PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Generalizing the unscented Kalman filter for state estimation

Quade Butler, Waleed Hilal, Brett Sicard, Youssef Ziada, S. Andrew Gadsden

Quade Butler, Waleed Hilal, Brett Sicard, Youssef Ziada, S. Andrew Gadsden, "Generalizing the unscented Kalman filter for state estimation," Proc. SPIE 12547, Signal Processing, Sensor/Information Fusion, and Target Recognition XXXII, 1254703 (14 June 2023); doi: 10.1117/12.2664227

Event: SPIE Defense + Commercial Sensing, 2023, Orlando, Florida, United States

Generalizing the Unscented Kalman Filter for State Estimation

Quade Butler[a], Waleed Hilal[a], Brett Sicard[a], Youssef Ziada[b], and S. Andrew Gadsden[a]

[a]McMaster University, 1280 Main St W, Hamilton, Canada
[b]Ford Motor Company, 36200 Plymouth Road, Livonia, USA

ABSTRACT

The recent generalized unscented transform (GenUT) is formulated into a recursive Kalman filter framework. The GenUT constrains $2n+1$ sigma points and their weights to match the first four statistical moments of a probability distribution. The GenUT integrates well into the unscented Kalman filter framework, creating what we call the generalized unscented Kalman filter (GUKF). The measurement update equations for the skewness and kurtosis are derived within. Performance of the GUKF is compared to the UKF under two studies: noise described by a Gaussian distribution and noise described by a uniform distribution. The GUKF achieves lower errors in state estimation when the UKF uses the heuristic tuning parameter $\kappa = 3 - n$. It is also stated that when the parameter κ is tuned to an optimal value, the UKF performs identically to the GUKF. The advantage here is that GUKF requires no such tuning.

Keywords: Generalized unscented Kalman filter, Generalized unscented transform, State estimation, Nonlinear filter

1. INTRODUCTION

State estimation in a Bayesian filtering framework is exactly solvable under specific assumptions. When the stochastic models for the system and measurement are linear, and the process and measurement noise are zero mean and uncorrelated, the well-known Kalman filter (KF) provides the optimal state estimate. However, most applications are nonlinear and require appropriate methods such as the *extended* Kalman filter (EKF) and the *unscented* Kalman filter (UKF). Despite their achievements, the EKF can struggle with highly nonlinear systems. The UKF (more specifically, the unscented transform) was designed around symmetric probability density functions (pdf), and could potentially struggle with asymmetric distributions. Thus, more general nonlinear filtering methods are needed.

Many works have approximated arbitrary distributions by Gaussian densities. As a result, the Bayesian filtering equations can be approximated using numerical integration methods. These include Gauss-Hermite quadrature.[1,2] In addition, Monte Carlo integration has been used as in the Gaussian particle filter.[3] When the Gaussian assumptions do not hold up well, others have used generalized Edgeworth series to represent approximate non-Gaussian densities.[4–6] Point-mass methods approximate the densities using point masses located on a rectangular grid.[7] This allows for evaluating the Bayesian filtering equations numerically using discrete nonlinear convolution.

Monte Carlo methods have been used for nonlinear filtering with great success. These include ensemble Kalman filters and particle filters.[8–10] Comparatively, Monte Carlo filters can handle more drastic nonlinearities over typical Kalman filter methods at the cost of more computational complexity.

The difficulty of filtering nonlinear systems is mainly due to the challenge of approximating nonlinear functions. The EKF achieves this by a first-order approximation.[11] Despite this accomplishment, the EKF only performs well on nonlinear systems that behave linearly in a local neighborhood, which is often not the case. Two improvements to the EKF include the iterated EKF[12] and the second-order EKF.[13] However, these methods

Further author information: (Send correspondence to S.A. Gadsden)
S.A. Gadsden: E-mail: gadsdesa@mcmaster.ca,
Q. Butler: E-mail: butlerq@mcmaster.ca

Signal Processing, Sensor/Information Fusion, and Target Recognition XXXII, edited by Ivan Kadar, Erik P. Blasch, Lynne L. Grewe, Proc. of SPIE Vol. 12547, 1254703
 · doi: 10.1117/12.2664227

are more computationally expensive, and in the case of the second-order EKF, numerical instability is a large concern. The central difference KF takes a spin on the EKF in that it does not require derivatives (i.e., calculation of Jacobian matrices).[14] Quadrature Kalman filters are another class of innovations-based nonlinear filters.[15–17] These methods use Gauss Hermite quadrature rules to compute Gaussian-weighted integrals numerically.

Another improvement to nonlinear filtering was made with UKF. Instead of approximating the nonlinear function of a state, the UKF approximates the state's transformed probability density.[18,19] This method generally outperforms the EKF at a similar computational cost. However, UKF performance tends to degrade if the modelled system is highly nonlinear. More Kalman filter variants have been proposed to improve upon the UKF by using specialized integral formulations to calculate higher-order statistical moments, such as the cubature KF.[20–22]

The basis of the present work is to investigate a filter than can capture higher-order statistical moments of the estimation error. It is important to recognize that there exists various forms or modifications of the unscented transform, sigma point selection algorithm, and the UKF that can achieve this.[19,23] However, many of these methods require more than $2n+1$ sigma points. In this work, we emphasize a filter that uses $2n+1$ sigma points.

2. THE UNSCENTED TRANSFORM AND THE UKF

2.1 The Unscented Transform

Consider a random variable $x \in \mathbb{R}^n$ and nonlinear transformation $y = g(x)$, $g : \mathbb{R}^n \to \mathbb{R}^m$. The unscented transform (UT) estimates the pdf of the nonlinearly transformed variate x by propagating a deterministic set of $2n+1$ sigma points through the transformation y. The sigma points are chosen such that they have the same mean and covariance as the prior distribution. The sigma points are given by

$$\begin{aligned} \mathcal{X}_0 &= \hat{x}\,, & W_0 &= \frac{\kappa}{n+\kappa}\,, \\ \mathcal{X}_i &= \hat{x} + \left(\sqrt{(n+\kappa)P_{xx}}\right)_i, & W_i &= \frac{1}{2(n+\kappa)}\,, \\ \mathcal{X}_{i+n} &= \hat{x} - \left(\sqrt{(n+\kappa)P_{xx}}\right)_i, & W_{i+n} &= \frac{1}{2(n+\kappa)}\,, \end{aligned} \tag{1}$$

where $\mathcal{X}_i$ is the ith sigma point, $\hat{x}$ is the mean of x, n is the dimension of x, P_{xx} is the covariance matrix of x, W_i is a weighting factor, and κ is a tuning parameter. The term $\left(\sqrt{(n+\kappa)P_{xx}}\right)_i$ is the ith column or row of the matrix $\sqrt{(n+\kappa)P_{xx}}$.

The sigma points are then propagated through the nonlinear transformation to yield the transformed sigma points,

$$\mathcal{Y}_i = g(\mathcal{X}_i). \tag{2}$$

The mean and covariance of y is then approximated as

$$\hat{y} = \sum_{i=0}^{2n} W_i \mathcal{Y}_i, \tag{3}$$

$$P_{yy} = \sum_{i=0}^{2n} W_i \left[\mathcal{Y}_i - \hat{y}\right]\left[\mathcal{Y}_i - \hat{y}\right]^T. \tag{4}$$

The unscented transform, as presented here, is capable of capturing the transformed mean and covariance with third-order accuracy (with respect to a Taylor series expansion). The sigma point selection in Eq. (1) was designed with the standard Gaussian distribution in mind. Notice that the sigma points are symmetric about the mean. This implies that the sigma points can capture higher-order moments and exactly captures all odd-ordered moments. The reason being that, for a symmetrical density like a Gaussian, the odd-ordered central moments are zero.

Higher-order statistical moments can be captured via the tuning parameter κ. Choosing $\kappa = 3 - n$ minimizes the difference between the moments of a standard Gaussian and the sigma points up to kurtosis.[24] However, the kurtosis is not guaranteed to be precisely matched. When the distribution is non-Gaussian, it can be uncertain whether these statistical qualities of the UT hold true. Although it has been shown that the first three moments could be captured by the UT which provides more information when estimating the mean.[25] Despite this, the work in Ref. 23 shows that the UT does not generalize to an arbitrary probability density.

2.2 The Unscented Kalman Filter

The UT is the core of the UKF which can outperform the EKF at a similar computational cost. The well-known UKF algorithm is presented in this section.

We consider here a discrete nonlinear process and measurement model with additive noise sequences,

$$\begin{aligned} x_k &= f(x_{k-1}) + w_{k-1} \,, \\ y_k &= h(x_k) + v_k \,, \end{aligned} \tag{5}$$

where k is the discrete time step, $x_k \in \mathbb{R}^n$ is the system state, $y_k \in \mathbb{R}^m$ is the measurement, f is the process model, h is the measurement model, $w_k \in \mathbb{R}^n$ is the process noise, and $v_k \in \mathbb{R}^m$ is the measurement noise. It is assumed that the noise vectors w_k and v_k are zero mean and uncorrelated, i.e.,

$$\begin{aligned} \mathbb{E}[w_i w_j^T] &= \delta_{ij} Q, \\ \mathbb{E}[v_i v_j^T] &= \delta_{ij} R, \\ \mathbb{E}[w_i v_j^T] &= 0, \end{aligned}$$

for all i, j. Here $\mathbb{E}[\cdot]$ denotes the expectation operator, δ_{ij} is the Kronecker delta function, Q is the process noise covariance, and R is the measurement noise covariance.

The UKF algorithm is as follows:

1. Initialization:

$$\begin{aligned} \hat{x}_{0|0} &= \mathbb{E}[x_0] \\ P_{0|0} &= \mathbb{E}[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T] \end{aligned}$$

2. Sigma point selection according to Eq. (1).

3. The prediction stage:

$$\begin{aligned} \mathcal{X}_{i,k|k-1} &= f\left(\mathcal{X}_{i,k-1}\right) \\ \hat{x}_{k|k-1} &= \sum_{i=0}^{2n} W_i \mathcal{X}_{i,k|k-1} \\ P_{k|k-1} &= \sum_{i=0}^{2n} W_i \left[\mathcal{X}_i - \hat{x}\right] \left[\mathcal{X}_i - \hat{x}\right]^T \\ \mathcal{Y}_{i,k|k-1} &= h\left(\mathcal{X}_{i,k-1}\right) \\ \hat{y}_{k|k-1} &= \sum_{i=0}^{2n} W_i \mathcal{Y}_{i,k|k-1} \end{aligned}$$

4. The update stage:

$$\begin{aligned}
P_\nu &= \sum_{i=0}^{2n} W_i \left(\mathcal{Y}_{i,k|k-1} - \hat{y}_{k|k-1}\right)\left(\mathcal{Y}_{i,k|k-1} - \hat{y}_{k|k-1}\right)^T \\
P_{xy} &= \sum_{i=0}^{2n} W_i \left(\mathcal{X}_{i,k|k-1} - \hat{x}_{k|k-1}\right)\left(\mathcal{Y}_{i,k|k-1} - \hat{y}_{k|k-1}\right)^T \\
\mathcal{K} &= P_{xy} P_\nu^{-1} \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + \mathcal{K}\left(y_k - \hat{y}_{k|k-1}\right) \\
P_{k|k} &= P_{k|K-1} - \mathcal{K} P_\nu \mathcal{K}^T
\end{aligned}$$

Here we have used the subscript notation $k|k-1$ to denote quantities calculated in the prediction stage and $k|k$ to denote quantities calculated in the update stage. $\hat{x}_{k|k}$ is the state estimate, $P_{k|k}$ is the estimation error covariance, $\mathcal{K}$ is the Kalman gain, P_ν is the innovation covariance, and P_{xy} is the cross-covariance between the estimation error and innovation.

It is important to notice that other UKF variants exist. These UKF filters use an augmented state vector that includes the process and measurement noise, and tend to perform very well. However, this UKF variant uses more than $2n+1$ sigma points. Thus, we are limiting the study to two sigma point filters that use $2n+1$ sigma points. The second filter will be introduced in the next section.

3. THE GENERALIZED UNSCENTED TRANSFORM AND THE GUKF

3.1 The Generalized Unscented Transform

The limitations of the UT are highlighted in Ref. 23. A summary includes the following. When comparing the Taylor series expansion of the true mean of the transformation $y = g(x)$, it can be seen that the unscented transform does not contain odd-powered moments (e.g. skewness). This introduces significant approximation errors in situations where the odd-powered moments of the distribution of x are non-zero and the transformation $y = g(x)$ is highly nonlinear. The fourth-order term also fails to capture a part of the true kurtosis even when the heuristic value of $\kappa = 3 - n$ is selected when assuming a Gaussian distribution. The UT performs optimally on Gaussian distributions because the skewness of a Gaussian distribution is zero. However, many other distributions (continuous and discrete) have non-zero skewness that the UT fails to capture. Similar issues as just mentioned can be seen when comparing the Taylor series expansion of the true covariance with the UT expansion of the covariance.

These limitations led to the generalized unscented transform (GenUT).[23] The GenUT aims to accurately capture the mean, covariance matrix, and the diagonal components of both the skewness and kurtosis tensors. It is noted that for an independent random vector, accurately matching the diagonal components of the skewness tensor implies an accurate matching of the entire skewness tensor.[23] This is achieved by selecting sigma points that have the flexibility to either be symmetric when x is symmetrically distributed (e.g. Gaussian) or asymmetric when x is asymmetrically distributed. This is achieved with $2n+1$ sigma points.

Let us now define the notation. We define the vectors $\breve{S} \in \mathbb{R}^n$ and $\breve{K} \in \mathbb{R}^n$ which contain the diagonal components of the skewness tensor and the kurtosis tensor, respectively,

$$\begin{aligned}
\breve{S} &= \left[S_{111}, S_{222}, \ldots, S_{nnn}\right]^T \\
\breve{K} &= \left[K_{1111}, K_{2222}, \ldots, K_{nnnn}\right]^T .
\end{aligned}$$

Let x be a vector, P be a square matrix, and k be some positive integer. We define the element-wise product

(Hadamard product) as $\odot$, such that

$$x^{\odot n} = \underbrace{x \odot x \odot \cdots \odot x}_{\text{n times}}$$

$$P^{\odot -n} = \left(\underbrace{P \odot P \odot \cdots \odot P}_{\text{n times}}\right)^{-1}$$

We also define element-wise division (Hadamard division) as $\oslash$.

The GenUT involves constraining the sigma points and their weights to match the first four moments of a distribution. With a prescribed mean $\hat{x}$, covariance P, skewness tensor diagonal $\breve{S}$, and kurtosis tensor diagonal $\breve{K}$, the sigma points are calculated as

$$\begin{aligned} \mathcal{X}_0 &= \hat{x}\,, & W_0\,, \\ \mathcal{X}_i &= \hat{x} - u_i \sqrt{P_{xx}}_i, & W_i'\,, \\ \mathcal{X}_{i+n} &= \hat{x} - v_i \sqrt{P_{xx}}_i, & W_{i+n}''\,, \end{aligned} \tag{6}$$

for $i = 1, 2, \ldots, n$. The weights are defined as $W' = [W_1, W_2, \ldots, W_n]^T$ and $W'' = [W_{1+n}, W_{2+n}, \ldots, W_{2n}]^T$. The parameter $u > 0$ is chosen by the filter designer or calculated as

$$u = \frac{1}{2}\left(-\sqrt{P}^{\odot -3}\breve{S} + \sqrt{4\sqrt{P}^{\odot -4}\breve{K} - 3\left(\sqrt{P}^{\odot -3}\breve{S}\right)^{\odot 2}}\right) \tag{7}$$

and

$$v = u + \sqrt{P}^{\odot -3}\breve{S}. \tag{8}$$

The weights are then calculated as

$$\begin{aligned} W'' &= 1 \oslash v \oslash (u + v)\,, \\ W' &= W'' \odot v \oslash u\,, \\ W_0 &= 1 - \sum_{i=1}^{2n} W_i\,, \end{aligned} \tag{9}$$

and the total weight vector is $W = \begin{bmatrix} W_0 & W'^T & W''^T \end{bmatrix}^T$.

If information is known about the sigma points beforehand, the GenUT allows for the sigma points to be further constrained. See Ref. 23 for more details.

3.2 The Generalized Unscented Kalman Filter

Extending the GenUT into a recursive filter framework is straightforward. However, we note a few differences when comparing to the UKF:

- The GenUT requires a prescribed skewness and kurtosis in addition to the mean and covariance.
- As a result of the above point, the diagonals of the skewness and kurtosis tensors need to be updated in the filter algorithm.

To achieve the second point, we require the following assumptions and definitions. First, it is assumed that the measurement model is linear in all states, or if nonlinear, can be formed into an m-vector, i.e., $H \in \mathbb{R}^m$. Then, the measurement model used for the update step is the diagonal of the measurement model:

$$\breve{H} = \text{diag}(H)\,.$$

Likewise, we define the following as the diagonals of previously defined terms

$$\begin{aligned}\mathcal{K}_d &= \operatorname{diag}(\mathcal{K})\,,\\ \breve{P} &= \operatorname{diag}(P)\,,\\ \breve{R} &= \operatorname{diag}(R)\,,\\ \breve{S} &= \operatorname{diag}(S)\,,\\ \breve{K} &= \operatorname{diag}(K)\,.\end{aligned}$$

The measurement skewness and measurement kurtosis are defined as

$$\begin{aligned}\breve{S}_m &= \operatorname{diag}(S_m)\,,\\ \breve{K}_m &= \operatorname{diag}(K_m)\,.\end{aligned}$$

Once again, we assume the process and measurement noise are zero mean and uncorrelated. The skewness and kurtosis update equations as

$$\breve{S}_{k|k} = (\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 3} \odot \breve{S}_{k|k-1} - \mathcal{K}_d^{\odot 3} \odot \breve{S}_m\,, \tag{10}$$

$$\breve{K}_{k|k} = (\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 4} \odot \breve{K}_{k|k-1} + 6(\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 2} \odot \breve{P}_{k|k-1} \odot \mathcal{K}_d^{\odot 2} \odot \breve{R} - \mathcal{K}_d^{\odot 4} \odot \breve{K}_m\,, \tag{11}$$

where $\mathbf{1} = \begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}^T \in \mathbb{R}^n$ and $\mathbf{0} = \begin{bmatrix}0 & 0 & \dots & 0\end{bmatrix}^T \in \mathbb{R}^n$. See Appendix A for the derivations of Eqs. (10) and (11).

Extending the GenUT into the UKF structure defines what we call the generalized unscented Kalman filter (GUKF). The GUKF algorithm is as follows:

1. Initialization:

$$\begin{aligned}\hat{x}_{0|0} &= \mathbb{E}[x_0]\\ P_{0|0} &= \mathbb{E}[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T]\\ \breve{S}_{0|0} &= \mathbb{E}[(x_0 - \hat{x}_0)^{\odot 3}]\\ \breve{K}_{0|0} &= \mathbb{E}[(x_0 - \hat{x}_0)^{\odot 4}]\end{aligned}$$

2. Sigma point selection according to Eqs. (6)-(9).

3. The prediction stage:

$$\begin{aligned}\mathcal{X}_{i,k|k-1} &= f\left(\mathcal{X}_{i,k-1}\right)\\ \hat{x}_{k|k-1} &= \sum_{i=0}^{2n} W_i \mathcal{X}_{i,k|k-1}\\ P_{k|k-1} &= \sum_{i=0}^{2n} W_i \left[\mathcal{X}_i - \hat{x}\right]\left[\mathcal{X}_i - \hat{x}\right]^T\\ \mathcal{Y}_{i,k|k-1} &= h\left(\mathcal{X}_{i,k-1}\right)\\ \hat{y}_{k|k-1} &= \sum_{i=0}^{2n} W_i \mathcal{Y}_{i,k|k-1}\end{aligned}$$

4. The update stage:

$$
\begin{aligned}
P_{\nu} &= \sum_{i=0}^{2n} W_i \left(\mathcal{Y}_{i,k|k-1} - \hat{y}_{k|k-1}\right)\left(\mathcal{Y}_{i,k|k-1} - \hat{y}_{k|k-1}\right)^T \\
P_{xy} &= \sum_{i=0}^{2n} W_i \left(\mathcal{X}_{i,k|k-1} - \hat{x}_{k|k-1}\right)\left(\mathcal{Y}_{i,k|k-1} - \hat{y}_{k|k-1}\right)^T \\
\mathcal{K} &= P_{xy} P_{\nu}^{-1} \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + \mathcal{K}\left(y_k - \hat{y}_{k|k-1}\right) \\
P_{k|k} &= P_{k|K-1} - \mathcal{K} P_{\nu} \mathcal{K}^T \\
\\
\mathcal{K}_d &= \mathrm{diag}(\mathcal{K}) \\
\breve{S}_{k|k} &= (\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 3} \odot \breve{S}_{k|k-1} - \mathcal{K}_d^{\odot 3} \odot \breve{S}_m \\
\breve{K}_{k|k} &= (\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 4} \odot \breve{K}_{k|k-1} + 6(\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 2} \odot \breve{P}_{k|k-1} \odot \mathcal{K}_d^{\odot 2} \odot \breve{R} - \mathcal{K}_d^{\odot 4} \odot \breve{K}_m
\end{aligned}
$$

The next section details the simulations and results comparing the UKF and GUKF.

4. SIMULATIONS AND RESULTS

4.1 System Model and Noise Properties

The system model is the Van der Pol oscillator in reverse time. This system is highly nonlinear and exhibits unstable limit cycle behaviour. When the initial state is outside of the limit cycle, the state will diverge. Conversely, when the initial state is inside the limit cycle, it will converge to zero over time due to nonlinear damping. The Van der Pol oscillator can be modelled by the set of first order differential equations

$$
\begin{aligned}
\dot{x}_1 &= -x_2 \,, \\
\dot{x}_2 &= -\mu(1 - x_1^2)x_2 + x_1 \,.
\end{aligned}
\tag{12}
$$

The setup of the system and its properties follows that of Ref. 26. In the following simulations, the system is discretized with a sampling period of 0.1 seconds and the damping parameter is set to $\mu = 0.2$.

4.2 Effect of Noise Distribution

These simulations are based on the premise that the filter designer is assuming that the probability distributions of the noise sequences are Gaussian. As such, the heuristic value of $\kappa = 3 - n$ will be used.[18] The simulations explore what happens when the assumed distribution is not Gaussian, but remains symmetrical. Performance is compared using noise generated from a Gaussian distribution and a uniform distribution. We begin by comparing the performance of the UKF and GUKF in estimating the states when perturbed by Gaussian process and measurement noise.

For each of the case studies in Sec. 4.2, the additive process and measurement noise sequences had the following properties:

- zero mean,
- noise covariances

$$
Q_{\text{actual}} = R_{\text{actual}} \begin{bmatrix} 1 \times 10^{-2} & 0 \\ 0 & 1 \times 10^{-2} \end{bmatrix} ,
$$

- measurement skewness and kurtosis diagonals

$$
\begin{aligned}
\breve{S}_m &= \begin{bmatrix} 0 & 0 \end{bmatrix}^T , \\
\breve{K}_m &= \begin{bmatrix} 3 & 3 \end{bmatrix}^T .
\end{aligned}
$$

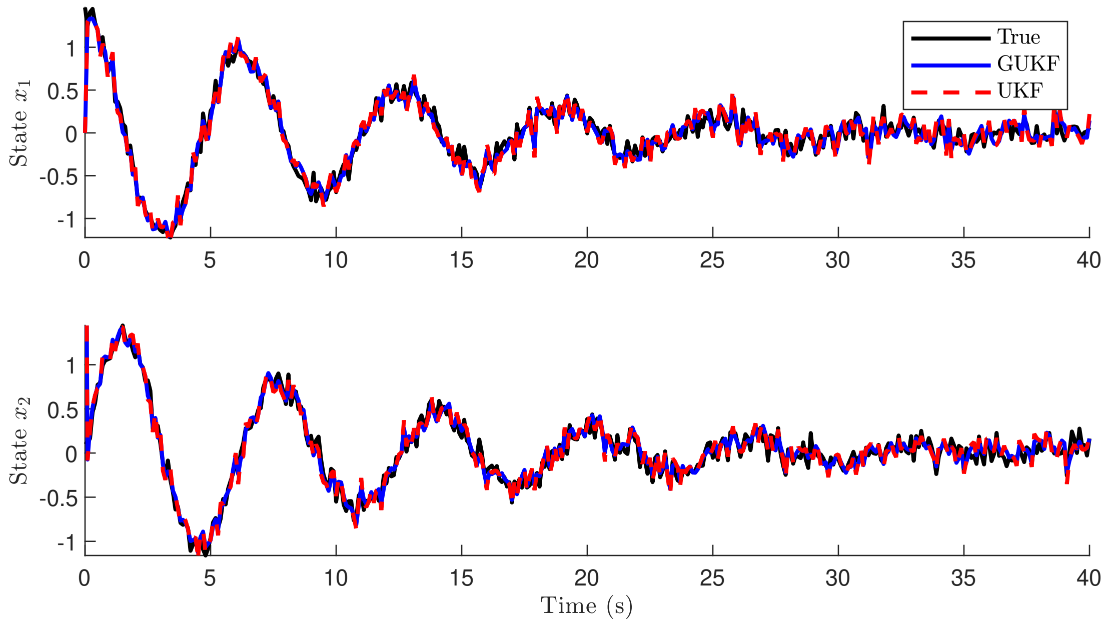

Figure 1. State estimates with Gaussian distributed process and measurement noise.

Additionally, the initial state, state estimate, process noise covariance, and measurement noise covariance are chosen as

$$x_0 = \begin{bmatrix} 1.4 \\ 0 \end{bmatrix}, \hat{x}_{0|0} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}, Q = Q_{\text{actual}}, R = R_{\text{actual}} .$$

The initial state error covariance is

$$P_{0|0} = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} .$$

The choice of initial choices of $P_{0|0}$ is adequate since the initial estimated state is far from the actual state. Also, as analysed by Ref. 26 for the same Van der Pol system in Eq. (5), the UKF is robust to the choices of $P_{0|0}$, Q, and R.

For the GUKF, the estimation error skewness and kurtosis are initialized as

$$\breve{S}_{0|0} = \begin{bmatrix} 0 & 0 \end{bmatrix}^T ,$$
$$\breve{K}_{0|0} = \begin{bmatrix} 3 & 3 \end{bmatrix}^T .$$

4.2.1 Gaussian Distributed Noise

In a theoretical sense, the UKF and GUKF should perform identically when the noise distributions are symmetrical. Thus, it is expected that for Gaussian noise both filters perform the same.

In this study, both the process and measurement noise sequences were sampled from a Gaussian distribution with zero mean and covariance as outlined in Sec. 4.2. This also means that the skewness was approximately zero and the standardized kurtosis was approximately 3. The state estimates and the state errors are plotted in Figures 1 and 2, respectively. The root-mean-squared errors are given in Table 1. The errors are quite similar, but the GUKF outperforms the UKF for the given noise characteristics by 7.21% for the first state (x_1).

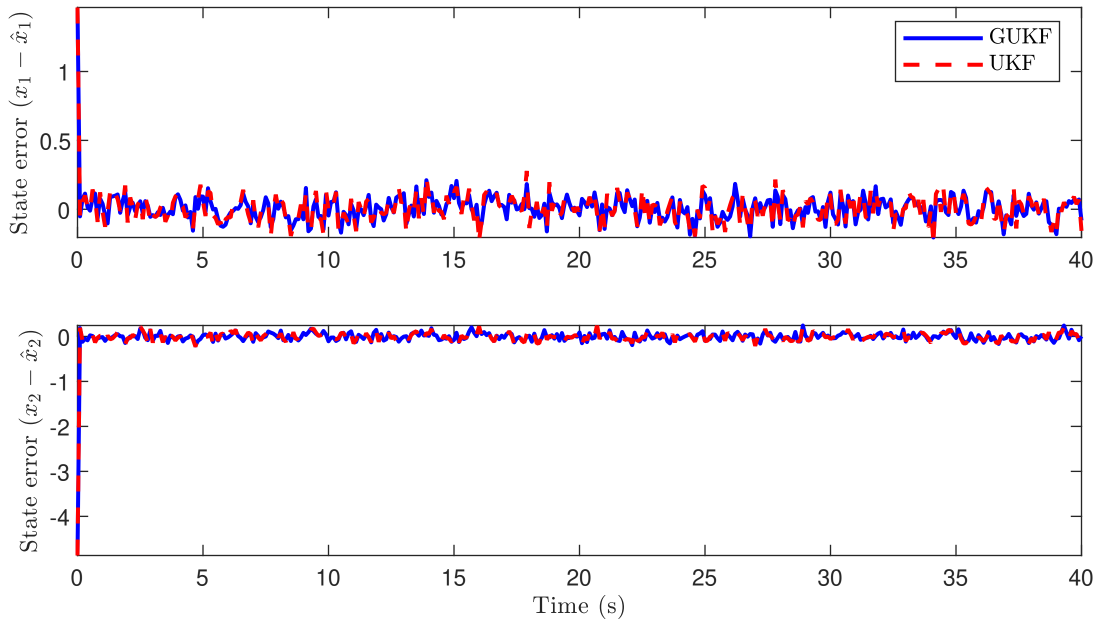

Figure 2. Estimation error for the UKF and GUKF with Gaussian distributed process and measurement noise.

Table 1. RMSE of State Estimates

Distribution	State	UKF	GUKF	% Improvement
Gaussian $\mathcal{N}(\mu, P)$	x_1	0.116	0.107	7.21%
	x_2	0.257	0.255	0.71%
Uniform(a, b)	x_1	0.115	0.109	5.17%
	x_2	0.273	0.271	0.77%

4.2.2 Uniformly Distributed Noise

For this study, the noise sequences were sampled from the distribution Uniform$(-0.175, 0.175)$, which gives zero mean and variance $\sigma^2 = 1 \times 10^{-2}$. The skewness was approximately zero and the standardized kurtosis approximately equal to 1.8.

The state estimates and their errors were similar to the Gaussian noise case. The RMSE errors we given in Table 1. We see that the GUKF shows a 5.17% improvement in estimating the first state. The elements of the covariance matrix and the Kalman gain are plotted in Figures 3 and 4, respectively. The GUKF achieved a lower covariance in both states. Interestingly, the GUKF Kalman gain for the state x_1 is quite different than that for the UKF. This can be interpreted as the GUKF relies more on the prediction than on the measurement innovation for state x_1.

4.3 Results

The results could indicate that the GUKF is an improvement over the UKF. However, the assumptions made about the simulation hinder the performance of the UKF. If the UKF parameter κ is tuned to an optimal value, the UKF and GUKF perform identically. The benefit here is that the GUKF does not require any tuning aside from the initialized statistics.

The results obtained here are quite similar to those found is Ref. 25. The conclusions from that work were that the linear update rule only uses the mean and covariance information to determine the Kalman gain. A Kalman gain that incorporates skewness and kurtosis information could lead to improvements in the GUKF. This Kalman gain could make use of the fact that the GUKF has the capability to predict the skewness and kurtosis at each discrete time step, much like the covariance prediction.

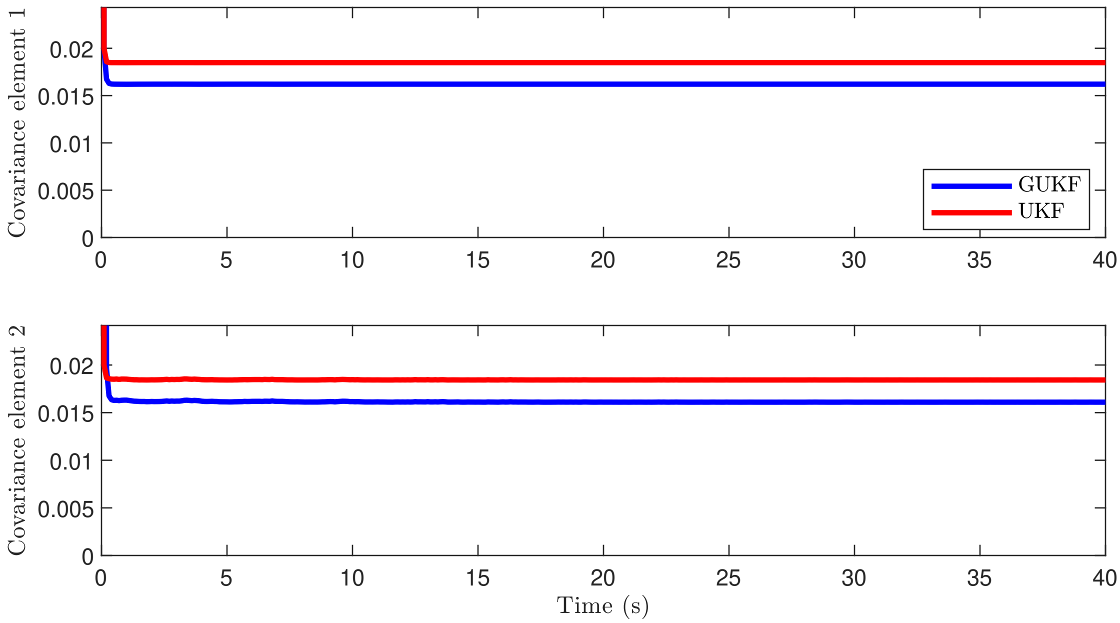

Figure 3. State error covariance of the UKF and GUKF with uniformly distributed process and measurement noise.

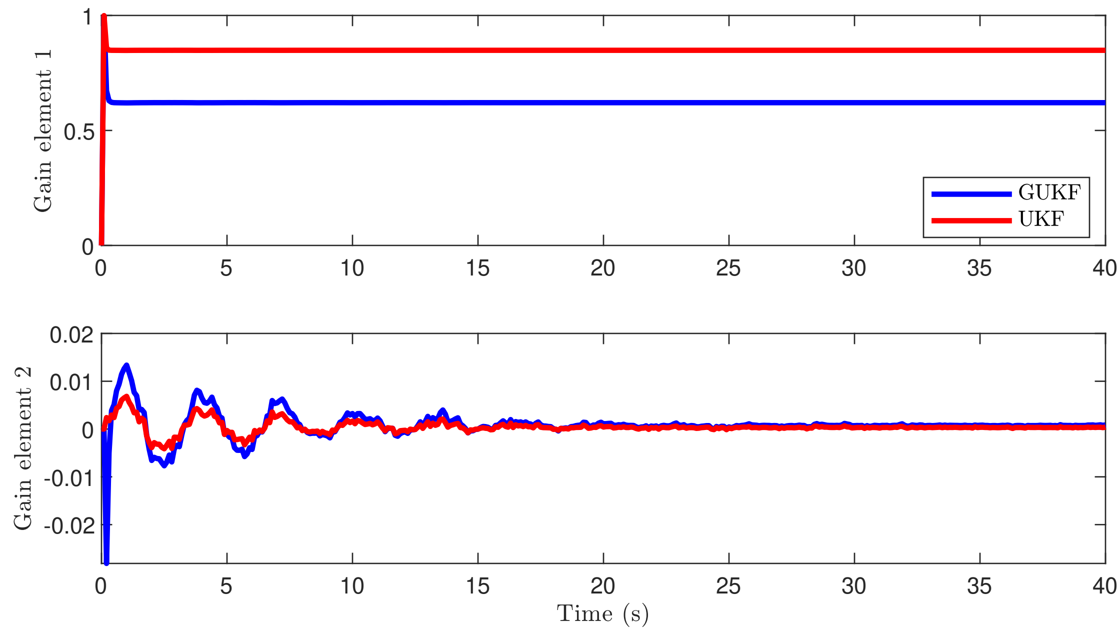

Figure 4. Kalman gain for the UKF and GUKF with uniformly distributed process and measurement noise.

5. CONCLUSION

The studies presented here show the potential of a generalized unscented Kalman filter. The GUKF requires measurement update equations for the diagonals of the skewness and kurtosis tensors. These updated equations and their derivations are provided.

Simulations were performed to compare the performance of the GUKF and UKF for Gaussian and non-Gaussian noise distributions. Improvements were made by the GUKF over the UKF for both cases of noise distributions. However, the UKF estimates can be made identical to the GUKF by properly tuning the parameter κ. The GUKF has the benefit that it does not need to be tuned.

Future work includes extending the GUKF to nonlinear measurements, non-additive noise models, and improving the skewness and kurtosis update through an improved Kalman gain. Additionally, asymmetric distributions should be used to test the GUKF in estimating the statistics of such a distribution. Under these circumstances, the performance of the UKF should deteriorate. The difficulty in using an asymmetric distribution is that they often have a non-zero mean.

APPENDIX A. DERIVATION OF SKEWNESS AND KURTOSIS UPDATE EQUATIONS

It is important that the diagonals of the skewness and kurtosis tensors are accurately updated so that the GenUT can generate an adequate set of sigma points for the GUKF. The following subsections outline the derivations for the update equations of the diagonals of the skewness and kurtosis tensors. Recall that is was assumed that the measurement model be linear in all states or it can be formulated into an m-vector, i.e., $H \in \mathbb{R}^m$. We represented the diagonal of the measurement model as $\breve{H} = \text{diag}(H)$.

To derive the skewness and kurtosis update equations, we require the equivalent diagonal form of the state estimation error. The state the estimation error is

$$\tilde{x}_{k|k} = x_k - \hat{x}_{k|k}.$$

Recalling that the state update equation is

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + \mathcal{K}(y_k - H\hat{x}_{k|k-1}),$$

where H is the linear measurement model and $y_k = Hx_k + v_k$, we can write the estimation error as

$$\begin{aligned}\tilde{x}_{k|k} &= x_k - \hat{x}_{k|k-1} - \mathcal{K}(Hx_k + v_k - H\hat{x}_{k|k-1}) \\ &= (I - \mathcal{K}H)\tilde{x}_{k|k-1} - \mathcal{K}v_k\,.\end{aligned}$$

Here, I is the identity matrix. We can extend this to only consider the diagonals of the required matrices and use element-wise products to get

$$\tilde{x}_{k|k} = (\mathbf{1} - \mathcal{K}_d \odot \breve{H}) \odot \tilde{x}_{k|k-1} - \mathcal{K}_d \odot v_k\,, \tag{13}$$

as needed.

A.1 Update Equation for the Diagonal of the Skewness Tensor

The skewness update equation is derived by expanding the third central moment of the estimation error as

$$\begin{aligned}\breve{S}_{k|k} &= \mathbb{E}\left[\tilde{x}_{k|k}^{\odot 3}\right] \\ &= \mathbb{E}\left[(x_k - \hat{x}_{k|k})^{\odot 3}\right] \\ &= \mathbb{E}\left[\left((\mathbf{1} - \mathcal{K}_d \odot \breve{H}) \odot \tilde{x}_{k|k-1} - \mathcal{K}_d \odot v_k\right)^{\odot 3}\right] \\ &= (\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 3} \odot \mathbb{E}\left[\tilde{x}_{k|k-1}^{\odot 3}\right] \\ &\quad - 3(\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 2} \odot \mathbb{E}\left[\tilde{x}_{k|k-1}^{\odot 2}\right] \odot \mathcal{K}_d \odot \mathbb{E}\left[v_k\right] \\ &\quad + 3(\mathbf{1} - \mathcal{K}_d \odot \breve{H}) \odot \mathbb{E}\left[\tilde{x}_{k|k-1}\right] \odot \mathcal{K}_d^{\odot 2} \odot \mathbb{E}\left[v_k^{\odot 2}\right] \\ &\quad - \mathcal{K}_d^{\odot 3} \odot \mathbb{E}\left[v_k^{\odot 3}\right]\,,\end{aligned} \tag{14}$$

where Eq. (13) was used on the third line. From the last line of Eq. (14) we can recognize the following expectations

$$\begin{aligned}
\mathbb{E}\left[\tilde{x}^{\odot 2}_{k|k-1}\right] &= \breve{P}_{k|k-1}\,,\\
\mathbb{E}\left[\tilde{x}^{\odot 3}_{k|k-1}\right] &= \breve{S}_{k|k-1}\,,\\
\mathbb{E}\left[v_k\right] &= \mathbf{0}\,,\\
\mathbb{E}\left[v^{\odot 2}_{k}\right] &= \breve{R}\,,\\
\mathbb{E}\left[v^{\odot 3}_{k}\right] &= \breve{S}_m\,.
\end{aligned}$$

Also, if the filter is assumed to be unbiased then we have

$$\mathbb{E}\left[\tilde{x}_{k|k-1}\right] = \mathbf{0}.$$

Using these relationships, we can reduce Eq. (14) to the update equation

$$\breve{\mathbf{S}}_{k|k} = (\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 3} \odot \breve{S}_{k|k-1} - \mathcal{K}_d^{\odot 3} \odot \breve{S}_m\,. \tag{15}$$

A.2 Update Equation for the Diagonal of the Kurtosis Tensor

The kurtosis update equation is derived via the fourth central moment about the estimation error,

$$\begin{aligned}
\breve{K}_{k|k} &= \mathbb{E}\left[\tilde{x}^{\odot 4}_{k|k}\right] \\
&= \mathbb{E}\left[(x_k - \hat{x}_{k|k})^{\odot 4}\right] \\
&= \mathbb{E}\left[\left((\mathbf{1} - \mathcal{K}_d \odot \breve{H}) \odot \tilde{x}_{k|k-1} - \mathcal{K}_d \odot v_k\right)^{\odot 4}\right] \\
&= (\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 4} \odot \mathbb{E}\left[\tilde{x}^{\odot 4}_{k|k-1}\right] \\
&\quad - 4(\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 3} \odot \mathbb{E}\left[\tilde{x}^{\odot 3}_{k|k-1}\right] \odot \mathcal{K}_d \odot \mathbb{E}\left[v_k\right] \\
&\quad + 6(\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 2} \odot \mathbb{E}\left[\tilde{x}^{\odot 2}_{k|k-1}\right] \odot \mathcal{K}_d^{\odot 2} \odot \mathbb{E}\left[v_k^{\odot 2}\right] \\
&\quad - 4(\mathbf{1} - \mathcal{K}_d \odot \breve{H}) \odot \mathbb{E}\left[\tilde{x}_{k|k-1}\right] \odot \mathcal{K}_d^{\odot 3} \odot \mathbb{E}\left[v_k^{\odot 3}\right] \\
&\quad - \mathcal{K}_d^{\odot 4} \odot \mathbb{E}\left[v_k^{\odot 4}\right]
\end{aligned} \tag{16}$$

Again, recognizing the expectations and including

$$\begin{aligned}
\mathbb{E}\left[\tilde{x}^{\odot 4}_{k|k-1}\right] &= \breve{K}_{k|k-1}\,,\\
\mathbb{E}\left[v^{\odot 4}_{k}\right] &= \breve{K}_m\,,
\end{aligned}$$

Eq. (16) reduces to the kurtosis update equation

$$\begin{aligned}
\breve{K}_{k|k} = (\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 4} &\odot \breve{K}_{k|k-1} \\
&+ 6(\mathbf{1} - \mathcal{K}_d \odot \breve{H})^{\odot 2} \odot \breve{P}_{k|k-1} \odot \mathcal{K}_d^{\odot 2} \odot \breve{R} \\
&- 4(\mathbf{1} - \mathcal{K}_d \odot \breve{H}) \odot \sqrt{\breve{P}_{k|k-1}} \odot \mathcal{K}_d^{\odot 3} \odot \breve{S}_m \\
&- \mathcal{K}_d^{\odot 4} \odot \breve{K}_m\,.
\end{aligned} \tag{17}$$

REFERENCES

[1] Kushner, H. and Budhiraja, A., "A nonlinear filtering algorithm based on an approximation of the conditional distribution," *IEEE Transactions on Automatic Control* **45**, 580–585 (Mar. 2000).

[2] Mcreynolds, S. R., "Multidimensional Hermite-Gaussian quadrature formulae and their application to nonlinear estimation," in [*Symposium on Nonlinear Estimation Theory and Its Applications*], (Jan. 1975). NTRS Author Affiliations: California Institute of Technology, Jet Propulsion Laboratory, Pasadena NTRS Meeting Information: Symposium on Nonlinear Estimation Theory and Its Applications; 1975-09-15 to 1975-09-17; undefined NTRS Document ID: 19760060700 NTRS Research Center: Legacy CDMS (CDMS).

[3] Genz, A. and Keister, B. D., "Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight," *Journal of Computational and Applied Mathematics* **71**, 299–309 (July 1996).

[4] W., S. H. and Stubberud, A. R., "Non-linear filtering by approximation of the a posteriori density," *International Journal of Control* **8**, 33–51 (July 1968).

[5] Srinivasan, K., "State estimation by orthogonal expansion of probability distributions," *IEEE Transactions on Automatic Control* **15**, 3–10 (Feb. 1970).

[6] Challa, S., Bar-Shalom, Y., and Krishnamurthy, V., "Nonlinear filtering via generalized Edgeworth series and Gauss-Hermite quadrature," *IEEE Transactions on Signal Processing* **48**, 1816–1820 (June 2000).

[7] Bucy, R. S. and Senne, K. D., "Digital synthesis of non-linear filters," *Automatica* **7**, 287–298 (May 1971).

[8] Evensen, G., "The Ensemble Kalman Filter: theoretical formulation and practical implementation," *Ocean Dynamics* **53**, 343–367 (Nov. 2003).

[9] Gordon, N. J., Salmond, D. J., and Smith, A. F. M., "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F (Radar and Signal Processing)* **140**, 107–113 (Apr. 1993).

[10] Carpenter, J., Clifford, P., and Fearnhead, P., "Improved particle filter for nonlinear problems," *IEE Proceedings - Radar, Sonar and Navigation* **146**, 2–7 (Feb. 1999).

[11] Jazwinski, A. H., [*Stochastic Processes and Filtering Theory*], Courier Corporation (Jan. 2007).

[12] Gelb, A., ed., [*Applied Optimal Estimation*], MIT Press (1974).

[13] Athans, M., Wishner, R., and Bertolini, A., "Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy measurements," *IEEE Transactions on Automatic Control* **13**, 504–514 (Oct. 1968).

[14] Schei, T. S., "A finite-difference method for linearization in nonlinear estimation algorithms," *Automatica* **33**, 2053–2058 (Nov. 1997).

[15] Ito, K. and Xiong, K., "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control* **45**, 910–927 (May 2000).

[16] Arasaratnam, I., Haykin, S., and Elliott, R. J., "Discrete-Time Nonlinear Filtering Algorithms Using Gauss–Hermite Quadrature," *Proceedings of the IEEE* **95**, 953–977 (May 2007).

[17] Arasaratnam, I. and Haykin, S., "Square-Root Quadrature Kalman Filtering," *IEEE Transactions on Signal Processing* **56**, 2589–2593 (June 2008).

[18] Julier, S. J. and Uhlmann, J. K., "New extension of the Kalman filter to nonlinear systems," in [*Signal Processing, Sensor Fusion, and Target Recognition VI*], **3068**, 182–193, SPIE (July 1997).

[19] Julier, S. and Uhlmann, J., "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE* **92**, 401–422 (Mar. 2004).

[20] Arasaratnam, I. and Haykin, S., "Cubature Kalman Filters," *IEEE Transactions on Automatic Control* **54**, 1254–1269 (June 2009).

[21] Arasaratnam, I., *Cubature Kalman Filtering Theory & Applications*, PhD thesis (Apr. 2009).

[22] Arasaratnam, I., Haykin, S., and Hurd, T. R., "Cubature Kalman Filtering for Continuous-Discrete Systems: Theory and Simulations," *IEEE Transactions on Signal Processing* **58**, 4977–4993 (Oct. 2010).

[23] Ebeigbe, D., Berry, T., Norton, M. M., Whalen, A. J., Simon, D., Sauer, T., and Schiff, S. J., "A generalized unscented transformation for probability distributions," *ArXiv* (2021).

[24] Julier, S., Uhlmann, J., and Durrant-Whyte, H., "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control* **45**, 477–482 (Mar. 2000).

[25] Julier, S. J., "Skewed approach to filtering," in [*Signal and Data Processing of Small Targets 1998*], **3373**, 271–282, SPIE (Sept. 1998).

[26] Kandepu, R., Foss, B., and Imsland, L., "Applying the unscented kalman filter for nonlinear state estimation," *Journal of process control* **18**(7-8), 753–768 (2008).