International Journal of Control, Automation, and Systems (2015) 13(3):652-661 ISSN:1598-6446 eISSN:2005-4092
DOI 10.1007/s12555-014-0114-4 http://www.springer.com/12555

Algorithm of Gaussian Sum Filter based on High-order UKF for Dynamic State Estimation

Lei Wang and Xianghong Cheng*

Abstract: In this work we consider the state estimation problem in nonlinear/non-Gaussian systems. A new version of Gaussian sum estimation algorithm is developed here based on high-order unscented Kalman filter (HUKF). A sigma point selection method, high-order unscented transformation (HUT) technique is proposed for the HUKF, which can approximate the Gaussian distributions more accurately. We present the systematic formulation of Gaussian filters and develop efficient and accurate numerical integration of the optimal filter. We then go on to extend the use of the HUKF to discrete-time, nonlinear systems with additive, possibly non-Gaussian noise. The resulting filtering algorithm, called the Gaussian sum high-order unscented Kalman filter (GS-HUKF) approximates the predicted and posterior densities as a finite number of weighted sums of Gaussian densities. It is corroborated in the theoretical analysis and the simulation that the proposed Gaussian sum HUKF has integrated advantages with respect to computational accuracy and time complexity for nonlinear non-Gaussian filtering problems.

Keywords: Gaussian Sum, high-order unscented transformation, nonlinear/non-Gaussian, probability density function, UKF.

1. INTRODUCTION

In recent years, the analysis and making of inferences about a dynamic system arise in a wide variety of applications in many disciplines. Examples include tracking the channel state information of a rapidly changing wireless channel, radar-based tracking of ships and aircraft, estimating the volatility of financial instruments using stock market data, and many others [1-4]. Under the assumption that the underlying dynamic state space model (DSSM) is linear and all of the probability densities are Gaussian and known, the Kalman filter is the optimal solution to the recursive Bayesian estimation problem. However, when we are confronted with a nonlinear filtering problem, we have to abandon the idea of seeking an optimal or analytical solution and be content with a suboptimal solution to the Bayesian filter [5]. In computational terms, suboptimal solutions to the posterior density can be obtained using one of two approximate approaches. One is the sigma-point Kalman filters, which is based on deterministic sampling methods for the propagation of Gaussian random variables through nonlinear systems. The extended Kalman filter (EKF) [6], the central difference Kalman filter (CDKF) [7], the unscented Kalman filter (UKF) [8], the quadrature Kalman filter (QKF) [9,10] and the cubature Kalman filter (CKF) [11] fall under this category. Another group of approximate solution is known as Sequential Monte Carlo methods (SMC) or particle filters [12,13]. The SMC methods, although are the least restrictive with respect to the assumptions, are more computationally expensive than Gaussian approximations.

In the linear/nonlinear systems, all the Gaussian filtering methods have a common assumption that the predicted state must satisfy Gaussian distribution. However, in the actual systems, the statistical characteristics often are not fully known [14]. Gaussian sum filters (GSF) enable a more accurate representation of the nonlinearities and non-Gaussianity in the dynamics and measurement models. The main operation of the Gaussian sum filter is the application of Kalman-filter based prediction and correction steps. In reference to Bayesian inference, the nonlinear non-Gaussian DSSM can be modeled as a bank of parallel nonlinear Gaussian noise dynamic state space models.

The original GSF is constituted by a bank of parallel EKF [15,16], which can approximate non-Gaussian and nonlinear DSSM. However, practices show that EKF is difficult to implement and tune, only reliable for systems that are almost linear on the time scale of the updates and even prone to divergence. The Gaussian sum particle filter (GS-PF) was suggested making use of the idea of the Bayesian sampling [17]. The GS-PF is asymptotically optimal in the number of random samples, which means that equipped with the computational

Manuscript received March 8, 2014; revised July 9, 2014; accepted August 23, 2014. Recommended by Associate Editor Hamid Reza Karimi under the direction of Editor Duk-Sun Shim.

This journal was supported by the National Natural Science Foundation of PR China (approval No. 61374215).

Lei Wang is with the School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China and Anhui Science and Technology University, Bengbu 233100, China (e-mail: frank_408@163.com).

Xianghong Cheng is with the School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: xhcheng@seu.edu.cn).

* Corresponding author.

ability to handle a large number of samples, the GS-PF is supposed to outperform analytical Gaussian sum filters. However, the disadvantage of random sampling-based filter, such as the GS-PF, is that we have to assume great computational burden, usually hundreds of times larger than analytical filters. Wu *et al*. [18] pointed out that we would better remember a practical rule: only when analytical tools do not suffice the requirements is the simulation-based method necessary. Arasaratnam [10] presented the GS-QKF to deal with nonlinear non-Gaussian problems which had a better performance than the GS-EKF and SMC methods. The limitation with this approach is that QKF suffers from the curse of dimensionality. The effect of curse of dimensionality may often become detrimental in high-dimensional state-space models with state-vectors of size 20 or more.

The motivation for this paper is to derive a more accurate filter that could be applied to solve nonlinear non-Gaussian filtering problems. We first develop the high-order unscented Kalman filter (HUKF). A mechanism of high-order unscented transformation (HUT) is described in which the selection of sigma point is according to property density function of variables. Secondly, the application of the HUKF has been extended by incorporating the conventional Gaussian sum filters. The resulting Gaussian sum high-order unscented Kalman filter (GS-HUKF) approximates the predicted and posterior densities as a finite number of weighted sums of Gaussian densities. Through our experiment study we found that the filters developed in the paper are more flexible and have integrated advantages with respect to time complexity and computational accuracy.

The rest of the paper is organized as follows: Section 2 derives the dynamic state space model and Bayesian filter theory in the Gaussian domain. Section 3 describes the high-order unscented transformation (HUT) method and its application to the unscented Kalman filter. The GS-HUKF filter, using the HUKF for the subfilter, is derived in Section 4. Section 5 validates the theoretical analyses and illustrates the proposed algorithm using two numerical examples and finally, Section 6 gives some concluding remarks.

2. PROBLEM DESCRIPTION

We first consider the discrete-time dynamic system with uncertain initial conditions and discrete-time measurement model. The n-dimensional system state x_k, with initial probability density $p(x_0)$, evolves over time as an indirect or partially observed first order Markov process according to the conditional probability density $p(x_k|x_{k-1})$. The observations z_k are conditionally independent given the state and are generated according to the conditional probability density $p(z_k|x_k)$. The dynamic state-space model can be written as a set of nonlinear system equations:

$$\begin{cases} x_k = f(x_{k-1}) + w_{k-1}, & k = 1,\cdots,N, \\ z_k = h(x_k) + v_k, & k = 1,\cdots,N. \end{cases} \quad (1)$$

where w_k is the process noise that drives the dynamic system through the nonlinear state transition function f, and v_k is the observation or measurement noise corrupting the observation of the state through the nonlinear observation function h. The nonlinear function $f(\cdot)$ and $h(\cdot)$ captures the state model and the measurement model. In the Bayesian networks, the problem of variables estimation can be reformulated as that how to recursively compute the posterior density as new observations arrive. The optimal method to recursively update the posterior density as new observations arrive is given by the recursive Bayesian estimation algorithm. The posterior density can be expanded and factored into the following recursive update form [12,15]:

$$p(x_k \mid z_{1:k-1}) = \int p(x_{k-1} \mid z_{1:k-1}) p(x_k \mid x_{k-1}) dx_{k-1}, \quad (2)$$

$$p(x_k \mid z_{1:k}) = \frac{p(x_k \mid z_{1:k-1})\, p(z_k \mid x_k)}{p(z_k \mid z_{1:k-1})}. \quad (3)$$

Despite the optimal recursive solution, the multi-dimensional integrals in (2) are usually only tractable for linear Gaussian systems in which case the close-form recursive solution is given by the Kalman filter. For most general real-world systems, however, the multi-dimensional integrals are intractable and approximate solutions must be used [19]. As discussed in Section 1, sigma-point Kalman filters based on deterministic sampling approaches can be utilized to address these issues. However, these filters may fail in certain nonlinear non-Gaussian problems with multi-modal and/or heavy tailed posterior distributions. Difficulties with the more general setting arise from the occurrence of non-Gaussian process and/or measurement noise. The focus of this paper is on developing a high-order unscented Kalman filter to assimilate nonlinear Gaussian systems. Then we use the high-order unscented Kalman filter to construct the GSF which can approximate the statistical distribution of the non-Gaussian process and/or measurement noise.

3. HIGH-ORDER UNSCENTED KALMAN FILTER

3.1. The unscented transformation

The unscented transformation (UT) is used to calculate the statistics of a random variable which undergoes a nonlinear transformation. It builds on the principle that it is easier to approximate a probability distribution than arbitrary nonlinear function [8]. Consider the propagation of a n dimensional random variable x through an arbitrary nonlinear function,

$$y = g(x). \quad (4)$$

Suppose x has mean $\bar{X}$ and covariance P_x. To calculate the statistics (first two moments) of y using the UT, we proceed as follows: firstly a set of $2n+1$ weighted samples, called sigma-points, $S_i=\{\omega_i, \zeta_i\}$ are deterministically chosen so that they completely capture the true mean and covariance of the prior random

variable x. A selection scheme that satisfies this requirement is:

$$\omega_i = \begin{cases} \kappa/(n+\kappa), & i=0; \\ 1/2(n+\kappa), & i\neq 0; \end{cases} \tag{5}$$

$$\begin{cases} \zeta_0 = \bar{X}, \\ \zeta_i = \bar{X} + (\sqrt{(n+\kappa)P_x})_i, & i=1,\ldots,n, \\ \zeta_i = \bar{X} - (\sqrt{(n+\kappa)P_x})_i, & i=n+1,\ldots,2n, \end{cases} \tag{6}$$

where ω_i is the weight associated with the ith sigma-point and $\sum_{i=0}^{2n}\omega_i = 1$. κ is called the scale coefficient, which can adjust the distance of the sigma points to the mean value and $\sigma_i = (\sqrt{(n+\kappa)P_x})_i$ denotes the ith column of the square root matrix. The numerically efficient Cholesky factorization method is typically used to calculate the matrix square root. Since the matrix square-root of positive semidefinite matrix is not unique, any ortho-normal rotation of the sigma-point set is a valid set. Furthermore, if desired, one can derive a selection scheme that captures higher order moment information such as skew or kurtosis. This, however, will in general require a larger set of sigma-points. Ponomareva [20] considered an augmented method to extend traditional unscented transformation to capture 3rd and 4th moments. Tenne [21] has considered the problem of capturing higher order moments and developed a sigma point selection algorithm that captures the first eight moments of a symmetric one-dimensional distribution using only five points. In the next section, a novel high-order unscented transformation is developed to address the accuracy of linearization by providing a more direct and extensible mechanism for transforming mean, covariance and more high-order moments information.

3.2. High-order unscented transformation

In the unscented Kalman filter, a set of sigma points are carefully selected, which exhibits the same statistical properties to a certain degree as the true distribution of the state. The multidimensional Gaussian estimation can be resolved into some one-dimensional Gaussian estimating problems. Therefore, we can consider the probability density function of the one-dimensional Gaussian distribution at first.

Consider a scalar random variable x having a Gaussian probability density $N(x;0,1)$,

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{7}$$

The expected value of the function $g(x)$ can be approximated as

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x)p(x)dx \approx \sum_{i=1}^{m} g(\sigma_i)\omega_i, \tag{8}$$

where σ_i and ω_i are the sigma point and its associated weight. Provided $g(x)$ is integrable on interval $[-\infty \sim +\infty]$, $p(x)$ here can be thought as the weights of all the continuous points and the right side of the equation can be thought as finite sum of σ_i. Our main objective is to choose the sigma points so that we can obtain a desired level of accuracy at a reasonable computational cost in terms of the number of integrand evaluations required. The function $g(\cdot)$ changes for various systems, thus the sigma points will be determined by the probability density function $p(x)$. Naturally, poles of the probability density function $p(x)$ and its derived functions may be a good choice. We can easily get the poles of the function $p^{(i)}(x)(i \geq 0)$ by calculating zeros of its derived functions $p^{(i+1)}(x)$ as follows:

$$p'(x) = 0 \Rightarrow xe^{-x^2/2} = 0 \Rightarrow x = 0, \tag{9}$$

$$p''(x) = 0 \Rightarrow (x^2-1)e^{-x^2/2} = 0 \Rightarrow x = \pm 1, \tag{10}$$

$$p^{(3)}(x) = 0 \Rightarrow (x^3-3x)e^{-x^2/2} = 0 \Rightarrow x = 0,\ \pm\sqrt{3}, \tag{11}$$

$$\begin{aligned} p^{(4)}(x) = 0 &\Rightarrow (x^4-6x^2+3)e^{-x^2/2} = 0 \\ &\Rightarrow x = \pm\sqrt{3\pm\sqrt{6}}, \end{aligned} \tag{12}$$

$$\begin{aligned} p^{(5)}(x) = 0 &\Rightarrow (x^5-10x^3+15x)e^{-x^2/2} = 0 \\ &\Rightarrow x = 0,\ \pm\sqrt{5\pm\sqrt{10}}. \end{aligned} \tag{13}$$

In this way, poles of higher derivative functions can be got. It can be seen from (9)-(13) that: the function $p(x)$ has one pole $x = 0$. If we use it as sigma point, it will be the same with EKF in which the state distribution is propagated analytically through the first-order linearization of the nonlinear function. The first derivative function $p'(x)$ has two symmetry poles $x = \pm 1$. Poles of the second derivative function $p''(x)$ are $x = 0,\ \pm\sqrt{3}$, which are identical with the sigma points computed by ordinary UT. The third derivative function $p^{(3)}(x)$ has two groups of symmetry poles $x = \pm\sqrt{3\pm\sqrt{6}}$. Poles of the fourth derivative function $p^{(4)}(x)$ are $x = 0,\ \pm\sqrt{5\pm\sqrt{10}}$. Fig. 1 depicts the distribution of these poles in Cartesian Coordinates, where σ_i denotes poles of the ith derivative function.

However, the probability density function $p(x)$ has countless derivative functions. To find out which set of

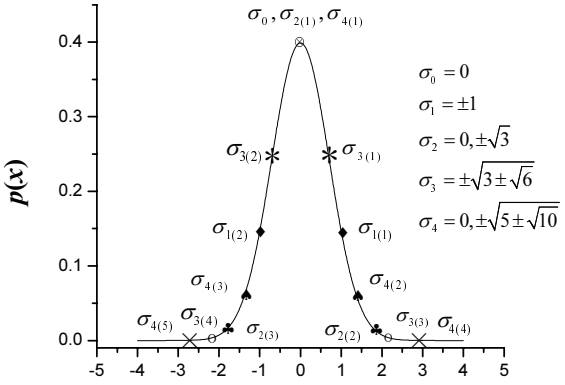


Fig. 1. Sigma points for standard Gaussian distribution.

poles is the most proper one is very important. In the statistics, the 68-95-99.7 rule states that for a normal distribution, nearly all the values lie within 3 standard deviations of the mean. If a point deviates from the average far away, the associated weight will be so small that can be ignored. Therefore, it's not necessarily true that the more points to choose, the better the approximation will be. It should be a compromise between computational complexity and performance gain.

In the following section, how to get the associated weights of each sigma point is introduced. At first, the relationships between the central moments of a normal Gaussian distribution [22] are presented. Suppose variable ς has a Gaussian density $f(\varsigma) \sim N(0, \Omega^2)$, then the kth central moment of ς meet

$$\Sigma_{\varsigma}^{k} = \begin{cases} \Omega^{k}(k-1)(k-3)\cdots 3\times 1, & k = 2,4\cdots 2L, \\ 0, & k = 1,3\cdots 2L-1, \end{cases} \tag{14}$$

where Σ_{ς}^{k} denotes the kth central moment of variable ς, Ω is the variance.

The scalar random variable x in (7) has a Gaussian probability density $N(x;0,1)$ and the required sigma points should have the same statistical characteristics with it, so we can get

$$\begin{cases} \overline{x} = \overline{\sigma} = 0 \\ \Sigma_{x}^{k} = \Sigma_{\sigma}^{k} = 1, \end{cases} \quad k = 1,2,\cdots 2L, \tag{15}$$

where $\overline{x}$ is the mean of variable x, $\overline{\sigma}$ is the mean of sigma points, Σ_{x}^{k} and Σ_{σ}^{k} are respectively the kth central moments of variable x and sigma points. Mean and the kth central moment of sigma points can be computed by

$$\begin{cases} \overline{\sigma} = \sum_{i=1}^{\Upsilon} \omega_i \sigma_i, \\ \Sigma_{\sigma}^{k} = \sum_{i=1}^{\Upsilon} \omega_i (\sigma_i - \overline{\sigma})^{k}, \end{cases} \quad k = 1,2,\cdots 2L, \tag{16}$$

where γ is the number of sigma points. According to (14) -(16), it can be obtained that

$$\begin{cases} \omega_0 + 2\sum_{i=1}^{\Upsilon} \omega_i = 1 \\ \begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \cdots & \sigma_{\Upsilon}^2 \\ \sigma_1^4 & \sigma_2^4 & \cdots & \sigma_{\Upsilon}^4 \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_1^{2L} & \sigma_2^{2L} & \cdots & \sigma_{\Upsilon}^{2L} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \cdots \\ \omega_{\Upsilon} \end{bmatrix} = \begin{bmatrix} 1/2 \\ 3/2 \\ \cdots \\ \dfrac{1\times 3\times\cdots\times(2L-1)}{2} \end{bmatrix}. \end{cases} \tag{17}$$

The associated weights of the sigma points will be convenient to find according to (17).

3.3. Algorithm of high-order unscented Kalman

The HUKF, which consists of augmented sigma points with separate weights, remains the same structure as the ordinary unscented Kalman filter. Sigma points of one-dimensional variable can be promoted to multiple dimensional problems. Suppose x is a n_x dimensional variable with a Gaussian density having mean $\overline{X}$ and variance P_x, the sigma points can be calculated by

$$\begin{cases} \zeta_0 = \overline{X}, \\ \zeta_i = \overline{X} + \sigma_i(\sqrt{P_x})_i, & i = 1,\ldots,n, \\ \zeta_i = \overline{X} - \sigma_i(\sqrt{P_x})_{i-n}, & i = n+1,\ldots,2n, \end{cases} \tag{18}$$

where σ_i is the ith sigma point of one-dimensional Gaussian variable, ζ_i is the ones of multidimensional problem, $(\sqrt{P_x})_i$ denotes the ith column of the square root matrix. According to (17) and (18), high-order moment of the distribution can be captured with a fixed, small number of points. Unlike the Ponomareva and Tenne method, the proposed HUT method avoids to solve nonlinear equations directly. We choose the sigma points by calculating the poles of Gaussian probability density function $p(x)$ and its derived functions. The associated weights can be got by (17). In this way, a set of high-order unscented Kalman filters can be designed. The complete HUKF algorithm that updates the mean $\hat{x}_k$ and covariance P_{x_k} of the Gaussian approximation to the posterior distribution of the states will now be presented:

3.3.1 Time update

1) Assume at time k that the posterior density function $p(x_{k-1} \mid z_{k-1}) = N(\hat{x}_{k-1|k-1}, P_{k-1|k-1})$ is known. Factorize

$$P_{k-1|k-1} = S_{k-1|k-1} S_{k-1|k-1}^{T}.$$

2) Evaluate the HUT sigma points (i = 1,2,…,m)

$$\mathrm{X}_{i,k-1|k-1} = S_{k-1|k-1}\sigma_i + \hat{x}_{k-1|k-1}.$$

3) Evaluate the propagated HUT sigma points (i = 1,2, …, m)

$$\mathrm{X}_{i,k|k-1}^{*} = f(\mathrm{X}_{i,k-1|k-1}, \mu_{k-1}).$$

4) Estimate the predicted state

$$\hat{x}_{k|k-1} = \sum_{i=1}^{m} \omega_i \mathrm{X}_{i,k|k-1}^{*}.$$

5) Estimate the predicted error covariance

$$P_{k|k-1} = Q_{k-1} + \sum_{i=1}^{m} \omega_i (\mathrm{X}_{i,k|k-1}^{*} - \hat{x}_{k|k-1})(\mathrm{X}_{i,k|k-1}^{*} - \hat{x}_{k|k-1})^{T}.$$

3.3.2 Measurement update

1) Factorize

$$P_{k|k-1} = S_{k|k-1} S_{k|k-1}^{T}.$$

2) Evaluate the HUT sigma points (i = 1,2,…,m)

$$\mathrm{X}_{i,k|k-1} = S_{k|k-1}\sigma_i + \hat{x}_{k|k-1}.$$

3) Evaluate the propagated HUT sigma points (i = 1,2, …,*m*)

$$Z_{i,k|k-1} = h(\mathrm{X}_{i,k|k-1}, \mu_k).$$

4) Estimate the predicted measurement

$$\hat{z}_{k|k-1} = \sum_{i=1}^{m} \omega_i Z_{i,k|k-1}.$$

5) Estimate the innovation covariance matrix

$$P_{zz,k|k-1} = R_k + \sum_{i=1}^{m} \omega_i (Z_{i,k|k-1} - \hat{z}_{k|k-1})(Z_{i,k|k-1} - \hat{z}_{k|k-1})^T .$$

6) Estimate the cross-covariance matrix

$$P_{xz,k|k-1} = \sum_{i=1}^{m} \omega_i (X_{i,k|k-1} - \hat{x}_{k|k-1})(Z_{i,k|k-1} - \hat{z}_{k|k-1})^T .$$

7) Estimate the Kalman gain

$$W_k = P_{xz,k|k-1} P_{zz,k|k-1}^{-1}.$$

8) Estimate the updated state

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + W_k (z_k - \hat{z}_{k|k-1}).$$

9) Estimate the corresponding error covariance

$$P_{k|k} = P_{k|k-1} - W_k P_{zz,k|k-1} W_k^T .$$

4. GS-HUKF ALGORITHM

In this section we present a further refinement of the HUKF called the Gaussian sum high-order unscented Kalman filter. The GS-HUKF combines the HUKF and Gaussian sum filter for the time update and measurement update step. The GS-HUKF uses a finite Gaussian mixture model (GMM) representation of the posterior filtering density, which is recovered from the weighted Gaussian items based on measurement update stage.

4.1. Principle of the GSF

The theoretical determination of a weighted Gaussian sum approximation for a non-Gaussian density is well detailed in [15]. It can be shown that as the number of Gaussian components increases, the Gaussian sum approximation converges uniformly to any probability density function. Here, we only briefly review the principle idea [23]. For a given density function $P(x)$, an approximation $P_m(x)$ can be described by:

$$P_m(x) \approx \sum_{i=1}^{m} \rho_i N(x; u_i, \Sigma_i), \tag{19}$$

where m is the number of mixing components, ρ_i are the mixing weights and $N(x; u_i, \Sigma_i)$ are Gaussian density functions with mean u_i and positive definite covariance matrices Σ_i.

To choose the mixture parameters, such as the weights ρ_i, the means u_i and variances Σ_i, that give the optimal approximation $P_m(x)$ to the density function $P(x)$, we have to minimize a given distance function, for example, the L^n norm:

$$\left\| P - P_m \right\|^n = \int_{-\infty}^{+\infty} | P(x) - \sum_{i=1}^{m} \rho_i N(x - u_i, \Sigma_i) |^n dx \tag{20}$$

with n usually chosen equal to 2.

4.2. The GS-HUKF algorithm

Non-Gaussian noise densities can be approximated empirically by Gaussian-sums as closely as possible. Consequently, for a state-space model with Gaussian-sum additive noise sources, such as the one described in (1), it is possible to obtain both the predicted and posterior densities as Gaussian-sums. The proposed GS-HUKF filter is, in some sense, a steady state of the GSF. It consists of the following steps:

1) Modeling of Non-Gaussian Density. Assume at time k that the additive process and measurement noise are both available as approximate Gaussian sums:

$$P(w_k) \approx \sum_{i=1}^{I_k} \beta_{ki} p_i(w_k) = \sum_{i=1}^{I_k} \beta_{ki} N(w_k; \overline{w}_{ki}, Q_{ki}), \tag{21}$$

$$P(v_k) \approx \sum_{j=1}^{J_k} \mu_{kj} p_j(v_k) = \sum_{j=1}^{J_k} \mu_{kj} N(v_k; \overline{v}_{kj}, R_{kj}), \tag{22}$$

where I_k, J_k are the number of components, β_{ki} and μ_{kj} are the mixing weights satisfying:

$$\sum_{i=1}^{I_k} \beta_{ki} = 1 \quad \text{and} \quad \sum_{j=1}^{J_k} \mu_{kj} = 1.$$

In addition, the prior density at time zero is also assumed to be a Gaussian sum:

$$P(x_0) \approx \sum_{\tau=1}^{T_0} \alpha_{0\tau} p_\tau(x_0) = \sum_{\tau=1}^{T_0} \alpha_{0\tau} N(x_0; x_{0\tau}, P_{0\tau}), \tag{23}$$

where α_{0i} are nonnegative constants and $\sum_{\tau=1}^{T_0} \alpha_{0\tau} = 1$. The GS-HUKF is required to compute the time and measurement update steps at each sampling instant and these two steps are derived in the sequel.

2) Time update. Suppose at time k the Gaussian sum approximation of the posterior density $P(x_{k-1}|z_{1:k-1})$ is known and given by:

$$P(x_{k-1} \mid z_{1:k-1}) = \sum_{\tau=1}^{T_{k-1}} \alpha_{(k-1|k-1)\tau} N(x_{k-1}; x_{(k-1|k-1)\tau}, P_{(k-1|k-1)\tau}). \tag{24}$$

For a process noise model of (21), the transition prior $P(x_k \mid x_{k-1})$ can be obtained as

$$P(x_k \mid x_{k-1}) = \sum_{i=1}^{I_k} \beta_{ki} N(x_k; f(x_{k-1}) + \overline{w}_{ki}, Q_{ki}). \tag{25}$$

Combined with (2), the predicted probability density function of x_k can be calculated by

$$P(x_k \mid z_{1:k-1}) = \int P(x_k \mid x_{k-1}) P(x_{k-1} \mid z_{k-1}) dx_{k-1}$$

$$= \sum_{\tau=1}^{T_{k-1}} \sum_{i=1}^{I_k} a_{(k-1|k-1)\tau} \beta_{ki} \int N(x_k; f(x_{k-1}) + \overline{w}_{ki}, Q_{ki}) \times N(x_{k-1}; \hat{x}_{(k|k-1)\tau}, P_{(k|k-1)\tau}) dx_{k-1}. \quad (26)$$

The integral on the right side is approximated by a Gaussian sum in x_k using time update of the HUKF. The predicted density $P(x_k|z_{1:k-1})$ is therefore approximated as

$$P(x_k \mid z_{1:k-1}) = \sum_{r=1}^{T_{k-1}I_k} \alpha_{(k|k-1)r} N(x_{k|k-1}; \hat{x}_{(k|k-1)r}, P_{(k|k-1)r}), \quad (27)$$

where

$$\alpha_{(k|k-1)\tau i} = \alpha_{(k-1|k-1)\tau} \beta_{ki},$$

$$\hat{x}_{(k|k-1)\tau i} = \overline{w}_{ki} + \sum_{\lambda=1}^{m} \omega_\lambda X^*_{(k|k-1)\tau i},$$

$$P_{(k|k-1)\tau i} = Q_{ki} + \sum_{\lambda=1}^{m} \omega_\lambda (X^*_{(k|k-1)\tau i} - \hat{x}_{(k|k-1)\tau i} - \overline{w}_{ki}) \times (X^*_{(k|k-1)\tau i} - \hat{x}_{(k|k-1)\tau i} - \overline{w}_{ki})^T,$$

$$X^*_{(k|k-1)\tau i} = f(\hat{x}_{(k-1|k-1)\tau} + \sqrt{P_{(k-1|k-1)\tau}}\sigma_\lambda).$$

Step of time update ends. Prediction probability density is expressed by $T_{k-1} I_k$ Gaussian items in (27).

3) Measurement update. Assume that the measurement noise sequence is non-Gaussian and modeled by a Gaussian mixture as given by (22). The likelihood probability density $P(z_k|x_k)$ can be described as

$$P(z_k \mid x_k) = \sum_{j=1}^{J_k} \mu_{kj} N(z_{k|k-1}; h(x_{k|k-1}) + \overline{v}_{kj}, R_{kj}). \quad (28)$$

The posterior density $P(x_k|z_{1:k})$ can be approximated by a Gaussian sum after receiving the measurement z_k using the measurement update step of the HUKF:

$$P(x_k \mid z_{1:k}) \approx \sum_{r=1}^{T_{k-1}I_k} \sum_{j=1}^{J_k} \alpha_{(k|k)r} N(x_k; \hat{x}_{(k|k)r}, P_{(k|k)r}) = \sum_{r=1}^{T_{k-1}I_kJ_k} \alpha_{(k|k)r} N(x_k; \hat{x}_{(k|k)r}, P_{(k|k)r}), \quad (29)$$

where

$$\alpha_{(k|k)\tau ij} = \frac{a_{(k-1|k-1)\tau} \Omega_{k\tau i} \mu_{kj}}{\sum_{j=1}^{J_k} \sum_{r=1}^{T_{k-1}I_k} a_{(k-1|k-1)\tau} \Omega_{k\tau i} \mu_{kj}}, \quad (30)$$

$$\hat{x}_{(k|k),\tau i} = \hat{x}_{(k|k-1)\tau} + K_{k,\tau i}(z_k - \hat{z}_{(k|k-1)\tau}), \quad (31)$$

$$P_{(k|k)\tau ij} = P_{(k|k-1)\tau i} - K_{k,\tau i} P_{zz,\tau i} K_{k,\tau i}, \quad (32)$$

with

$$K_{k,\tau i} = P_{xz,\tau ij} P^{-1}_{zz,\tau ij},$$

$$P_{xz,\tau ij} = \sum_{\lambda=1}^{m} \omega_\lambda (X_{\lambda(k|k-1)\tau} - \hat{x}_{(k|k-1)\tau}) \times (Z_{\lambda(k|k-1)\tau} - \hat{z}_{(k|k-1)\tau} - \overline{v}_{kj})^T,$$

$$P_{zz,\tau ij} = R_{kj} + \sum_{\lambda=1}^{m} \omega_\lambda (Z_{\lambda(k|k-1)\tau} - \hat{z}_{(k|k-1)\tau} - \overline{v}_{kj}) \times (Z_{\lambda(k|k-1)\tau} - \hat{z}_{(k|k-1)\tau} - \overline{v}_{kj})^T,$$

$$\hat{z}_{(k|k-1)\tau ij} = \overline{v}_{kj} + \sum_{\lambda=1}^{m} \omega_\lambda Z_{\lambda(k|k-1)\tau},$$

$$Z_{\lambda(k|k-1)\tau} = h(X_{\lambda(k|k-1)\tau}),$$

$$X_{\lambda(k|k-1)\tau} = \sqrt{P_{(k|k-1)\tau}}\sigma_\lambda + \hat{x}_{(k|k-1)\tau},$$

$$\Omega_{k\tau i} = N(z_k; \hat{z}_{(k|k-1)\tau i}, P_{zz,\tau i}).$$

Then the final state estimate $\hat{x}_{k|k}$ in the minimum-mean-squared-error sense and the associated covariance $P_{k|k}$ can be computed by:

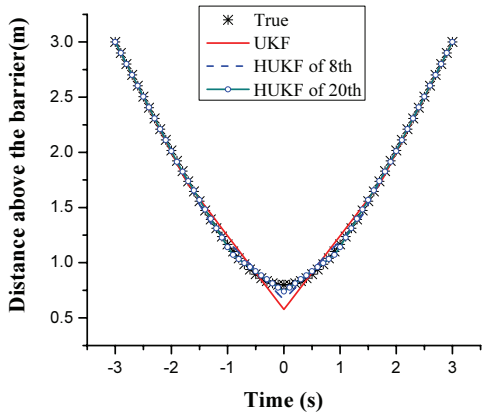
$$\hat{x}_{k|k} = \sum_{r=1}^{T_{k-1}I_kJ_k} \alpha_{(k|k)r} \hat{x}_{(k|k)r}, \quad (33)$$

$$P_{k|k} = \sum_{r=1}^{T_{k-1}I_kJ_k} \alpha_{(k|k)r} P_{(k|k)r}. \quad (34)$$

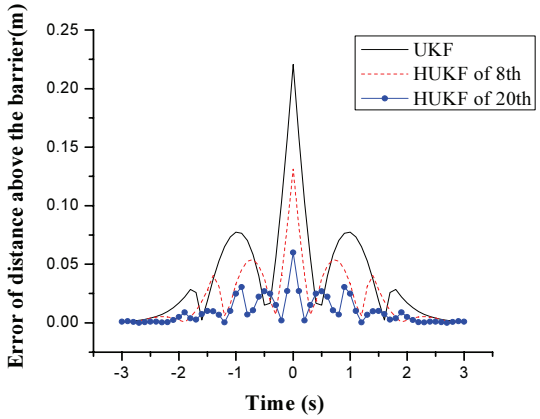
It has been shown that the number of mixing components in the GMM representation grows from T_{k-1} to $T_{k-1} \times I_k$ in the time update step and from $T_{k-1} \times I_k$ to $T_{k-1} \times I_k \times J_k$ in the subsequent measurement update step. Over time, this will lead to an exponential increase in the total number of mixing components and must be addressed by a mixing-component reduction scheme. In addition, some weights in the Gaussian sum approximation, i.e., some $\alpha_{(k|k-1)r}$ in (27) and some $\alpha_{(k|k)r}$ in (29) may be sufficiently small compared to the others so that they can be simply neglected or combined into a single term without seriously affecting the approximation. Another possible strategy is to conduct probability density function re-approximation, i.e., one uses a new Gaussian mixture model, with the specified number of Gaussian distributions, to approximate the prior or the posterior probability density function that has already been expressed in terms of a Gaussian sum approximation. To estimate the parameters of the new Gaussian mixture model (i.e., weights, means and covariances of individual Gaussian distributions), we adopt the expectation-maximization (EM) algorithm [10,24], which is an iterative method, to decrease the computational cost of the GS-HUKF algorithm.

5. SIMULATION

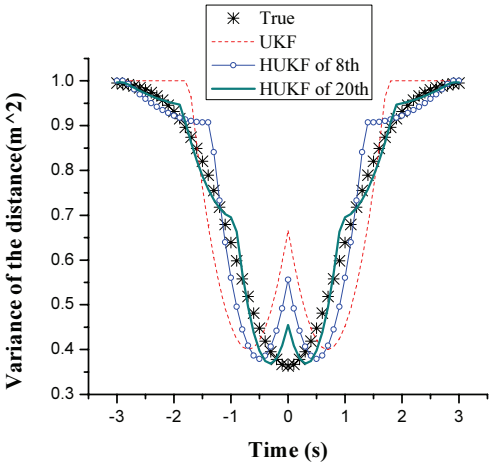
The proposed high-order unscented transform and GS-HUKF algorithm were validated by two examples. In the first example, the ordinary UT and HUT were applied to a simple one dimensional function with a discontinuous first derivative. In the second example, the performances of GS-HUKF were assessed by means of Monte Carlo simulations, GS-UKF, GS-QKF and GS-PF, by applying the algorithms for a univariate nonstationary growth model in the presence of heavy-tailed additive gamma distributed process noise [9].



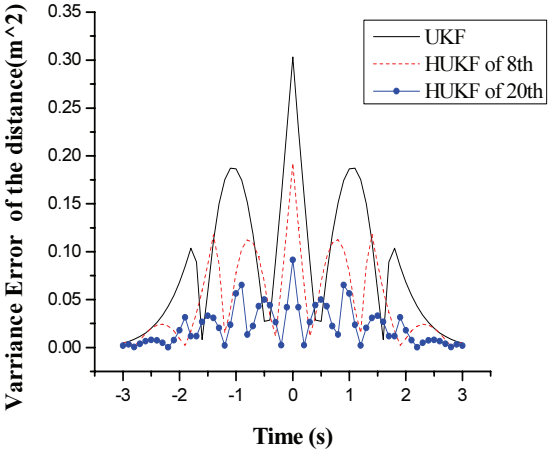
(a) Mean distance above the barrier. It can be seen that the approximated results of ordinary UKF (solid red), HUKF of 8th (dashed blue) and HUKF of 20th (circle green) can match the analytic solution (*) well. However, the HUKF of 20th get the best estimation.



(b) Error of the estimated mean distance. The differences were calculated between approximations and the analytic solution.



(c) Variance of the distance above the barrier. Same with the Mean distance, the approximated results of ordinary UKF (dot red), HUKF of 8th (circle blue) and HUKF of 20th (solid green) can match the analytic solution (*) well, the HUKF of 20th has the best estimation.



(d) Error of the estimated variance of the distance. The differences between estimated results and the analytic solution were calculated.

Fig. 2. The analytical solution and the estimated results.

5.1. Distance of a particle above a barrier

A particle with an initial velocity of -1 strikes a barrier at the origin at time zero and bounces elastically [8,18]. Thus,

$$y = g(x,t) = | x + v_0 t | . \tag{35}$$

The original distance from the particle to the barrier is assumed to be normally distributed with unit variance. The exact values of the distance and its variance can be calculated by

$$\bar{y} = E(Y) = \int_{-\infty}^{v_0 t} (v_0 t - z) \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz + \int_{v_0 t}^{+\infty} (z - v_0 t) \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

$$= v_0 t \times erf\left(\frac{v_0 t}{\sqrt{2P_x}} \right) + \sqrt{\frac{2P_x}{\pi}} e^{-\frac{1}{2}\frac{(v_0 t)^2}{P_x}}, \tag{36}$$

$$P_y = E\left\{ (Y - \bar{y})^2 \right\} = P_x + (v_0 t)^2 - \bar{y}^2. \tag{37}$$

The goal of the approximation is to estimate the mean $\bar{y}$ and the variance P_y as accurately as possible using a function of the sigma points and the weights ω_i. Using sigma points got by HUT in Table 1, the estimation result is shown in Fig. 2. Figs. 2(a) and 2(b) illustrate the mean distance above the barrier and error of the estimated mean distance, respectively. Figs. 2(c) and 2(d) show variance of the distance above the barrier and error of the estimated variance of the distance, respectively.

Table 1. The sigma points of HUT.

HUT	Sigma point	Weight	NO.
HUT of 4th	$\sigma_1 = \sqrt{3}$	$\omega_0 = \frac{2}{3}, \quad \omega_1 = \frac{1}{6}$	3
HUT of 8th	$\sigma_1 = \sqrt{5-\sqrt{10}}$ $\sigma_2 = \sqrt{5+\sqrt{10}}$	$\omega_0 = \frac{8}{15}$ $\omega_1 = (7+2\sqrt{10})/60$ $\omega_2 = (7-2\sqrt{10})/60$	5
HUT of 20th	$\sigma_1 = 0.9289$ $\sigma_2 = 1.8760$ $\sigma_3 = 2.8651$ $\sigma_4 = 3.9362$ $\sigma_5 = 5.1880$	$\omega_0 = 0.3694$ $\omega_1 = 0.2422$ $\omega_2 = 0.0661$ $\omega_3 = 0.0067$ $\omega_4 = 1.9567e-4$ $\omega_5 = 8.1218e-7$	11

As shown in Table 1, each set of sigma points are symmetric one-dimensional distribution. Sigma points and relevant weights for the HUT of 4th are same with the ordinary UT which can capture the first 4th moments of a nonlinear function at most. HUT of 8th has five sigma points and can capture the first 8th moments at most. Eleven sigma points are needed to capture the first 20th moments.

Fig. 2 shows the analytical mean, deviation and the approximate solutions of the distance above the barrier. It can be seen that all the unscented filters reasonably approximate the mean and variance of the particles distance, whereas the HUKF of 20th can better describe the deviation due to its capability of matching the higher order moments of nonlinear functions. The number of the sigma points is proportionate to the estimation accuracy. However, the computational complexity will also increase. So, proper number of points should be selected based on the specific problem in hand.

5.2. Univariate nonstationary growth model

Univariate nonstationary growth model is widely accepted in econometrics and it is formulated as

$$x_k = 0.5x_{k-1} + 25\frac{x_{k-1}}{1+x_{k-1}^2} + 8\cos(1.2(k-1)) + w_{k-1}, \quad (40)$$

$$z_k = \frac{x_k^2}{20} + v_k, \qquad k = 1, \cdots, N, \quad (41)$$

where the process noise w_k is assumed to be distributed as a heavy-tailed gamma function given by $w_k \sim \Gamma(3, 2)$, where Γ denotes the gamma distribution, and v_k is zero-mean Gaussian white noise with variance 1. The reference data are generated using $x_0 = 0.1$ and $N = 100$. The state x_k is to be estimated from the measurement data z_k.

The performance is compared using the mean-square error (MSE) defined by

$$MSE = \frac{1}{N}\sum_{n=1}^{N}(x_n - x_{n|n})^2. \quad (42)$$

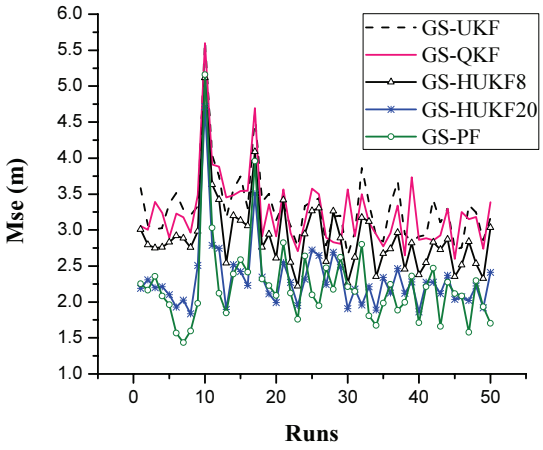


Fig. 3. MSEs of each filter across 50 random runs.

Table 2. Estimation result averaged over 50 Monte Carlo runs.

Algorithm	Mean	Variance	Time(s)
GS-UKF	3.2987	0.2331	1.48
GS-QKF	3.2835	0.2317	1.50
GS-HUKF8	2.8951	0.2469	1.69
GS-HUKF20	2.3100	0.2265	2.55
GS-PF	2.2271	0.2446	38.18

The MSE of 50 Monte Carlo runs are plotted in Fig. 3. In each run, the actual initial state x_0 is assumed to be a uniformly distributed random variable in the interval [0, 1]. We compare the performance of GS-HUKF against GS-UKF, GS-QKF and GS-PF. In the algorithm of GS-HUKF, sigma points are chosen from Table 1. 3-point QKF is used in GS-QKF. 300 particles SIR-PF is adopted in GS-PF. It can be seen that MSEs of GS-UKF and GS-QKF are larger than other filters. There is no discernible difference in accuracy between them. That is because UKF and 3-point QKF are analytically identical to each other for the one dimensional systems [18]. However, the limitation of QKF for practical use is that it suffers from the curse of dimensionality. The GS-HUKF8 has a noticeable improvement over GS-UKF and GS-QKF. The GS-PF and GS-HUKF20 have better tracking performances than GS-UKF, GS-QKF and GS-HUKF8.

Table 2 summarizes the performance of the different filters averaged over 50 randomly initialized Monte Carlo runs. It shows the mean and variances of the MSE of the state estimates as well as the average processing time of each filter. Since the processing time for each algorithm is directly related to its computational complexity, Table 2 clearly indicates that the GS-HUKF algorithms have the same order of computational complexity as the GS-UKF, but much better estimation performance. Conversely, for the same level of estimation performance as the GS-UKF, the GS-PF realizes this at a much higher computational cost. The best performance is achieved by the GS-HUKF20 which better models the non-Gaussian nature of the process noise distribution.

6. CONCLUTION

The motivation for this paper is to derive a more accurate nonlinear non-Gaussian filter that could be applied to solve nonlinear non-Gaussian dynamic state space problems. One significant contribution lies in the proposed high-order unscented Kalman filters, which enable the estimation more accurately by choosing the sigma points based on the HUT. The new high-order unscented transformation design the sigma points according to the property density function of variables which is more flexible and has integrated advantages with respect to time complexity and computational accuracy. In addition, the Gaussian sum high-order unscented Kalman filter algorithm for nonlinear and non-Gaussian DSSM is presented. Simulation results exhibit a significant improvement of the GS-HUKFs over other nonlinear filtering methods, namely, GS-UKF, GS-QKF and GS-PF, to solve nonlinear non-Gaussian filtering problems.

REFERENCES

[1] J. T. Horwood and A. B. Poore, "Adaptive Gaussian sum filters for space surveillance," *IEEE Trans. on Automatic Control*, vol. 56, no. 8, pp. 1777-1790, August 2011.

[2] W. Yi, M. Morelande, L. Kong, and J. Yang, "A computationally efficient particle filter for multi-target tracking using an independence approximation," *IEEE Trans. on Signal Processing*, vol. 61, no. 4, pp. 843-856, February 2013.

[3] P. Wu, X. Li, and Y. Bo, "Iterated square root unscented Kalman filter for maneuvering target tracking using TDOA measurements," *International Journal of Control, Automation, and Systems*, vol. 11, no. 4, pp. 761-767, August 2013.

[4] K.-P. Lim, "Ranking market efficiency for stock markets: a nonlinear perspective," *Physica A: Statistical Mechanics and its Applications*, vol. 376, no. 10, pp. 445-454, March 2007.

[5] H. J. Kushner, "Approximations to optimal nonlinear filters," *IEEE Trans. on Automatic Control*, vol. 12, no. 5, pp. 546-556, October 1967.

[6] H. Cox, "On the estimation of state variables and parameters for noisy dynamic systems," *IEEE Trans. on Automatic Control*, vol. 9, no. 1, pp. 5-12, January 1964.

[7] M. Nørgaard, N. K. Poulsen, and O. Ravn, "New developments in state estimation for nonlinear systems," *Automatica*, vol. 36, no. 11, pp. 1627-1638, November 2000.

[8] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Trans. on Automatic Control*, vol. 45, no. 3, pp. 477-482, March 2000.

[9] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. on Automatic Control*, vol. 45, no. 8, pp. 910-927, May 2000.

[10] I. Arasaratnam, S. Haykin, and R. J. Elliott, "Discrete-time nonlinear filtering algorithms using Gauss-Hermite quadrature," *Proc. of the IEEE*, vol. 95, no. 5, pp. 953-977, May 2007.

[11] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Trans. on Automatic Control*, vol. 54, no. 6, pp. 1254-1269, June 2009.

[12] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 4, pp. 107-113, April 1993.

[13] J. Carpener, P. Clifford, and P. Fearnhead, "Improved particle filter for nonlinear problems," *IEE Proc.-Radar, Sonar and Navigation*, vol. 146, no. 2, pp. 1-7, February 1999.

[14] H. Zhang, G. Feng, and C. Han, "Linear estimation for random delay systems," *Systems & Control Letters*, vol. 60, no. 7, pp. 450-459, July 2011.

[15] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. on Automatic Control*, vol. 17, no. 4, pp. 439-448, August 1972.

[16] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM review*, vol. 26, no. 2, pp.195-239, April 1984.

[17] J. H. Kotecha and P. M. Djuric, "Gaussian sum particle filtering," *IEEE Trans. on Signal Processing*, vol. 51, no. 10, pp. 2602-2612, October 2003.

[18] Y. Wu, D. Hu, M. Wu, and X. Hu, "A numerical-integration perspective on Gaussian filters," *IEEE Trans. on Signal Processing*, vol. 54, no. 8, pp. 2910-2921, August 2006.

[19] C. Han, W. Wang, and Y. Zhang, "White noise estimators for networked systems with packet dropouts," *International Journal of Control, Automation, and Systems*, vol. 60, no. 7, pp. 1187-1195, December 2013.

[20] K. Ponomareva, P. Date, and Z. Wang, "A new unscented Kalman filter with higher order moment-matching," *Proc. of Mathematical Theory of Networks and Systems*, pp. 1609-1613, 2010.

[21] D. Tenne and T. Singh, "The higher order unscented filter," *Proc. of the American Control Conference*, pp. 2441-2446, 2003.

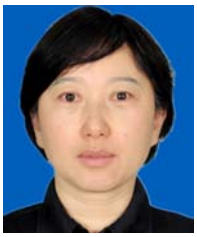
[22] M. Fisz, *Probability Theory and Mathematical Statistics*, Wiley Press, Ann Arbor, 1963.

[23] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall Press, Englewood Cliffs, 1979.

[24] G. Terejanu, P. Singla, and T. Singh, "Adaptive Gaussian sum filter for nonlinear Bayesian estimation," *IEEE Trans. on Automatic Control*, vol. 56, no. 9, pp. 2151-2156, September 2011.

Lei Wang received his M.S. degree from Wuhan University of Science and Technology, China, in 2009. He is currently a teaching assistant at Anhui Science and Technology University and has been pursuing his Ph.D. degree since 2011 at the School of Instrument Science and Engineering, Southeast University, China.

His research interests include nonlinear filtering and estimation, sensor fusion, and statistical signal processing.



Xianghong Cheng received her M.S. and Ph.D. degrees both in Precision Instrument and Mechanics from the Southeast University, China, in 1991 and 1998, respectively. She is currently a Professor in the School of Instrument Science and Engineering, Southeast University. As the first author or corresponding author she has published more than 40 peer reviewed technical papers in archival journals and conference proceedings. Her research interests include navigation system, general estimation theory, nonlinear optimal control theory, and nonlinear filtering.