lipschitzcondition and Lipschitz Continuity*. Brighton, UK: Univ. of Sussex, 2006 [Online]. Available: http://www.maths.sussex.ac.uk/Staff/OL/Teaching/ADE-2006/Docs/HO-03.pdf
[41] M. A. Epelman, *Continuous Optimization Methods*. Ann Arbor, MI, USA: Univ. of Michigan, 2007 [Online]. Available: http://www-personal.umich.edu/ mepelman/teaching/IOE511/511notes.pdf
[42] B. Bell and F. Cathey, "The iterated Kalman filter update as a Gauss-Newton method," *IEEE Trans. Autom. Control*, vol. 38, no. 2, pp. 294–297, 1993.
[43] R. Van der Merwe and E. Wan, "The square-root unscented Kalman filter for state and parameter-estimation," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Salt Lake City, UT, USA, 2001, vol. 6, pp. 3461–3464.
[44] X. Hu, Y. Hu, and B. Xu, "Energy balanced scheduling for target tracking in wireless sensor networks," *ACM Trans. Sens. Netw.*, vol. 11, no. 1, pp. 21:1–21:29, 2014.
[45] Y. Sarkka, *Bayesian Estimation of Time-Varying Processes: Discrete-time Systems*. Aalto, Finland: Aalto Univ., 2011.
[46] T. Amemiya, *Advanced Econometrics*. Cambridge, U.K.: Harvard Univ. Press, 1985.



Xiaoqing Hu received the B.S. and M.S. degrees from Guangdong University of Technology, China, in 2007 and 2010, respectively, and the Ph.D. degree from South China University of Technology, China, in 2013. From 2011 to 2013, he was a visiting Ph.D. student at the department of Electrical and Computer Engineering Department, University of Wisconsin-Madison in USA. Since 2013, he has been with the Institute of Acoustics, Chinese Academy of Sciences, as a post-doctoral researcher. His current research interests include wireless sensor networks, digital signal processing and intelligent control.

Ming Bao received the B.E. degree from Wuhan University of Transportation Technology, China, in 1996, and his Ph.D. degree in signal and information processing from the Institute of Acoustics, Chinese Academy of Sciences, Beijing, in 2008. Since 1999 he has been with the Institute of Acoustics, Chinese Academy of Sciences, he is currently a Professor. His main research areas are His main research areas lie in pattern recognition, machine learning intelligence acoustic signal processing, binaural interaction modeling, and function sensing.

Xiao-Ping Zhang (M'97–SM'02) received B.S. and Ph.D. degrees from Tsinghua University, in 1992 and 1996, respectively, both in Electronic Engineering. He holds an MBA in Finance, Economics and Entrepreneurship with Honors from the University of Chicago Booth School of Business, Chicago, IL.

Since Fall 2000, he has been with the Department of Electrical and Computer Engineering, Ryerson University, where he is now Professor, Director of Communication and Signal Processing Applications Laboratory (CASPAL). He has served as Program Director of Graduate Studies. He is cross appointed to the Finance Department at the Ted Rogers School of Management at Ryerson University. His research interests include statistical signal processing, multimedia content analysis, sensor networks and electronic systems, computational intelligence, and applications in bioinformatics, finance, and marketing.

Dr. Zhang is a registered Professional Engineer in Ontario, Canada, and a member of Beta Gamma Sigma Honor Society. He is the general chair for MMSP'15. He is the publicity chair for ICME'06 and program chair for ICIC'05 and ICIC'10. He served as guest editor for Multimedia Tools and Applications, and the International Journal of Semantic Computing. He is a tutorial speaker in ACMMM2011, ISCAS2013, ICIP2013 and ICASSP2014. He is currently an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE SIGNAL PROCESSING LETTERS AND FOR JOURNAL OF MULTIMEDIA.

Luyang Guan received B.S. degree from Nankai University, China, in 2002, and Ph.D. degree from the Institute of Acoustics, Chinese Academy of Sciences in 2008. From 2008 to 2009, he was a senior research fellow at University of Huddersfield. Since 2011, he have been an associate professor in Institute of Acoustic, Chinese Academy of Sciences. His current research interests include acoustic and vibration signal processing, target detection and recognition.

Yu-Hen Hu (M'83–SM'88–F'99) received the B.S.E.E. degree from National Taiwan University, Taipei, Taiwan, R.O.C., in 1976 and the M.S. and Ph.D. degrees, both in electrical engineering, from University of Southern California, Los Angeles, in 1980 and 1982, respectively. Currently, he is a Professor with the Electrical and Computer Engineering Department, University of Wisconsin-Madison. Previously, he was with the Electrical Engineering Department, Southern Methodist University, Dallas, TX. His research interests include multimedia signal processing, design methodology and implementation of signal processing algorithms and systems, and neural network signal processing.

Dr. Hu served as an associate editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, the Journal of VLSI Signal Processing, and the European Journal of Applied Signal Processing. He served as Secretary of the IEEE Signal Processing Society, the Board of Governors of the IEEE Neural Networks Council, chair of IEEE Signal Processing Society, and the IEEE Neural Network Signal Processing Technical Committee.