Unscented Filtering and Nonlinear Estimation

SIMON J. JULIER, MEMBER, IEEE, AND JEFFREY K. UHLMANN, MEMBER, IEEE

Invited Paper

The extended Kalman filter (EKF) is probably the most widely used estimation algorithm for nonlinear systems. However, more than 35 years of experience in the estimation community has shown that is difficult to implement, difficult to tune, and only reliable for systems that are almost linear on the time scale of the updates. Many of these difficulties arise from its use of linearization. To overcome this limitation, the unscented transformation (UT) was developed as a method to propagate mean and covariance information through nonlinear transformations. It is more accurate, easier to implement, and uses the same order of calculations as linearization. This paper reviews the motivation, development, use, and implications of the UT.

Keywords—*Estimation, Kalman filtering, nonlinear systems, target tracking.*

I. INTRODUCTION

This paper considers the problem of applying the Kalman filter (KF) to nonlinear systems. Estimation in nonlinear systems is extremely important because almost all practical systems—from target tracking [1] to vehicle navigation, from chemical process plant control [2] to dialysis machines—involve nonlinearities of one kind or another. Accurately estimating the state of such systems is extremely important for fault detection and control applications. However, estimation in nonlinear systems is extremely difficult. The optimal (Bayesian) solution to the problem requires the propagation of the description of the full probability density function (pdf) [3]. This solution is extremely general and incorporates aspects such as multimodality, asymmetries, discontinuities. However, because the form of the pdf is not restricted, it cannot, in general, be described using a finite number of parameters. Therefore, any practical estimator must use an approximation of some kind. Many different types of approximations have been developed; unfortunately, most are either computationally unmanageable or require special assumptions about the form of the process and observation models that cannot be satisfied in practice. For these and other reasons, the KF remains the most widely used estimation algorithm.

The KF only utilizes the first two moments of the state (mean and covariance) in its update rule. Although this is a relatively simple state representation, it offers a number of important practical benefits.

1) The mean and covariance of an unknown distribution requires the maintenance of only a small and constant amount of information, but that information is sufficient to support most kinds of operational activities (e.g., defining a validation gate for a search region for a target). Thus, it is a successful compromise between computational complexity and representational flexibility. By contrast, the complete characterization of an evolving error distribution requires the maintenance of an unbounded number of parameters. Even if it were possible to maintain complete pdf information, that information may not be operationally useful (e.g., because the exploitation of the information is itself an intractable problem).
2) The mean and covariance (or its square root) are linearly transformable quantities. For example, if an error distribution has mean $\bar{\mathbf{x}}$ and covariance $\mathbf{\Sigma}_x$, the mean and covariance of the distribution after it has undergone the linear transformation $\mathbf{T}$ is simply $\mathbf{T}\bar{\mathbf{x}}$ and $\mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T$. In other words, mean and covariance estimates can be maintained effectively when subjected to linear and quasilinear transformations. Similar results do not hold for other nonzero moments of a distribution.
3) Sets of mean and covariance estimates can be used to characterize additional features of distribution, e.g., significant modes. Multimodal tracking methods based on the maintenance of multiple mean and covariance estimates include multiple-hypothesis tracking [4], sum-of-Gaussian filters [5], and Rao–Blackwellized particle filters [6].

Manuscript received March 14, 2003; revised November 27, 2003.

S. J. Julier is with IDAK Industries, Jefferson City, MO 65109 USA (e-mail: sjulier@idak.com).

J. K. Uhlmann is with the Department of Computer Engineering and Computer Science, University of Missouri–Columbia, Columbia, MO 65211 USA (e-mail: uhlmannj@missouri.edu).

Digital Object Identifier 10.1109/JPROC.2003.823141

0018-9219/04$20.00 © 2004 IEEE

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

The most common application of the KF to nonlinear systems is in the form of the extended KF (EKF) [7], [8]. Exploiting the assumption that all transformations are quasi-linear, the EKF simply linearizes all nonlinear transformations and substitutes Jacobian matrices for the linear transformations in the KF equations. Although the EKF maintains the elegant and computationally efficient recursive update form of the KF, it suffers a number of serious limitations.

1) Linearized transformations are only reliable if the error propagation can be well approximated by a linear function. If this condition does not hold, the linearized approximation can be extremely poor. At best, this undermines the performance of the filter. At worst, it causes its estimates to diverge altogether. However, determining the validity of this assumption is extremely difficult because it depends on the transformation, the current state estimate, and the magnitude of the covariance. This problem is well documented in many applications such as the estimation of ballistic parameters of missiles [1], [9]–[12] and computer vision [13]. In Section II-C we illustrate its impact on the near-ubiquitous nonlinear transformation from polar to Cartesian coordinates.
2) Linearization can be applied only if the Jacobian matrix exists. However, this is not always the case. Some systems contain discontinuities (for example, the process model might be jump-linear [14], in which the parameters can change abruptly, or the sensor might return highly quantized sensor measurements [15]), others have singularities (for example, perspective projection equations [16]), and in others the states themselves are inherently discrete (e.g., a rule-based system for predicting the evasive behavior of a piloted aircraft [17]).
3) Calculating Jacobian matrices can be a very difficult and error-prone process. The Jacobian equations frequently produce many pages of dense algebra that must be converted to code (e.g., see the Appendix to [18]). This introduces numerous opportunities for human coding errors that may undermine the performance of the final system in a manner that cannot be easily identified and debugged—especially given the fact that it is difficult to know what quality of performance to expect. Regardless of whether the obscure code associated with a linearized transformation is or is not correct, it presents a serious problem for subsequent users who must validate it for use in any high integrity system.

The unscented transformation (UT) was developed to address the deficiencies of linearization by providing a more direct and explicit mechanism for transforming mean and covariance information. In this paper we describe the general UT mechanism along with a variety of special formulations that can be tailored to the specific requirements of different nonlinear filtering and control applications. The structure of this paper is as follows. Section II reviews the relationship between the KF and the EKF for nonlinear systems and motivates the development of the UT. An overview of the UT is provided in Section III and some of its properties are discussed. Section IV discusses the algorithm in more detail and some practical implementation considerations are considered in Section V. Section VI describes how the transformation can be applied within the KF's recursive structure. Section VII considers the implications of the UT. Summary and conclusions are given in Section VIII. The paper also includes a comprehensive series of Appendixes which provide detailed analyses of the performance properties of the UT and further extensions to the algorithm.

II. Problem Statement

A. Applying the KF to Nonlinear Systems

Many textbooks derive the KF as an application of Bayes' rule under the assumption that all estimates have independent, Gaussian-distributed errors. This has led to a common misconception that the KF can only be strictly applied under Gaussianity assumptions. However, Kalman's original derivation did not apply Bayes' rule and does not require the exploitation of any specific error distribution informaton beyond the mean and covariance [19].

To understand the the limitations of the EKF, it is necessary to consider the KF recursion equations. Suppose that the estimate at time step $n-1$ is described by the mean $\boldsymbol{\mu}_{n-1}$ and covariance $\mathbf{K}_{n-1}$. It is assumed that this is *consistent* in the sense that [7]

$$\mathbf{K}_n - \mathrm{E}\left[\mathbf{e}_n \mathbf{e}_n^T\right] \geq \mathbf{0} \tag{1}$$

where $\mathbf{e}_n = \mathbf{x} - \boldsymbol{\mu}_n$ is the estimation error.[1]

The KF consists of two steps: prediction followed by update. In the prediction step, the filter propagates the estimate from a previous time step $n-1$ to the current time step n. The prediction is given by

$$\begin{aligned}\hat{\boldsymbol{\mu}}_n &= \mathrm{E}[\mathbf{f}[\boldsymbol{\mu}_{n-1}, \mathbf{u}_n, \mathbf{b}_n, n]] \\ \hat{\mathbf{K}}_n &= \mathrm{E}\left[\hat{\mathbf{e}}_n \hat{\mathbf{e}}_n^T\right]\end{aligned}$$

The update (or measurement update) can be derived as the linear minimum mean-squared error estimator [20]. Given that the mean is to be updated by the linear rule

$$\begin{aligned}\boldsymbol{\mu}_n &= \hat{\boldsymbol{\mu}}_n + \mathbf{W}_n \boldsymbol{\nu}_n \\ \boldsymbol{\nu}_n &= \mathbf{y}_n - \hat{\mathbf{y}}_n \\ \hat{\mathbf{y}}_n &= \mathrm{E}[\mathbf{g}[\boldsymbol{\mu}_n, \mathbf{v}_n]]\end{aligned}$$

the weight (gain) matrix $\mathbf{W}_n$ is chosen to minimize the trace of the updated covariance $\mathbf{K}_n$. Its value is calculated from

$$\mathbf{W}_n = \hat{\boldsymbol{K}}_n^{xy} \hat{\boldsymbol{S}}_n^{-1}$$

where $\hat{\boldsymbol{K}}_n^{xy}$ is the cross covariance between the error in $\hat{\boldsymbol{\mu}}_n$ and the error in $\hat{\mathbf{y}}_n$ and $\hat{\boldsymbol{S}}_n$ is the covariance of $\boldsymbol{\nu}_n$.

[1] A *conservative* estimate replaces the inequality with a strictly greater than relationship.

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

Using this weight, the updated covariance is

$$\mathbf{K}_n = \hat{\mathbf{K}}_n - \mathbf{W}_n \hat{\mathbf{S}}_n \mathbf{W}_n^T.$$

Therefore, the KF update equations can be applied if several sets of expectations can be calculated. These are the predicted state $(\hat{\boldsymbol{\mu}}_n, \hat{\mathbf{K}}_n)$, the predicted observation $(\hat{\mathbf{y}}_n, \hat{\mathbf{S}}_n)$, and the cross covariance between the prediction and the observation $(\hat{\mathbf{K}}_n^{xy})$. When all of the system equations are linear, direct substitution into the above equations gives the familiar linear KF equations. When the system is nonlinear, methods for approximating these quantities must be used. Therefore, the problem of applying the KF to a nonlinear system becomes one of applying nonlinear transformations to mean and covariance estimates.

B. Propagating Means and Covariances Through Nonlinear Transformations

Consider the following problem. A random variable $\mathbf{x}$ has mean $\bar{\mathbf{x}}$ and covariance $\boldsymbol{\Sigma}_x$. A second random variable, $\mathbf{z}$, is related to $\mathbf{x}$ through to the nonlinear transformation

$$\mathbf{z} = \mathbf{h}[\mathbf{x}]. \tag{2}$$

The problem is to calculate a consistent estimate of $\mathbf{z}$ with mean $\bar{\mathbf{z}}$ and covariance $\boldsymbol{\Sigma}_z$. Both the prediction and update steps of a KF can be written in this form.[2]

Taking the multidimensional Taylor series expansion

$$\begin{aligned}\mathbf{z} &= \mathbf{h}[\bar{\mathbf{x}} + \mathbf{e}] \\ &= \mathbf{h}[\bar{\mathbf{x}}] + \mathbf{D_e h} + \frac{\mathbf{D_e^2 h}}{2!} + \frac{\mathbf{D_e^3 h}}{3!} + \frac{\mathbf{D_e^4 h}}{4!} + \cdots \end{aligned} \tag{3}$$

where the $\mathbf{D_e h}$ operator evaluates the total differential of $\mathbf{h}[\cdot]$ when perturbed around a nominal value $\bar{\mathbf{x}}$ by $\mathbf{e}$. The ith term in the Taylor series for $\mathbf{h}[\cdot]$ is

$$\frac{\mathbf{D_e^i h}}{i!} = \frac{1}{i!}\left(\sum_{j=1}^{N_x} e_j \frac{\partial}{\partial x_j}\right)^i \mathbf{h}[\mathbf{x}]\Bigg|_{\mathbf{x}=\bar{\mathbf{x}}} \tag{4}$$

where e_j is the jth component of $\mathbf{e}$. Therefore, the ith term in the series is an ith-order polynomial in the coefficients of $\mathbf{e}$, whose coefficients are given by derivatives of $\mathbf{h}[\cdot]$.

In Appendix I we derive the full expression for the mean and covariance using this series. Linearization assumes that all second and higher order terms in the Taylor series expansion are negligible, i.e.,

$$\bar{\mathbf{z}} \approx \mathbf{h}[\bar{\mathbf{x}}] + \mathbf{D_e h}.$$

Taking outer products and expectations, and by exploiting the assumption that the estimation error is approximately zero-mean, the mean and covariance are

$$\bar{\mathbf{z}} = \mathbf{h}[\bar{\mathbf{x}}] \tag{5}$$

$$\boldsymbol{\Sigma}_z = \mathbf{A}(\mathbf{x})\boldsymbol{\Sigma}_x \mathbf{A}^T(\mathbf{x}) \tag{6}$$

[2]The prediction corresponds to the case when $\mathbf{x} = \boldsymbol{\mu}_{n-1}$, $\mathbf{z} = \hat{\boldsymbol{\mu}}_{n-1}$, and $\mathbf{h}[\cdot] = \mathbf{f}[\cdot]$. The update step corresponds to the case when $\mathbf{x} = \hat{\boldsymbol{\mu}}_{n-1}$, $\mathbf{z} = \hat{\mathbf{y}}_n$, and $\mathbf{h}[\cdot] = \mathbf{g}[\cdot]$.

where $\mathbf{A}(\mathbf{x})$ is the Jacobian of $\mathbf{h}[\cdot]$ and the fact that $\mathrm{E}[\mathbf{D_e h}(\mathbf{D_e h})^T] = \mathbf{A}(\mathbf{x})\boldsymbol{\Sigma}_x \mathbf{A}^T(\mathbf{x})$ has been used.

However, the full Taylor series expansion of this function, given in Appendix I, shows that these quantities contain higher order terms that are a function of the statistics of $\mathbf{e}$ and higher derivatives of the nonlinear transformation. In some situations, these terms have negligible effects. In other situations, however, they can significantly degrade estimator performance. One dramatic and practically important example is the transformation from polar to Cartesian coordinates [12].

C. Polar to Cartesian Coordinate Transformations

One of the most important and ubiquitous transformations is the conversion from polar to Cartesian coordinates. This transformation forms the foundation for the observation models of many sensors, from radar to laser range finders. A sensor returns polar information (r, θ) in its local coordinate frame that has to be converted into an (x, y) position estimate of the target position in some global Cartesian coordinate frame

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}. \tag{7}$$

Problems arise when the bearing error is significant. As an example, a range-optimized sonar sensor can provide fairly good measurements of range (2-cm standard deviation) but extremely poor measurements of bearing (standard deviation of 15°) [21]. The effects of the large error variances on the nonlinearly transformed estimate are shown in Fig. 1, which shows the results for a target whose true position is (0, 1). Fig. 1(a) plots several hundred (x, y) samples. These were derived by taking the actual (r, θ) value of the target location, adding Gaussian zero-mean noise terms to each component, and then converting to (x, y) using (7). As can be seen, the points lie on a "banana"-shaped arc. The range error causes the points to lie in a band, and the bearing error causes this region to be stretched around the circumference of a circle. As a result, the mean does not lie at $(0, 1)$ but is actually located closer to the origin. This is confirmed in Fig. 1(b), which compares the mean and covariance of the converted coordinates using Monte Carlo sampling and linearization. The figure plots the 1σ contours calculated by each method. The 1σ contour is the locus of points $\{\mathbf{y} : (\mathbf{y} - \bar{\mathbf{y}})\mathbf{P}_y^{-1}(\mathbf{y} - \bar{\mathbf{y}}) = 1\}$ and is a graphical representation of the size and orientation of $\boldsymbol{\Sigma}_y$.

Compared to the "true" result, the linearized estimate is biased and inconsistent. This is most evident in the y direction. The linearized mean is at 1.0 m but the true mean is at 96.7 cm. Because it is a bias that arises from the transformation process itself, the same error with the same sign will be committed every time a coordinate transformation takes place. Even if there were no bias, the transformation would still be inconsistent because it underestimates the variance in the y component.[3]

[3]It could be argued that these errors arise because the measurement errors are unreasonably large. However, Lerro [12] demonstrated that, in radar tracking applications, the transformations can become inconsistent when the standard deviation of the bearing measurement is less than a degree.

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

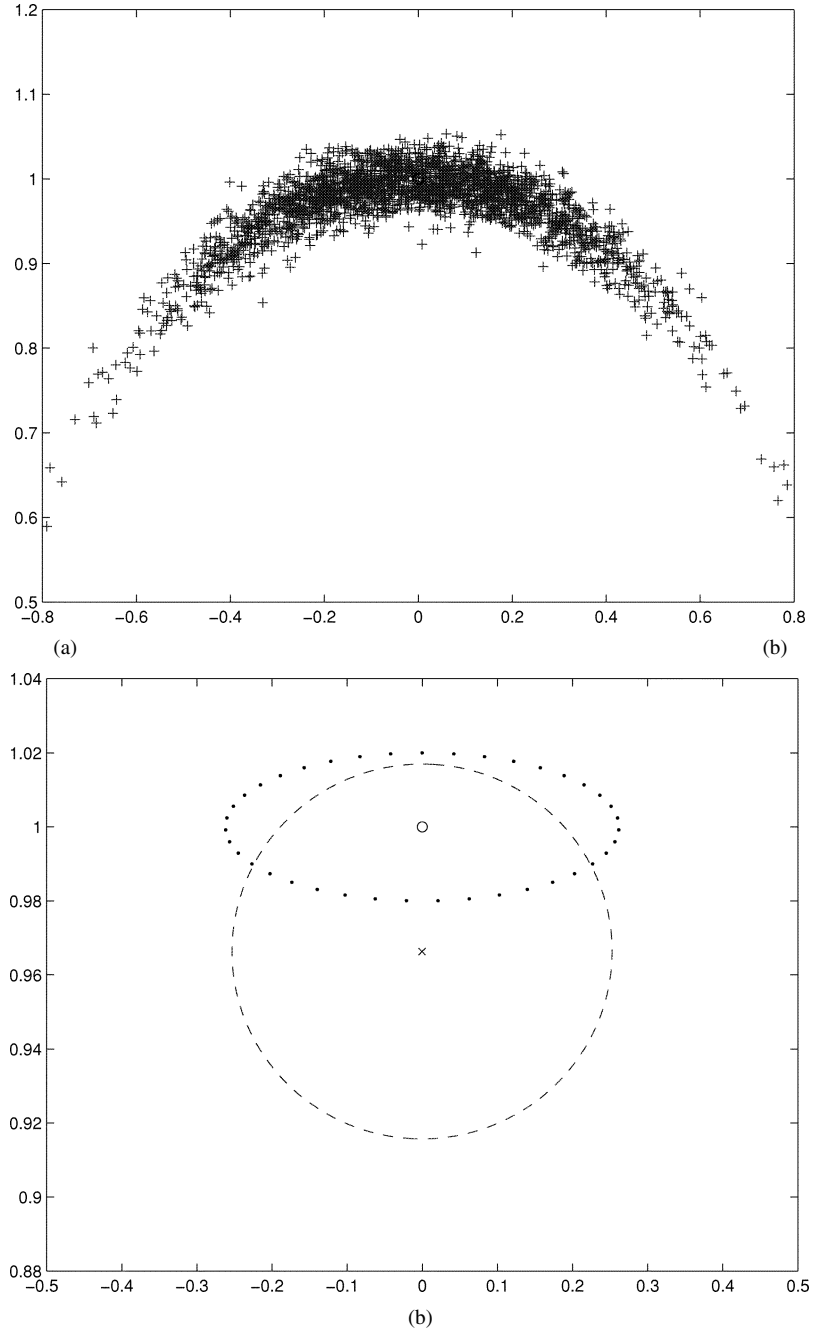

Fig. 1. The true nonlinear transformation and the statistics calculated by Monte Carlo analysis and linearization. Note that the scaling in the x and y axes are different. (a) Monte Carlo samples from the transformation and the mean calculated through linearization. (b) Results from linearization. The true mean is at $\times$ and the uncertainty ellipse is solid. Linearization calculates the mean at $\circ$ and the uncertainty ellipse is dashed.

There are several strategies that could be used to address this problem. The most common approach is to apply linearization and "tune" the observation covariance by padding it with a sufficiently large positive definite matrix that the transformed estimate becomes consistent.[4] However, this approach may unnecessarily increase the assumed uncertainty in some directions in the state space, and it does nothing to address the problem of the bias. A second approach would be to perform a detailed analysis of the transformations and derive precise closed-form solutions for the transformed mean and covariance under specific distribution assumptions. In the case of the polar-to-Cartesian transformation of an assumed Gaussian distributed observation estimate, such closed-form solutions do exist [12], [23]. However, exact solutions can only be derived in special cases with highly specific assumptions. A slightly more general approach is to note that linearization errors arise from an implicit truncation of the Taylor series description of the true transformed estimate. Therefore, maintaining higher order terms may lead to better results. One of the first to attempt this was

[4]This process is sometimes euphemistically referred to as "injecting stabilizing noise" [22].

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

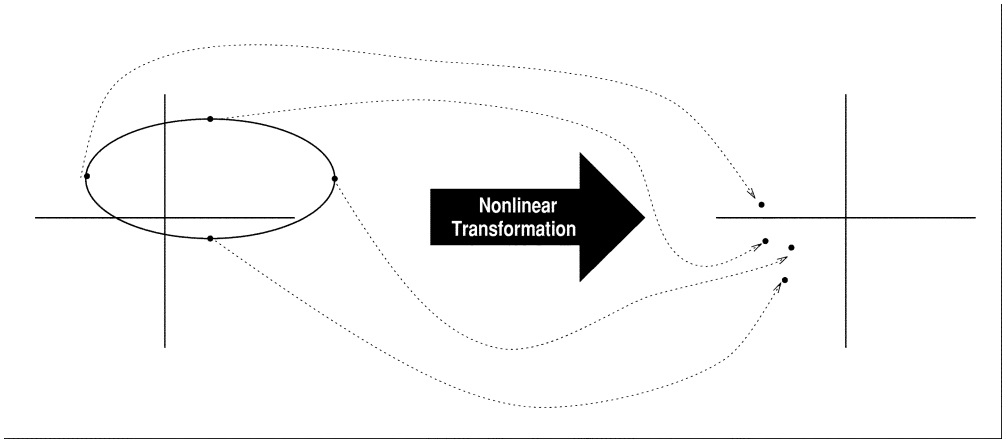

Fig. 2. The principle of the UT.

Athans [9], who developed a second-order Gaussian filter. This filter assumes that the model is piecewise quadratic and truncates the Taylor series expansion after its second term. However, to implement this filter, the Hessian (tensor of second-order derivatives) must be derived. Typically, deriving the Hessian is even more difficult than deriving a Jacobian, especially for a sophisticated, high-fidelity system model. Furthermore, it is not clear under what conditions the use of the Hessian will yield improved estimates when the strict assumption of Gaussianity is violated.

In summary, the KF can be applied to nonlinear systems if a consistent set of predicted quantities can be calculated. These quantities are derived by projecting a prior estimate through a nonlinear transformation. Linearization, as applied in the EKF, is widely recognized to be inadequate, but the alternatives incur substantial costs in terms of derivation and computational complexity. Therefore, there is a strong need for a method that is provably more accurate than linearization but does not incur the implementation nor computational costs of other higher order filtering schemes. The UT was developed to meet these needs.

III. The Unscented Transformation

A. Basic Idea

The UT is founded on the intuition that *it is easier to approximate a probability distribution than it is to approximate an arbitrary nonlinear function or transformation* [24]. The approach is illustrated in Fig. 2—a set of points (*sigma points*) are chosen so that their mean and covariance are $\bar{\mathbf{x}}$ and $\mathbf{\Sigma}_x$. The nonlinear function is applied to each point, in turn, to yield a cloud of transformed points. The statistics of the transformed points can then be calculated to form an estimate of the nonlinearly transformed mean and covariance.

Although this method bares a superficial resemblance to particle filters, there are several fundamental differences. First, the sigma points are *not* drawn at random; they are deterministically chosen so that they exhibit certain specific properties (e.g., have a given mean and covariance). As a result, high-order information about the distribution can be captured with a fixed, small number of points. The second difference is that sigma points can be weighted in ways that are inconsistent with the distribution interpretation of sample points in a particle filter. For example, the weights on the points do not have to lie in the range $[0, 1]$.

A set of sigma points $\mathcal{S}$ consists of $p+1$ vectors and their associated weights $\mathcal{S} = \{i = 0, 1, \ldots, p : \mathbf{x}^{(i)}, W^{(i)}\}$. The weights $W^{(i)}$ can be positive or negative but, to provide an unbiased estimate, they must obey the condition

$$\sum_{i=0}^{p} W^{(i)} = 1. \tag{8}$$

Given these points, $\bar{\mathbf{y}}$ and $\mathbf{\Sigma}_y$ are calculated as follows.

1) Instantiate each point through the function to yield the set of transformed sigma points

$$\mathbf{z}^{(i)} = \mathbf{h}\left[\mathbf{x}^{(i)}\right].$$

2) The mean is given by the weighted average of the transformed points

$$\bar{\mathbf{z}} = \sum_{i=0}^{p} W^{(i)} \mathbf{z}^{(i)}. \tag{9}$$

3) The covariance is the weighted outer product of the transformed points

$$\mathbf{\Sigma}_z = \sum_{i=0}^{p} W^{(i)} \left\{\mathbf{z}^{(i)} - \bar{\mathbf{z}}\right\} \left\{\mathbf{z}^{(i)} - \bar{\mathbf{z}}\right\}^T. \tag{10}$$

The statistics of any other function can be calculated in a similar manner.

One set of points that satisfies the above conditions consists of a symmetric set of $2N_x$ points that lie on the $\sqrt{N_x}$th covariance contour [24]

$$\begin{aligned} \mathbf{x}^{(i)} &= \bar{\mathbf{x}} + (\sqrt{N_x \mathbf{\Sigma}_x})_i \\ W^{(i)} &= 1/2N_x \\ \mathbf{x}^{(i+N_x)} &= \bar{\mathbf{x}} - (\sqrt{N_x \mathbf{\Sigma}_x})_i \\ W_{i+N_x} &= 1/2N_x \end{aligned} \tag{11}$$

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

where $(\sqrt{N_x \boldsymbol{\Sigma}_x})_i$ is the ith row or column[5] of the matrix square root of $N_x \boldsymbol{\Sigma}_x$ (the original covariance matrix multiplied by the number of dimensions), and $W^{(i)}$ is the weight associated with the ith point.

Despite its apparent simplicity, the UT has a number of important properties.

1) Because the algorithm works with a finite number of sigma points, it naturally lends itself to being used in a "black box" filtering library. Given a model (with appropriately defined inputs and outputs), a standard routine can be used to calculate the predicted quantities as necessary for any given transformation.
2) The computational cost of the algorithm is the same order of magnitude as the EKF. The most expensive operations are calculating the matrix square root and the outer products required to compute the covariance of the projected sigma points. However, both operations are $O(N_x^3)$, which is the same as evaluating the $N_x \times N_x$ matrix multiplications needed to calculate the EKF predicted covariance.[6] This contrasts with methods such as Gauss–Hermite quadrature [26] for which the required number of points scales geometrically with the number of dimensions.
3) Any set of sigma points that encodes the mean and covariance correctly, including the set in (11), calculates the projected mean and covariance correctly to the second order (see Appendixes I and II). Therefore, the estimate implicitly includes the second-order "bias correction" term of the truncated second-order filter, but without the need to calculate any derivatives. Therefore, the UT is *not* the same as using a central difference scheme to calculate the Jacobian.[7]
4) The algorithm can be used with discontinuous transformations. Sigma points can straddle a discontinuity and, thus, can approximate the effect of a discontinuity on the transformed estimate. This is discussed in more detail in Section VII.

The improved accuracy of the UT can be demonstrated with the polar-to-Cartesian transformation problem.

B. The Demonstration Revisited

Using sigma points determined by (11), the performance of the UT is shown in Fig. 3. Fig. 3(a) plots the set of transformed sigma points $\mathbf{z}^{(i)}$. The original set of points $\mathbf{x}^{(i)}$ were symmetrically distributed about the origin and the nonlinear transformation has changed the distribution into a triangle with a point at the center. The mean and covariance of the UT, compared to the true and linearized values, is shown in Fig. 3(b). The UT mean is extremely close to that of the true transformed distribution. This reflects the effect of the second-order bias correction term that is implicitly and automatically incorporated into the mean via the UT. However, the UT covariance estimate underestimates the true covariance of the actual transformed distribution. This is because the set of points described above are only accurate to the second order. Therefore, although the mean is predicted much more accurately, the UT predicted covariance is of the same order of accuracy as linearization. The transformed estimate could be made consistent by adding stabilising noise to increase $\boldsymbol{\Sigma}_x$; however, the UT framework provides more direct mechanisms that can greatly improve the accuracy of the estimate. These are considered next.

IV. Exploiting Higher Order Information

The example in the previous section illustrates that the UT, using the sigma point selection algorithm in (11), has significant implementation and accuracy advantages over linearization. However, because the UT offers enough flexibility to allow information beyond mean and covariance to be incorporated into the set of sigma points, it is possible to pick a set that exploits any additional known information about the error distribution associated with an estimate.

A. Extending the Symmetric Set

Suppose a set of points is constructed to have a given mean and covariance, e.g., according to (11). If another point equal to the given mean were added to the set, then the mean of the set would be unaffected, but the remaining points would have to be scaled to maintain the given covariance. The scaled result is a different sigma set, with different higher moments, but with the same mean and covariance. As will be shown, weighting this newly added point provides a parameter for controlling some aspects of the higher moments of the distribution of sigma points without affecting the mean and covariance. By convention, let $W^{(0)}$ be the weight on the mean point, which is indexed as the zeroth point. Including this point and adjusting the weights so that the normality, mean, and covariance constraints are preserved, the new point distribution becomes[8]

$$\begin{aligned}
\mathbf{x}^{(0)} &= \bar{\mathbf{x}} \\
W^{(0)} &= W^{(0)} \\
\mathbf{x}^{(i)} &= \bar{\mathbf{x}} + \left(\sqrt{\frac{N_x}{1-W^{(0)}}\boldsymbol{\Sigma}_x}\right)_i \\
W^{(i)} &= \frac{1-W^{(0)}}{2N_x} \\
\mathbf{x}^{(i+N_x)} &= \bar{\mathbf{x}} - \left(\sqrt{\frac{N_x}{1-W^{(0)}}\boldsymbol{\Sigma}_x}\right)_i \\
W^{(i+N_x)} &= \frac{1-W^{(0)}}{2N_x}.
\end{aligned} \tag{12}$$

[5]If the matrix square root $\mathbf{A}$ of $\mathbf{P}$ is of the form $\mathbf{P} = \mathbf{A}^T\mathbf{A}$, then the sigma points are formed from the *rows* of $\mathbf{A}$. However, if the matrix square root is of the form $\mathbf{P} = \mathbf{A}\mathbf{A}^T$, the *columns* of $\mathbf{A}$ are used.

[6]The matrix square root should be calculated using numerically efficient and stable methods such as the Cholesky decomposition [25].

[7]Lefebvre recently argued for an alternative interpretation of the UT as being an example of least-squares regression [27]. Although regression analysis can be used to justify UT's accuracy benefits over linearization, it does not provide a prescriptive framework, e.g., for deriving the extensions to higher order information described in Section IV.

[8]In the scalar case, this distribution is the same as the perturbation result described by Holtzmann [28]. However, the generalization of Holtzmann's method to multiple dimensions is accurate only under the assumption that the errors in each dimension are independent of one another.

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

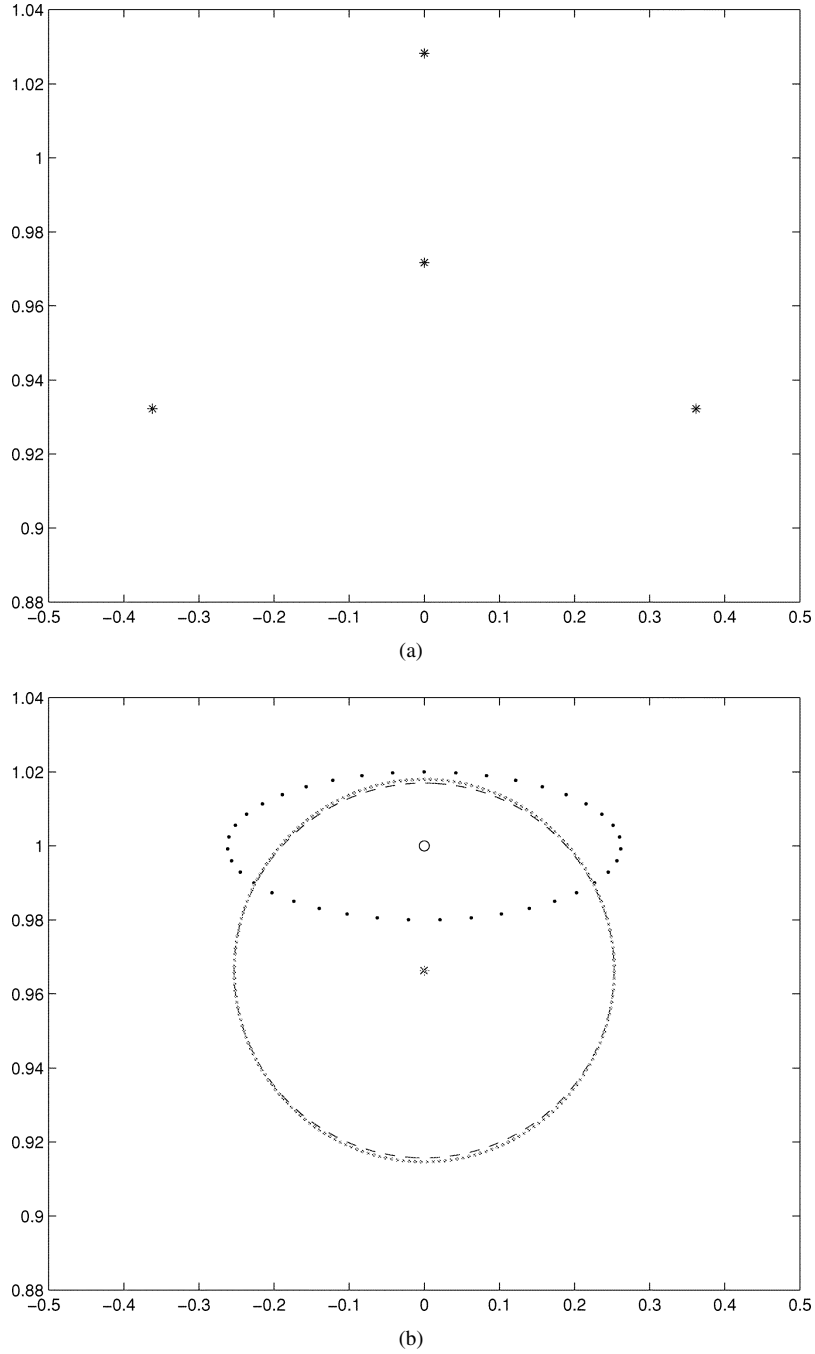

Fig. 3. The mean and standard deviation ellipses for the true statistics, those calculated through linearization and those calculated by the UT. (a) The location of the sigma points which have undergone the nonlinear transformation. (b) The results of the UT compared to linearization and the results calculated through Monte Carlo analysis. The true mean is at $\times$ and the uncertainty ellipse is dotted. The UT mean is at $+$ and is the solid ellipse. The linearized mean is at $\circ$ and its ellipse is also dotted.

The value of $W^{(0)}$ controls how the other positions will be repositioned. If $W^{(0)} \geq 0$, the points tend to move further from the origin. If $W^{(0)} \leq 0$ (a valid assumption because the UT points are *not* a pdf and so do not require nonnegativity constraints), the points tend to be closer to the origin.

The impact of this extra point on the moments of the distribution are analyzed in detail in Appendix II. However, a more intuitive demonstration of the effect can be obtained from the example of the polar-to-Cartesian transformation.

Given the fact that the assumed prior distribution (sensor error) is Gaussian in the local polar coordinate frame of the sensor, and using the analysis in Appendix II, $W^{(0)} = 1/3$ can be justified because it guarantees that some of the fourth-order moments are the same as in the true Gaussian case. Fig. 4(a) shows the points generated with the augmented set. The additional point lies at the mean calculated by the EKF $(0, 1)$. The effect is shown in more detail in Fig. 4(b). These dramatic improvements come from the

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

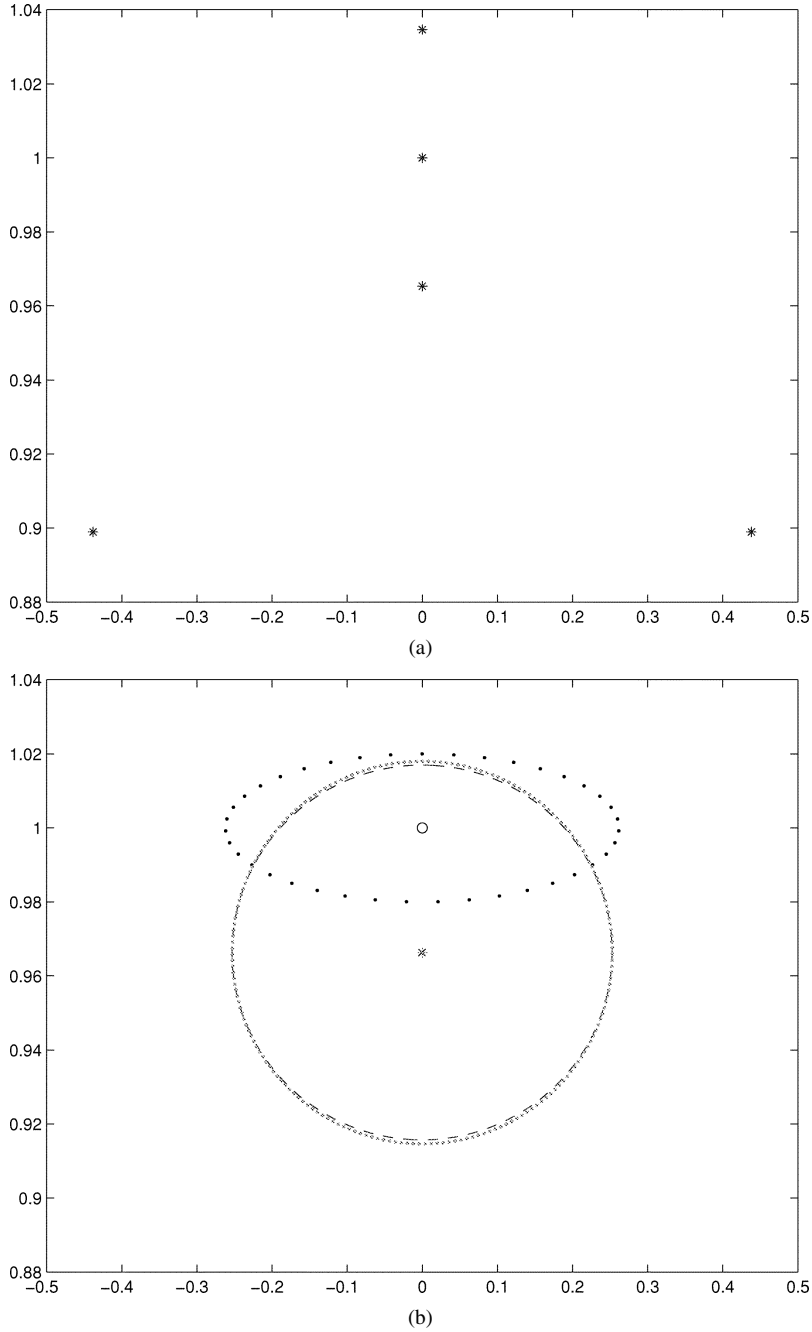

Fig. 4. The mean and standard deviation ellipses for the true statistics, those calculated through linearization and those calculated by the UT. (a) The location of the sigma points which have undergone the nonlinear transformation. (b) The results of the UT compared to linearization and the results calculated through Monte Carlo analysis. The true mean is at $\times$ and the uncertainty ellipse is dotted. The UT mean is at $+$ and is the solid ellipse. The linearized mean is at $\circ$ and its ellipse is also dotted.

fact that the extra point exploits control of the higher order moments.[9]

B. General Sigma Point Selection Framework

The extension in the previous section shows that knowledge of higher order information can be partially incorporated into the sigma point set. This concept can be generalized so that the UT can be used to propagate *any* higher order information about the moments.

Because no practical filter can maintain the full distribution of the state, a simpler distribution of the form $p(\mathbf{x})$ may be *heuristically assumed* at time step k. If $p(\mathbf{x})$ is chosen to be computationally or analytically tractable to work with, and if it captures the critical salient features of the true distribution, then it can be used as the basis for a good approximate solution to the nonlinear transformation problem. One possi-

[9]Recently, Nørgaard developed the DD2 filtering algorithm, which is based on Stirling's interpolation formula [29]. He has proved that it can yield more accurate estimates than the UT with the symmetric set. However, this algorithm has not been generalized to higher order information.

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

bility explored by Kushner is to assume that all distributions are Gaussian [30], i.e., $p(\mathbf{x})$ is a Gaussian distributed random variable with mean $\bar{\mathbf{x}}$ and covariance $\mathbf{\Sigma}_x$.

Although Gaussianity is not always a good assumption to make in many cases, it provides a good example to demonstrate how a set of sigma points can be used to capture, or *match*, different properties of a given distribution. This matching can be written in terms of a set of nonlinear constraints of the form

$$\mathbf{c}[\mathcal{S}, p_x(\mathbf{x})] = \mathbf{0}.$$

For example, the symmetric set of points given in (11) matches the mean and covariance of a Gaussian, and by virtue of its symmetry, it also matches the third moment (skew $=$ 0) as well. These conditions on $\mathbf{c}[\cdot,\cdot]$ can be written as

$$\mathbf{c}_1[\mathcal{S}, p_x(\mathbf{x})] = \sum_{i=0}^{p} W^{(i)}\mathbf{x}^{(i)} - \bar{\mathbf{x}}$$

$$\mathbf{c}_2[\mathcal{S}, p_x(\mathbf{x})] = \sum_{i=0}^{p} W^{(i)}\left(\mathbf{x}^{(i)} - \bar{\mathbf{x}}\right)\left(\mathbf{x}^{(i)} - \bar{\mathbf{x}}\right)^T - \mathbf{\Sigma}_x$$

$$\mathbf{c}_3[\mathcal{S}, p_x(\mathbf{x})] = \sum_{i=N_x}^{p} W^{(i)}\left(\mathbf{x}^{(i)} - \bar{\mathbf{x}}\right)^3.$$

The constraints are not always sufficient to uniquely determine the sigma point set. Therefore, the set can be refined by introducing a cost function $c[\mathcal{S}, p(\mathbf{x})]$ which penalizes undesirable characteristics. For example, the symmetric set given in (12) contains some degrees of freedom: the matrix square root and the value $W^{(0)}$. The analysis in Appendix II shows that $W^{(0)}$ affects the fourth and higher moments of the sigma point set. Although the fourth-order moments cannot be matched precisely, $W^{(0)}$ can be chosen to minimize the errors.

In summary, the general sigma point selection algorithm is

$$\min_{\mathcal{S}} c[\mathcal{S}, p_x(\mathbf{x})] \quad \text{subject to } \mathbf{c}[\mathcal{S}, p_x(\mathbf{x})] = \mathbf{0}. \tag{13}$$

Two possible uses of this approach are illustrated in Appendixes III and IV. Appendix III shows how the approach can be used to generate a sigma point set that contains the minimal number of sigma points needed to capture mean and covariance ($N_x + 1$ points). Appendix IV generates a set of $2N_x^2 + 1$ points that matches the first four moments of a Gaussian exactly. This set cannot precisely catch the sixth and higher order moments, but the points are chosen to minimize these errors.

Lerner recently analyzed the problem of stochastic numerical integration using *exact mononomials* [31]. Assuming the distribution is symmetric, he gives a general selection rule which, for precision 3, gives the set of points in (12) and, for precision 5, gives the fourth-order set of points given in Appendix IV. His method also provides exact solutions for arbitrary higher order dimensions. Tenne has considered the problem of capturing higher order moments without the assumption of symmetry [32]. He has developed a sigma point selection algorithm that captures the first eight moments of a symmetric one-dimensional distribution using only five points.

This section has discussed how the UT can be extended to incorporate more information about the distribution when this information is available. However, it is frequently the case that only the first two moments of the distribution are known, and imposing incorrect assumptions about these higher order terms might significantly degrade estimator performance. With the UT, it is possible to refine the point set to minimize the effects of the assumptions made about higher order terms.

V. Mitigating Unmodeled Effects

In general, the Taylor series expansion for a set of transformed sigma points contains an infinite number of terms. Although it is possible to match the series expansion for a given distribution up to a given order, the accuracy will be determined by the higher order terms. In most practical contexts there is no information available about these higher terms, so analytically any set of sigma points that matches the available information is as good as any other. However, there are other pragmatic considerations that may give preference to some sigma sets over others.

Suppose that only mean and covariance information is available for some N_x-dimensional state estimate. This information can be captured in a simplex set of N_x+1 sigma points (see Appendix III), which is a significantly smaller set—and, thus, computationally less burdensome—than the $2N_x$ set introduced in (11). Given that nothing is known about the true distribution associated with the estimate beyond the second central moment, it may seem that there is no reason to prefer the symmetric $2N_x$ set over the $N_x + 1$ simplex set. However, consider the following cases for the skew (third central moment) of the unknown distribution.

1) If the skew of the true distribution is zero, then the symmetric set of $2N_x$ sigma points will be more accurate. This is because the odd moments of a symmetric distribution are zero while the skew of a simplex set of points is not generally zero.
2) If the skew of the true distribution is nonzero, then the skew of the simplex set may fortuitously align with that of the true distribution and, thus, be more accurate than the symmetric set. On the other hand, the skew of the simplex set may be in the opposite direction of the true skew, thus causing it to produce much poorer results.

From the above considerations, it is not possible to determine which of the two sigma sets will be more accurate on average, but it *is* possible to conclude that the skewed set is likely to produce the poorest results in the worst case. This can be demonstrated by considering the effect of rotating a sigma point set. When a rotation is applied, it affects the pattern of higher order moments and, thus, can have a significant impact on performance.

These effects can be illustrated with the polar-to-Cartesian transformation which was first discussed in Section II-C.

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

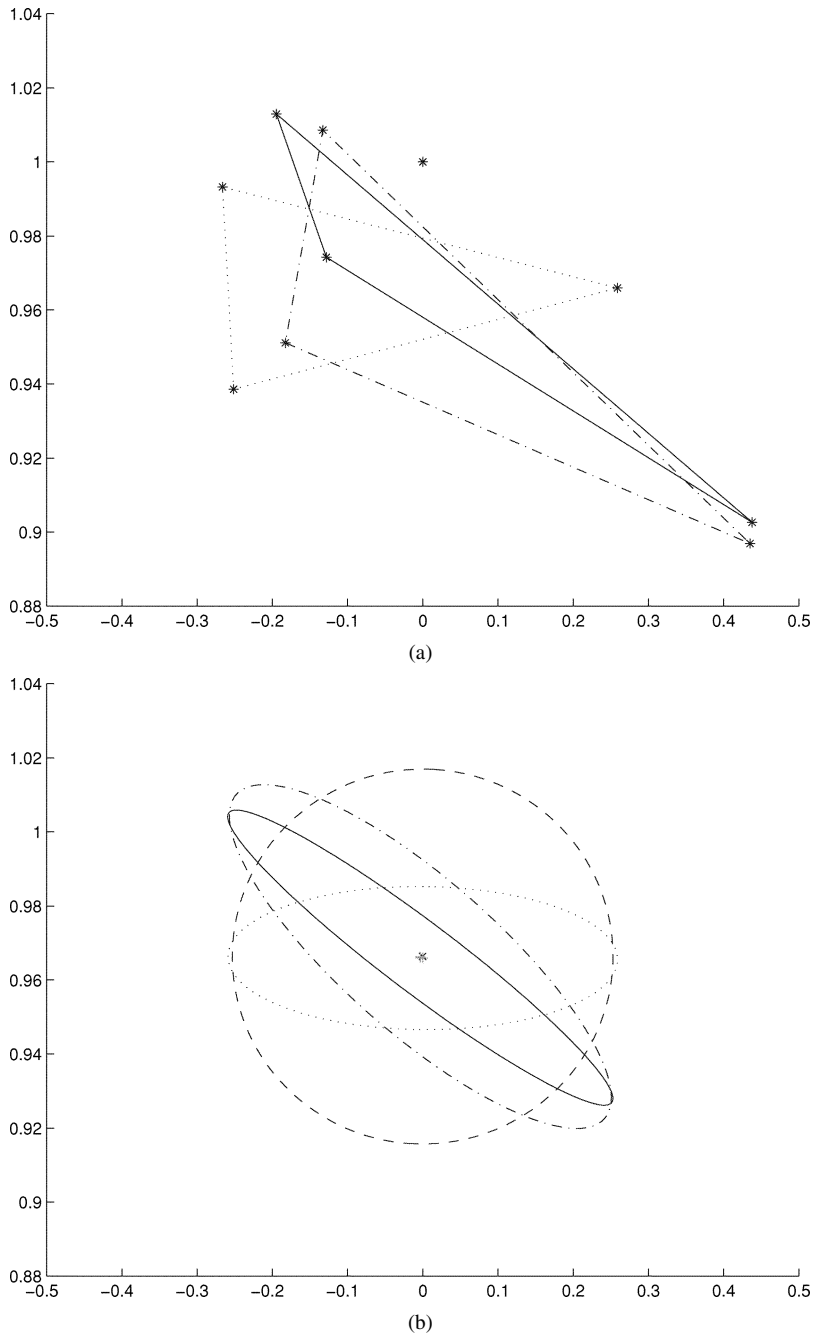

Fig. 5. The effect of rotation on the simplex sigma points. (a) Simplex sets after different rotations. Three sets of simplex sigma points have been rotated by $\theta = 0^\circ$ (solid), $\theta = 120^\circ$ (dot-dashed), and $\theta = 240^\circ$ (dotted), respectively. The zeroth point is unaffected by the rotation and lies at (0, 1) in all cases. (b) The mean and covariance calculated by each set. The dashed plot is the true transformed estimate obtained from Monte Carlo integration assuming a Gaussian prior. (a) Simplex sets after different rotations. (b) Mean and covariance plots.

Suppose that a set of simplex sigma points have been selected using $W^{(0)} = 0$. As explained in Appendix III, these points lie in a triangle. If the rotation matrix $\mathbf{R}$

$$\mathbf{R} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

is applied to each point in the set, the result is equivalent to rotating the triangle. This does not affect the first two moments, but it does affect the third and higher moments. The impact is shown in Fig. 5(a), which superimposes the plots of the sets of transformed sigma points $(\mathbf{z}^{(i)})$ for three different rotations. The impact of the rotations is illustrated in Fig. 5(b), which shows that the mean does not vary but the covariance can change significantly. This is partially explained by the analysis in Appendix I, which shows that the covariance is more sensitive to the higher order behavior of the series.

Thus, even without exact knowledge about the higher terms in the series expansion for the true but unknown distribution, it is still possible (and often necessary) to minimize large-magnitude errors resulting from, e.g., significant

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

but erroneous skew terms. If the first two moments are to be preserved, one mechanism for partially parameterizing the net effect of the higher moments is to reformulate the problem so that it can be written in terms of a different nonlinear transformation

$$\begin{aligned}\tilde{\mathbf{z}} &= \tilde{\mathbf{h}}[\boldsymbol{\mu}_n, \alpha, \mu] \\ &= \frac{\mathbf{h}[\bar{\mathbf{x}} + \alpha\mathbf{e}] - \mathbf{h}[\bar{\mathbf{x}}]}{\mu} + \mathbf{h}[\bar{\mathbf{x}}] \end{aligned} \tag{14}$$

where α is a scaling factor (which scales the "spread" of the distribution) and μ is a normalizing constant whose value is determined below. The reason for considering this form arises from its Taylor series expansion. From (3)

$$\tilde{\mathbf{z}} = \mathbf{h}[\bar{\mathbf{x}}] + \frac{\alpha}{\mu}\mathbf{D}_{\mathbf{e}}\mathbf{h} + \frac{\alpha^2}{\mu}\frac{\mathbf{D}_{\mathbf{e}}^2\mathbf{h}}{2!} + \frac{\alpha^3}{\mu}\frac{\mathbf{D}_{\mathbf{e}}^3\mathbf{h}}{3!} + \frac{\alpha^4}{\mu}\frac{\mathbf{D}_{\mathbf{e}}^4\mathbf{h}}{4!} + \cdots. \tag{15}$$

In other words, the series resembles that for $\mathbf{h}[\mathbf{x}]$ but the kth term is scaled by α^k/μ. If α has a sufficiently small value, the higher order terms have a negligible impact.

Assuming that $\mathrm{E}[\mathbf{e}] = \mathbf{0}$ and setting $\mu = \alpha^2$, it can be shown that [33]

$$\bar{\mathbf{z}} \approx \mathrm{E}[\tilde{\mathbf{z}}] \tag{16}$$

$$\boldsymbol{\Sigma}_z \approx \alpha^2 \mathrm{E}[(\tilde{\mathbf{z}} - \mathrm{E}[\tilde{\mathbf{z}}])(\tilde{\mathbf{z}} - \mathrm{E}[\tilde{\mathbf{z}}])^T] \tag{17}$$

where the results are approximate because the two are correct up to the second order. It should be noted that the mean includes the second-order bias correction term and so this procedure is not equivalent to using a central difference scheme to approximate Jacobians. (Rather than modifying the function itself, the same effect can be achieved by applying a simple postprocessing step to the sigma point selection algorithm [33]. This is discussed in detail in Appendix V).

One side effect of using the modified function is that it eliminates all higher order terms in the Taylor series expansion. The effects are shown in Fig. 6(a), and are very similar to the use of a symmetric sigma set discussed above. When higher order information is known about the distribution, it can be incorporated to improve the estimate. In Appendix VI we show that these terms can be accounted for by adding a term of the form $\{\mathbf{y}^{(0)} - \bar{\mathbf{y}}\}\{\mathbf{y}^{(0)} - \bar{\mathbf{y}}\}^T$ with a weight β to the covariance term

$$\mathbf{y}^{(i)} = \mathbf{h}\left[\mathbf{x}^{(i)}\right] \tag{18}$$

$$\bar{\mathbf{y}}' = \sum_{i=0}^{p} W^{(i)}\mathbf{y}^{(i)} \tag{19}$$

$$\boldsymbol{\Sigma}_z^* = \sum_{i=0}^{p} W^{(i)}\left\{\mathbf{y}^{(i)} - \bar{\mathbf{y}}\right\}\left\{\mathbf{y}^{(i)} - \bar{\mathbf{y}}\right\}^T \tag{20}$$

$$+ (1 + \beta - \alpha^2)\left\{\mathbf{y}^{(0)} - \bar{\mathbf{y}}\right\}\left\{\mathbf{y}^{(0)} - \bar{\mathbf{y}}\right\}^T. \tag{21}$$

The value of β is a function of the kurtosis. For a Gaussian, it is three, and the effects are shown in Fig. 6(b).

The previous two sections have considered how the UT can be applied to "one shot" transformations. We now describe how the UT can be embedded in the KF's recursive prediction and update structure.

VI. Applying the UT to Recursive Estimation

A. Application to the Basic Filtering Steps

To recap the requirements described in Section II, the KF consists of the following steps.

1) Predict the new state of the system $\hat{\boldsymbol{\mu}}_n$ and its associated covariance $\hat{\mathbf{K}}_n$. This prediction must take into account the effects of process noise.
2) Predict the expected observation $\hat{\mathbf{y}}_n$ and the innovation covariance $\hat{\boldsymbol{S}}_n$. This prediction should include the effects of observation noise.
3) Predict the cross covariance matrix $\hat{\boldsymbol{K}}_n^{xy}$.

These steps can be formulated by slightly restructuring the state vector and the process and observation models. The most general formulation augments the state vector with the process and noise terms to give an augmented $N_a = N_x + N_d + N_v$-dimensional vector

$$\bar{\mathbf{x}}_{n-1}^a = \begin{bmatrix} \bar{\mathbf{x}}_{n-1} \\ \mathbf{d}_k \\ \mathbf{v}_k \end{bmatrix}. \tag{22}$$

The process and observation models are rewritten as a function of $\bar{\mathbf{x}}_n^a$

$$\begin{aligned}\bar{\mathbf{x}}_n^a &= \mathbf{f}^a\left[\bar{\mathbf{x}}_{n-1}^a, \mathbf{u}_n\right] \\ \mathbf{y}_n &= \mathbf{g}^a\left[\bar{\mathbf{x}}_n^a, \mathbf{u}_n\right]\end{aligned} \tag{23}$$

and the UT uses sigma points that are computed from[10]

$$\boldsymbol{\mu}_{a,n} = \begin{pmatrix} \boldsymbol{\mu}_n \\ \mathbf{0}_{N_d \times 1} \\ \mathbf{0}_{N_v \times 1} \end{pmatrix} \quad \text{and}$$

$$\mathbf{K}_{a,n} = \begin{bmatrix} \mathbf{K}_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{d,n} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{v,n} \end{bmatrix}. \tag{24}$$

Although this method requires the use of additional sigma points, it incorporates the noises into the predicted state with

[10]If correlations exist among the noise terms, (24) can be generalized to draw the sigma points from the covariance matrix

$$\mathbf{K}_{a,n} = \begin{bmatrix} \mathbf{K}_n & \boldsymbol{\Sigma}_n^{xd} & \boldsymbol{\Sigma}_n^{xv} \\ \boldsymbol{\Sigma}_n^{dx} & \boldsymbol{\Sigma}_{d,n} & \boldsymbol{\Sigma}_n^{vd} \\ \boldsymbol{\Sigma}_n^{vx} & \boldsymbol{\Sigma}_n^{dv} & \boldsymbol{\Sigma}_{v,n} \end{bmatrix}.$$

Such covariance structures commonly arise in algorithms such as the Schmidt–Kalman filter [34])

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

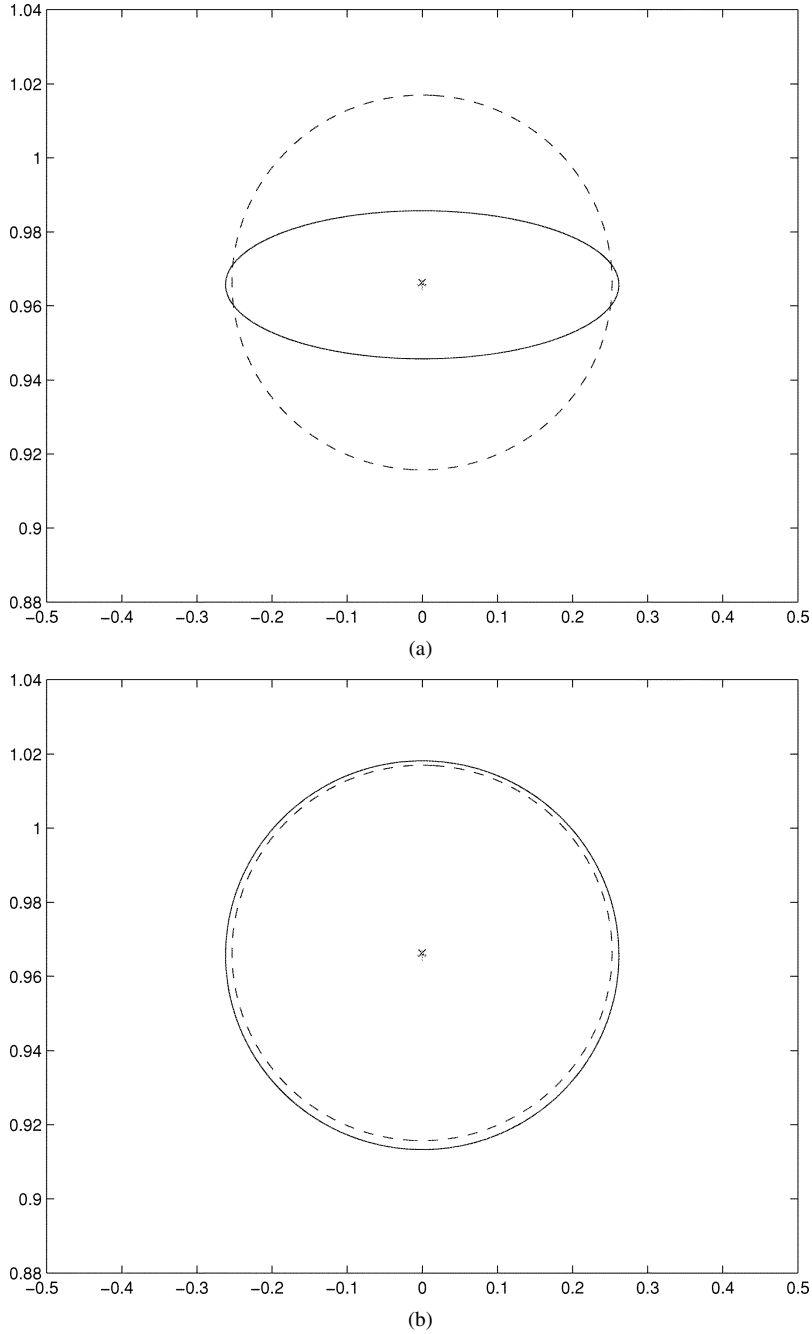

Fig. 6. The effect of the scaled UT on the mean and covariance. $\alpha = 10^{-3}, \beta = 2$, and $\mu = 10^{-6}$. Note that the UT estimate (solid) is conservative with respect to the true transformed estimate (dashed) assuming a Gaussian prior. (a) Without correction term. (b) With correction term.

the same level of accuracy as the propagated estimation errors. In other words, the estimate is correct to the second order and no derivatives are required.[11]

The unscented filter is summarized in Fig. 7. However, recall that this is a very general formulation and many specialized optimizations can be made. For example, if the process model is linear, but the observation model is not, the normal linear KF prediction equations can be used to calculate $\hat{\boldsymbol{\mu}}_n$ and $\hat{\mathbf{K}}_n$. The sigma points would be determined from the prediction distribution and would only be used to calculate $\hat{\mathbf{y}}_n, \hat{\boldsymbol{K}}_n^{xy}$, and $\hat{\boldsymbol{S}}_n$.

We now illustrate the use of the unscented KF in a stressing nonlinear application.

[11]It should be noted that this form, in common with standard EKF practice, approximates the impact of the process noise. Technically, the process noise is integrated through the discrete time interval. The approximation assumes that the process noise is sampled at the beginning of the time step and is held constant through the time step. Drawing parallels with the derivation of the continuous time filter from the discrete time filter, the dimensions of the process noise are correct if $\boldsymbol{\Sigma}_{d,n}$ is actually scaled by $1/t$.

B. Example

Consider the problem that is illustrated in Fig. 8. A vehicle enters the atmosphere at high altitude and at a very

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

1) The set of sigma points is created by applying a sigma point selection algorithm (e.g., (12) to the augmented system given by (24).
2) The transformed set is given by instantiating each point through the process model

$$\hat{\mathbf{x}}_{a,n}^{(i)} = \mathbf{f}\left[\mathbf{x}_{a,n}^{(i)}, \mathbf{u}_n\right]$$

3) The predicted mean is computed as

$$\hat{\mu}_{a,n} = \sum_{i=0}^{p} W^{(i)} \hat{\mathbf{x}}_{a,n}^{(i)}.$$

4) And the predicted covariance is computed as

$$\hat{\mathbf{K}}_{a,n} = \sum_{i=0}^{p} W^{(i)} \left\{\hat{\mathbf{x}}_{a,n}^{(i)} - \hat{\mu}_{a,n}\right\} \left\{\hat{\mathbf{x}}_{a,n}^{(i)} - \hat{\mu}_{a,n}\right\}^T.$$

5) Instantiate each of the prediction points through the observation model,

$$\hat{\mathbf{y}}_n^{(i)} = \mathbf{g}\left[\hat{\mathbf{x}}_{a,n}^{(i)}, \mathbf{u}_n\right].$$

6) The predicted observation is calculated by

$$\hat{\mathbf{y}}_n = \sum_{i=0}^{p} W^{(i)} \hat{\mathbf{y}}_n^{(i)}.$$

7) The innovation covariance is

$$\hat{\boldsymbol{S}}_n = \sum_{i=0}^{p} W^{(i)} \left\{\hat{\mathbf{y}}_n^{(i)} - \hat{\mathbf{y}}_n\right\} \left\{\hat{\mathbf{y}}_n^{(i)} - \hat{\mathbf{y}}_n\right\}^T.$$

8) The cross covariance matrix is determined by

$$\hat{\boldsymbol{K}}_n^{xy} = \sum_{i=0}^{p} W^{(i)} \left\{\hat{\mathbf{x}}_n^{(i)} - \hat{\mu}_n\right\} \left\{\hat{\mathbf{x}}_n^{(i)} - \hat{\mu}_n\right\}^T.$$

9) Finally, the update can be performed using the normal Kalman filter equations:

$$\begin{aligned}
\mu_n &= \hat{\mu}_n + \mathbf{W}_n \boldsymbol{\nu}_n \\
\mathbf{K}_n &= \hat{\mathbf{K}}_n - \mathbf{W}_n \hat{\boldsymbol{S}}_n \mathbf{W}_n^T \\
\boldsymbol{\nu}_n &= \mathbf{y}_n - \hat{\mathbf{y}}_n \\
\mathbf{W}_n &= \hat{\boldsymbol{K}}_n^{xy} \hat{\boldsymbol{S}}_n^{-1}.
\end{aligned}$$

Fig. 7. General formulation of the KF using the UT.

high speed. The position of the body is tracked by a radar which measures range and bearing. This type of problem has been identified by a number of authors [1], [9]–[11] as being particularly stressful for filters and trackers because of the strong nonlinearities exhibited by the forces which act on the vehicle. There are three types of forces in effect. The most dominant is aerodynamic drag, which is a function of vehicle speed and has a substantial nonlinear variation in altitude. The second type of force is gravity, which accelerates the vehicle toward the center of the earth. The final forces are random buffeting terms. The effect of these forces gives a trajectory of the form shown in Fig. 8: initially, the trajectory is almost ballistic; but as the density of the atmosphere increases, drag effects become important and the vehicle rapidly decelerates until its motion is almost vertical. The tracking problem is made more difficult by the fact that the drag properties of the vehicle might be only very crudely known.

In summary, the tracking system should be able to track an object that experiences a set of complicated, highly nonlinear forces. These depend on the current position and velocity of the vehicle as well as on certain characteristics which are not known precisely. The filter's state space consists of the position of the body (x_1 and x_2), its velocity (x_3 and x_4) and a parameter of its aerodynamic properties (x_5). The vehicle state dynamics are

$$\begin{aligned}
\dot{x}_1(k) &= x_3(k) \\
\dot{x}_2(k) &= x_4(k) \\
\dot{x}_3(k) &= D(k)x_3(k) + G(k)x_1(k) + v_1(k) \\
\dot{x}_4(k) &= D(k)x_4(k) + G(k)x_2(k) + v_2(k) \\
\dot{x}_5(k) &= v_3(k)
\end{aligned} \tag{25}$$

where $D(k)$ is the drag-related force term, $G(k)$ is the gravity-related force term, and $v.(k)$ is the process noise vector. The force terms are given by

$$D(k) = -\beta(k) \exp\left\{\frac{[R_0 - R(k)]}{H_0}\right\} V(k)$$

$$G(k) = -\frac{Gm_0}{r^3(k)}$$

and

$$\beta(k) = \beta_0 \exp x_5(k)$$

where $R(k) = \sqrt{x_1^2(k) + x_2^2(k)}$ (the distance from the center of the earth) and $V(k) = \sqrt{x_3^2(k) + x_4^2(k)}$ (speed).

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

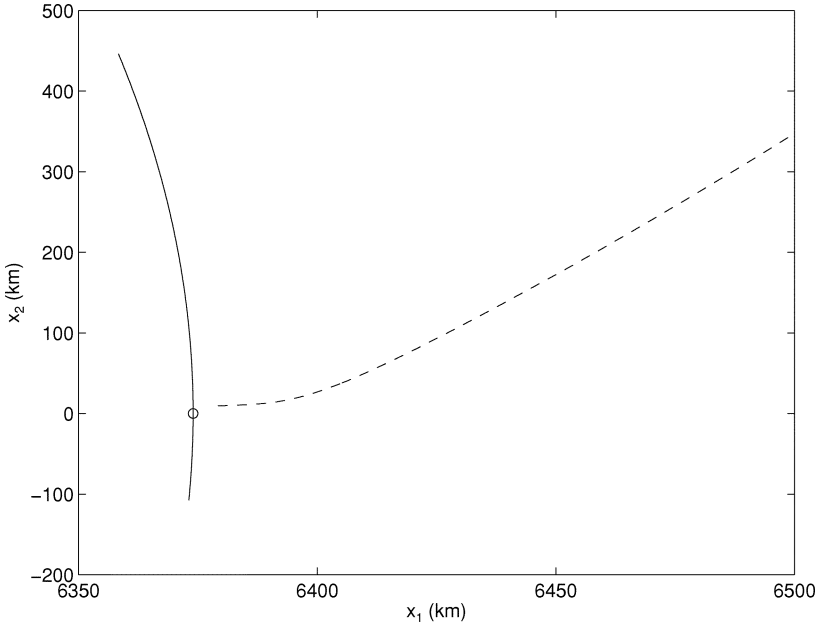

Fig. 8. The reentry problem. The dashed line is the sample vehicle trajectory and the solid line is a portion of the earth's surface. The position of the radar is marked by a ∘.

In this example $\beta_0 = -0.597\,83, H_0 = 13.406, Gm_0 = 3.9860 \times 10^5$, and $R_0 = 6374$ and reflect typical environmental and vehicle characteristics [10]. The parameterization of the ballistic coefficient $\beta(k)$ reflects the uncertainty in vehicle characteristics [1]. β_0 is the ballistic coefficient of a "typical vehicle" and it is scaled by $\exp(x_5(k))$ to ensure that its value is always positive. This is vital for filter stability.

The motion of the vehicle is measured by a radar that is located at (x_r, y_r). It is able to measure range r and bearing θ at a frequency of 10 Hz, where

$$r_r(k) = \sqrt{(x_1(k) - x_r)^2 + (x_2(k) - y_r)^2} + w_1(k)$$
$$\theta(k) = \tan^{-1}\left(\frac{x_2(k) - y_r}{x_1(k) - x_r}\right) + w_2(k)$$

$w_1(k)$ and $w_2(k)$ are zero-mean uncorrelated noise processes with variances of 1 m and 17 mrd, respectively [35]. The high update rate and extreme accuracy of the sensor means that a large quantity of extremely high quality data is available for the filter. The bearing uncertainty is sufficiently small that the EKF is able to predict the sensor readings accurately with very little bias.

The true initial conditions for the vehicle are

$$\mathbf{x}(0) = \begin{pmatrix} 6500.4 \\ 349.14 \\ -1.8093 \\ -6.7967 \\ 0.6932 \end{pmatrix}$$
$$\mathbf{P}(0) = \begin{bmatrix} 10^{-6} & 0 & 0 & 0 & 0 \\ 0 & 10^{-6} & 0 & 0 & 0 \\ 0 & 0 & 10^{-6} & 0 & 0 \\ 0 & 0 & 0 & 10^{-6} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

In other words, the vehicle's ballistic coefficient is twice the nominal value.

The vehicle is buffeted by random accelerations

$$\mathbf{Q}(k) = \begin{bmatrix} 2.4064 \times 10^{-5} & 0 & 0 \\ 0 & 2.4064 \times 10^{-5} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The initial conditions assumed by the filter are

$$\boldsymbol{\mu}_0 = \begin{pmatrix} 6500.4 \\ 349.14 \\ -1.8093 \\ -6.7967 \\ 0 \end{pmatrix}$$
$$\mathbf{K}_0 = \begin{bmatrix} 10^{-6} & 0 & 0 & 0 & 0 \\ 0 & 10^{-6} & 0 & 0 & 0 \\ 0 & 0 & 10^{-6} & 0 & 0 \\ 0 & 0 & 0 & 10^{-6} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The filter uses the nominal initial condition and, to offset for the uncertainty, the variance on this initial estimate is one.

Both filters were implemented in discrete time, and observations were taken at a frequency of 10 Hz. However, due to the intense nonlinearities of the vehicle dynamics equations, the Euler approximation of (25) was only valid for small time steps. The integration step was set to be 50 ms, which meant that two predictions were made per update. For the unscented filter, each sigma point was applied through the dynamics equations twice. For the EKF, it was necessary to perform an initial prediction step and relinearize before the second step.

The performance of each filter is shown in Fig. 9. This figure plots the estimated mean-squared estimation error (the

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

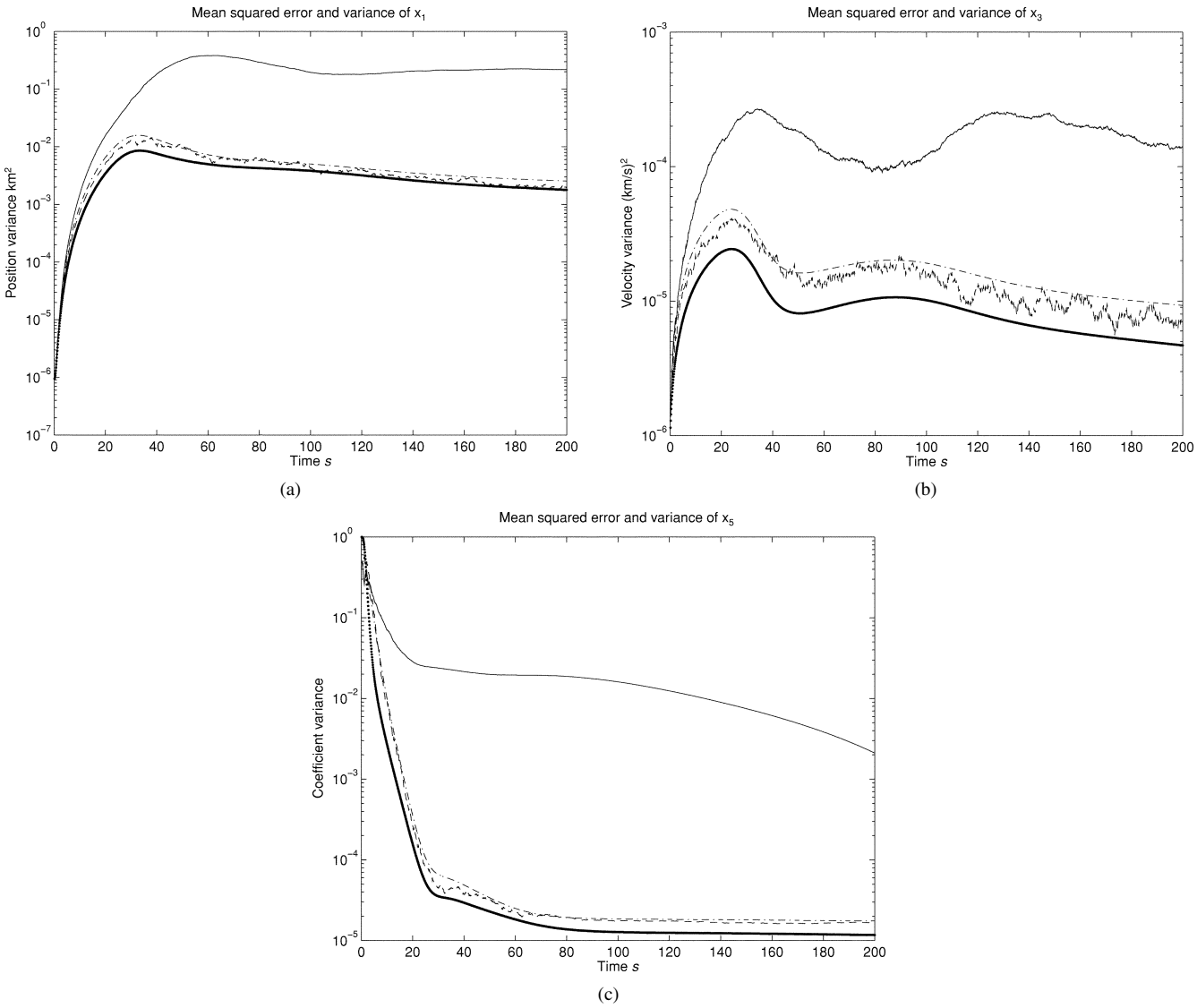

Fig. 9. Mean-squared errors and estimated covariances calculated by an EKF and an unscented Filter. In all the graphs, the solid line is the mean-squared error calculated by the EKF, and the dotted line is its estimated covariance. The dashed line is the unscented mean-squared error and the dot-dashed line its estimated covariance. (a) Results for x_1. (b) Results for x_3. (c) Results for x_5.

diagonal elements of $\mathbf{K}_n$) against actual mean-squared estimation error (which is evaluated using 100 Monte Carlo simulations). Only x_1, x_3, and x_5 are shown—the results for x_2 are similar to x_1, and x_4 is the same as that for x_3. In all cases it can be seen that the unscented filter estimates its mean-squared error very accurately, and it is possible to have confidence in the filter estimates. The EKF, however, is highly inconsistent: the peak mean-squared error in x_1 is 0.4 km^2, whereas its estimated covariance is over 100 times smaller. Similarly, the peak mean-squared velocity error is 3.4×10^{-4} km^{2}s^{-2}, which is over five times the true mean-squared error. Finally, it can be seen that x_5 for the EKF is highly biased, and this bias only slowly decreases over time. This poor performance is the direct result of linearization errors.

This section has illustrated the use of the UT in a KF for a nonlinear tracking problem. However, the UT has the potential to be used in other types of transformations as well.

VII. Further Applications of the UT

The previous section illustrated how the UT can be used to estimate the mean and covariance of a continuous nonlinear transformation more precisely. However, the UT is not limited to continuous distributions with well-defined Taylor series expansions.

A. Discontinuous Transformation

Consider the behavior of a two-dimensional (2-D) particle whose state consists of its position $\mathbf{x}(k) = [x(k), y(k)]^T$. The projectile is initially released at time 1 and travels at a constant and known speed v_x in the x direction. The objective is to estimate the mean position and covariance of the position at time 2 $[x(2), y(2)]^T$ where $\Delta T \triangleq t_2 - t_1$. The problem is made difficult by the fact that the path of the projectile is obstructed by a wall that lies in the "bottom right quarter-plane" $(x \geq 0, y \leq 0)$. If the projectile hits the wall,

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

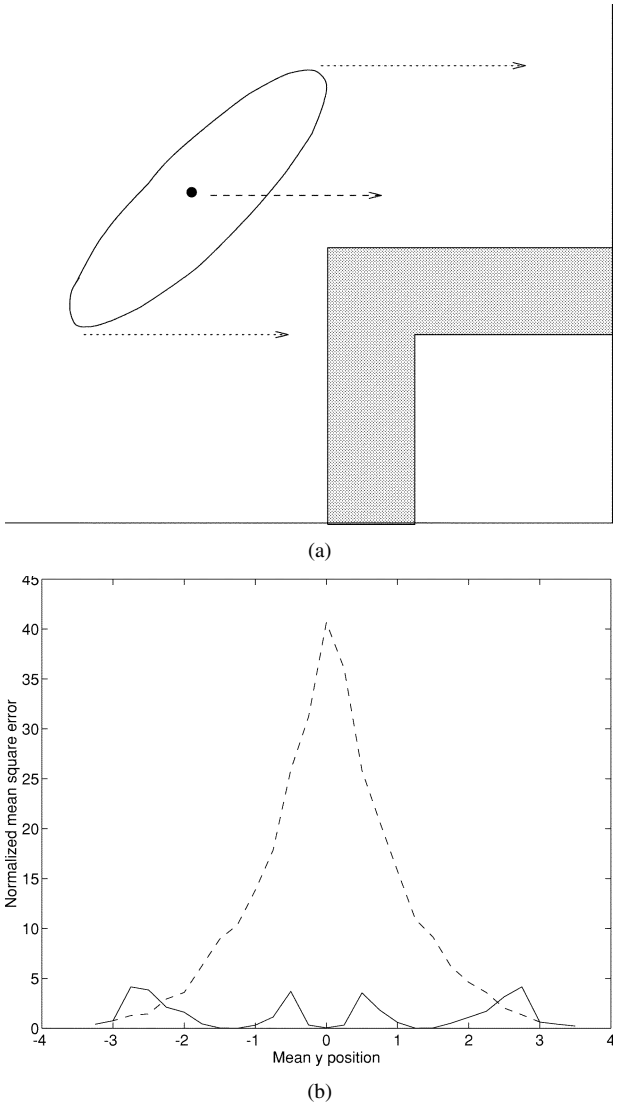

Fig. 10. Discontinuous system example: a particle can either strike a wall and rebound or continue to move in a straight line. The experimental results show the effect of using different start values for y. (a) Problem scenario. (b) Mean-squared prediction error against initial value of y. Solid is UF, dashed is EKF.

there is a perfectly elastic collision and the projectile is reflected back at the same velocity as it traveled forward. This situation is illustrated in Fig. 10(a), which also shows the 1σ covariance ellipse of the initial distribution.

The process model for this system is

$$x(2) = \begin{cases} x(1) + \Delta T v_x, & y(1) \geq 0 \\ -(x(1) + \Delta T v_x), & y(1) < 0 \end{cases} \tag{26}$$

$$y(2) = y(1). \tag{27}$$

At time step 1, the particle starts in the left half-plane ($x \leq 0$) with position $[x(1), y(1)]^T$. The error in this estimate is Gaussian, has zero mean, and has covariance $\mathbf{K}_1$. Linearized about this start condition, the system appears to be a simple constant velocity linear model.

The true conditional mean and covariance was determined using Monte Carlo simulation for different choices of the initial mean of y. The mean-squared error calculated by the EKF and by the UT for different values is shown in Fig. 10(b). The UT estimates the mean very closely, suffering only small spikes as the translated sigma points successively pass the wall. Further analysis shows that the covariance for the filter is only slightly larger than the true covariance, but conservative enough to account for the deviation of its estimated mean from the true mean. The EKF, however, bases its entire estimate of the conditional mean on the projection of the prior mean, so its estimates bear no resemblance to the true mean, except when most of the distribution either hits or misses the wall and the effect of the discontinuity is negligible.

B. Multilevel Sensor Fusion

The UT can also be used to bridge the gap between low-level filters based on the KF and high-level systems based on artificial intelligence or related mechanisms. The basic problem associated with multilevel data fusion is that different levels are typically concerned with different pieces of information represented in very different ways. For example, a low-level filter may maintain information about states relating to, e.g., wheel slip of a vehicle, that are only of indirect relevance to a high-level guidance system that maintains information relating to, e.g., whether the vehicle is veering toward the edge of the roadway. The multilevel data fusion problem has been decomposed into a set of hierarchical domains [36]. The lowest levels, Level 0 and Level 1 (object refinement), are concerned with quantitative data fusion problems such as the calculation of a target track. Level 2 (situation refinement) and Level 3 (threat refinement) apply various high-level data fusion and pattern recognition algorithms to attempt to glean strategic and tactical information from these tracks.

The difficulty lies in the fundamental differences in the representation and use of information. On the one hand, the low-level tracking filter only provides mean and covariance information. It does not specify an exact kinematic state from which an expert system could attempt to infer a tactical state (e.g., relating to "intents" or "goals"). On the other hand, an expert system may be able to accurately predict the behavior of a pilot under a range of situations. However, the system does not define a rigorous low-level framework for fusing its predictions with raw sensor information to obtain high-precision estimates suitable for reliable tracking. The practical solution to this problem has been to take the output of standard control and estimation routines, discretize them into a more symbolic form (e.g., "slow" or "fast"), and process them with an expert/fuzzy rule base. The results of such processing are then converted into forms that can be processed by conventional process technology.

One approach for resolving this problem, illustrated in Fig. 11, is to combine the different data fusion algorithms together into a single, composite data fusion algorithm that takes noise-corrupted raw sensor data and provides the inferred high level state. From the perspective of the track estimator, the higher level fusion rules are considered to be arbitrary, nonlinear transformations. From the perspective of the higher level data fusion algorithms, the UT converts

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

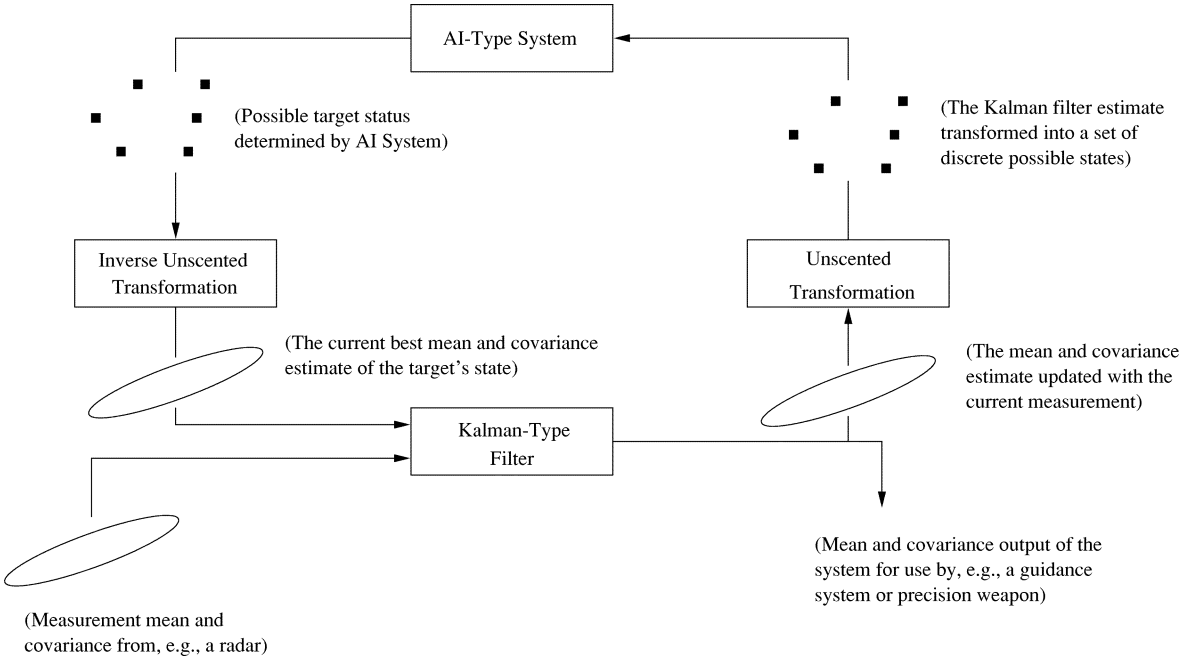

Fig. 11. Possible framework for multilevel information fusion using the UT.

the output from the low-level tracker into a set of vectors. Each vector is treated as a *possible* kinematic state which is processed by the higher level fusion algorithms. In other words, the low-level tracking algorithms do not need to understand the concept of higher level constructs, such as maneuvers, whereas the higher level algorithms do not need to understand or produce probabilistic information.

Consider the problem of tracking an aircraft. The aircraft model consists of two components—a kinematic model, which describes the trajectory of the aircraft for a given set of pilot inputs, and an expert system, which attempts to infer current pilot intentions and predict future pilot inputs. The location of the aircraft is measured using a radar tracking system.

Some sigma points might imply that the aircraft is making a rapid acceleration, some might indicate a moderate acceleration, and yet others might imply that there is no discernible acceleration. Each of the state vectors produced from the UT can be processed individually by the expert system to predict a possible future state of the aircraft. For some of the state vectors, the expert system will signal an evasive maneuver and predict the future position of the aircraft accordingly. Other vectors, however, will not signal a change of tactical state and the expert system will predict that the aircraft will maintain its current speed and bearing. The second step of the UT consists of computing the mean and covariance of the set of predicted state vectors from the expert system. This mean and covariance gives the predicted state of the aircraft in a form that can then be fed back to the low-level filter. The important observation to be made is that this mean and covariance reflects the probability that the aircraft will maneuver *even though the expert system did not produce any probabilistic information and the low-level filter knows nothing about maneuvers*.

VIII. Summary and Conclusion

In this paper we have presented and discussed the UT. We explored how it can be used to transform statistical information through a nonlinear transformation. Although it has some similarity to particle filters, there are two important distinctions. First, the sigma points are chosen deterministically from the statistics of the transformation; therefore, second-order properties of the distribution can be propagated with only a small amount of statistical information. Second, the approximation itself can be interpreted more generally than as a probability distribution, and we have shown how this generality can be exploited.

The UT method has found a number of applications in high-order, nonlinear coupled systems including navigation systems for high-speed road vehicles [37], [38], public transportation systems [39], data assimilation systems [40], and underwater vehicles [41]. Square root filters can be formulated in terms of discrete sets of points (as demonstrated in [42]), and iterated filters can be constructed using the predictions [43]. The algorithm has been extended to capture the first four moments of a Gaussian distribution [44] and the first three moments of an arbitrary distribution [45].

More recently, other authors have explored extensions and refinements to the basic algorithm and concepts. Van Der Merwe [46] has developed a square root formulation of the UT which propagates the mean and square root of the covariance matrix, rather than the covariance matrix itself. Van Der Merwe has extended the concept of sigma point filters to work with particle filters and sums of Gaussians [47].

Although this paper has considered many aspects of the various ways in which the UT can be tailored to address the subtleties of particular applications or performance concerns,

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

it is important to recognize that the basic UT algorithm is conceptually very simple and easy to apply. In this respect the UT has the same appeal as linearization for the EKF, but unlike linearization the UT provides sufficient accuracy to be applied in many highly nonlinear filtering and control applications.

Appendix I
Taylor Series and Moments

This Appendix analyzes the performance of the prediction with respect to the Taylor series expansion and moments. $\bar{\mathbf{z}}$ is the expected value of (3)

$$\begin{aligned}\bar{\mathbf{z}} &= \mathrm{E}[\mathbf{h}[\bar{\mathbf{x}}+\mathbf{e}]] \\ &= \mathbf{h}[\bar{\mathbf{x}}]+\mathrm{E}\left[\mathbf{D_e h}+\frac{\mathbf{D_e^2 h}}{2!}+\frac{\mathbf{D_e^3 h}}{3!}+\frac{\mathbf{D_e^4 h}}{4!}+\cdots\right] \quad (28)\end{aligned}$$

where the ith term in the series is given by

$$\begin{aligned}&\mathrm{E}\left[\frac{\mathbf{D_e^i h}}{i!}\right] \\ &= \frac{1}{i!}\mathrm{E}\left[\left(\sum_{j=1}^{n}\tilde{x}_j\frac{\partial}{\partial x_j}\right)^i\right]\mathbf{h}[\bar{\mathbf{x}}] \\ &= \frac{1}{i!}\left(m_{1\ldots11}\frac{\partial^i\mathbf{h}}{\partial {x_1}^i}+m_{1\ldots12}\frac{\partial^i\mathbf{h}}{\partial {x_1}^{i-1}\partial x_2}+\cdots\right) \quad (29)\end{aligned}$$

and $m_{c_1c_2\ldots c_i}$ is the ith-order moment $m_{c_1c_2\ldots c_i} = \mathrm{E}[e_{c_1}e_{c_2}\ldots e_{c_i}]$. Therefore, if the mean is to be correctly calculated to the mth order, both the derivatives of $\mathbf{h}[\cdot]$ and the moments of $\mathbf{e}$ must be known up to the mth order.

The covariance $\mathbf{\Sigma}_z$ is given by

$$\mathbf{\Sigma}_z = \mathrm{E}[\{\mathbf{z}-\bar{\mathbf{z}}\}\{\mathbf{z}-\bar{\mathbf{z}}\}^T].$$

Substituting from (3) and (28)

$$\begin{aligned}\mathbf{z}-\bar{\mathbf{z}} &= \mathbf{h}[\bar{\mathbf{x}}+\mathbf{e}]-\mathrm{E}[\mathbf{h}[\bar{\mathbf{x}}+\mathbf{e}]] \\ &= \mathbf{D_e h}+\frac{\mathbf{D_e^2 h}}{2!}+\frac{\mathbf{D_e^3 h}}{3!}+\frac{\mathbf{D_e^4 h}}{4!} \\ &\quad -\mathrm{E}\left[\frac{\mathbf{D_e^2 h}}{2!}+\frac{\mathbf{D_e^4 h}}{4!}+\cdots\right]\end{aligned}$$

Taking outer products and expectations and exploiting the fact that $\mathbf{e}$ is symmetric, the odd terms all evaluate to zero and the covariance is given by

$$\begin{aligned}\mathbf{\Sigma}_z &= \mathbf{A}(\mathbf{x})\mathbf{\Sigma}_x\mathbf{A}^T(\mathbf{x}) \\ &\quad +\mathrm{E}\left[\frac{\mathbf{D_e h}\left(\mathbf{D_e^3 h}\right)^T}{3!}+\frac{\mathbf{D_e^2 h}\left(\mathbf{D_e^2 h}\right)^T}{2\times 2!}\right. \\ &\quad \left.+\frac{\mathbf{D_e^3 h}\left(\mathbf{D_e h}\right)^T}{3!}\right] \\ &\quad -\mathrm{E}\left[\frac{\mathbf{D_e^2 h}}{2}\right]\mathrm{E}\left[\frac{\mathbf{D_e^2 h}}{2}\right]^T+\cdots \quad (30)\end{aligned}$$

where $\mathrm{E}[\mathbf{D_e h}(\mathbf{D_e h})^T] = \mathbf{A}(\mathbf{x})\mathbf{\Sigma}_x\mathbf{A}^T(\mathbf{x})$ has been used.

In other words, the mth-order term in the covariance series is calculated correctly only if $\mathbf{h}[\cdot]$ and the moments of $\mathbf{e}$ are known up to the $2m$th order.

Comparing this with (28), it can be seen that if the moments and derivatives are known to a given order, the order of accuracy of the mean estimate is higher than that of the covariance.

Appendix II
Moment Characteristics of the Symmetric Sigma Point Sets

This Appendix analyzes the properties of the higher moments of the symmetric sigma point sets. This Appendix analyzes the set given in (12). The set presented in (11) corresponds to the special case $W^{(0)} = 0$.

Without loss of generality, assume that the prior distribution has a mean $\mathbf{0}$ and covariance $\mathbf{I}$. An affine transformation can be applied to rescale and translate this set to an arbitrary mean and covariance. Using these statistics, the points selected by (12) are symmetrically distributed about the origin and lie on the coordinate axes of the state space. Let $m_{c_1c_2\ldots c_i}$ be the ith-order moment

$$\begin{aligned}m_{c_1c_2\ldots c_k} &= \mathrm{E}\left[x_{c_1}x_{c_2}\ldots x_{c_k}\right] \\ &= \sum_{i=0}^{p}W^{(i)}x_{c_1}^{(i)}x_{c_2}^{(i)}\ldots x_{c_k}^{(i)} \quad (31)\end{aligned}$$

where $x_{c_j}^{(i)}$ is the c_jth component of the ith sigma point.

From the orthogonality and symmetry of the points, it can be shown that only the even-ordered moments where all indices are the same $(c_1 = c_2 = \cdots = c_k)$ are nonzero. Furthermore, each moment has the same value given by

$$\begin{aligned}m_{c_k\ldots c_k} &= \sum_{i=0}^{p}W^{(i)}\left(x_{c_1}^{(i)}\right)^k \quad (32) \\ &= \frac{1-W^{(0)}}{N_x}\left(\frac{N_x}{1-W^{(0)}}\right)^{k/2}. \quad (33)\end{aligned}$$

Consider the case when the sigma points approximate a Gaussian distribution. From the properties of a Gaussian it can be shown that [22]

$$\mathrm{E}\left[x_1^{\alpha_1}\times\cdots\times {x_n}^{\alpha_n}\right] = \prod_{i=1}^{n}\mathrm{E}\left[{x_i}^{\alpha_i}\right]$$

where the moment contains α_i repetitions of each element. (For example, the moment $m_{11\,223\,333}$ has $\alpha_1 = 2, \alpha_2 = 2, \alpha_3 = 4$). For each element

$$\mathrm{E}[x_i^{\alpha_i}] = \begin{cases}0, & \alpha_i \text{ is odd} \\ 1\times 3\times\cdots\times(\alpha_i-1), & \alpha_i \text{ is even}\end{cases}.$$

Comparing this moment scaling with that from the symmetric sigma point set, two differences become apparent. First, the "scaling" of the points are different. The moments of a Gaussian increase factorially whereas the moments of the sigma point set increases geometrically. Second, the sigma point moment set "misses out" many of the moment terms. For example, the Gaussian moment $\mathrm{E}[x_1^2x_2^2] = 1$ whereas the sigma point set moment is $\mathrm{E}[x_1^2x_2^2] = 0$. These differences become more marked as the order increases.

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

However, these effects are mitigated by the fact that the terms in the Taylor series have inverse factorial weights.

Assuming that the contributions of terms diminish as the order increases, the emphasis should be to minimize the error in the lowest order in the series. The first errors occur in the fourth order. For a Gaussian distribution, $\mathrm{E}[x_i^4] = 3$. The symmetric sigma point set can achieve the same value with $W^{(0)} = 1 - N_x/3$.

Because the fourth and higher order moments are not precisely matched, the choice of matrix square root affects the errors in the higher order terms by adjusting the way in which the errors are distributed amongst the different states. This issue is discussed in more detail in Appendix V.

Appendix III
Simplex Sigma Point Sets

The computational costs of the UT are proportional to the number of sigma points. Therefore, there is a strong incentive to reduce the number of points used. This demand come from both small-scale applications such as head tracking, where the state dimension is small but real time behavior is vital, to large-scale problems, such as weather forecasting, where there can be many thousands or tens of thousands of states [48].

The minimum number of sigma points are the smallest number required which have mean and covariance $\bar{\mathbf{x}}$ and $\mathbf{\Sigma}_x$. From the properties of the outer products of vectors

$$\min(p+1, N_x) \geq \text{rank}\left[\sum_{i=0}^{p} W^{(i)}\left(\mathbf{x}^{(i)} - \bar{\mathbf{x}}\right)\left(\mathbf{x}^{(i)} - \bar{\mathbf{x}}\right)^T\right].$$

Therefore, at least $N_x + 1$ points are required. This result makes intuitive sense. In two dimensions, any two points are colinear and so their variance cannot exceed rank 1. If a third point is added (forming a triangle), the variance can equal 2. Generalizing, a N_x-dimensional space can be matched using a *simplex* of $N_x + 1$ vertices.

Although any N_x points are sufficient, additional constraints must be supplied to determine which points should be used. In [49] we chose a set of points that minimize the skew (or third-order moments) of the distribution. Recently, we have investigated another set that places all of the sigma points at the same distance from the origin [50]. This choice is motivated by numerical and approximation stability considerations. This spherical simplex sigma set of points consists of points that lie on a hypersphere.

The basic algorithm is as follows.

1) Choose $0 \leq W_0 \leq 1$.
2) Choose weight sequence
$$W^{(i)} = (1 - W_0)/(N_x + 1).$$
3) Initialize vector sequence as
$$\tilde{\mathbf{x}}_{(0)}1 = [0]$$
$$\tilde{\mathbf{x}}^{(1)}1 = \left[-\frac{1}{\sqrt{2W_1}}\right]$$

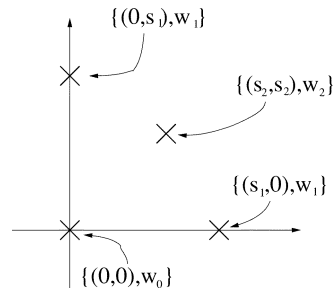

Fig. 12. Fourth-order sigma points.

and
$$\tilde{\mathbf{x}}^{(2)}1 = \left[\frac{1}{\sqrt{2W_1}}\right].$$

4) Expand vector sequence for $j = 2, \ldots, N_x$ according to
$$\mathbf{x}^{(i)} = \begin{cases} \begin{bmatrix} \tilde{\mathbf{x}}^{(0)}j - 1 \\ 0 \end{bmatrix}, & \text{for } i = 0 \\ \begin{bmatrix} \tilde{\mathbf{x}}^{(i)}j - 1 \\ -\frac{1}{\sqrt{j(j+1)W_1}} \end{bmatrix}, & \text{for } i = 1, \ldots, j \\ \begin{bmatrix} \mathbf{0}_j \\ \frac{j}{\sqrt{j(j+1)W_1}} \end{bmatrix}, & \text{for } i = j+1. \end{cases}$$

This algorithm has two notable two features. First, the weight on each sigma point (apart from the zeroth point) is the same and is proportional to $(1 - W_0)/(N_x + 1)$. Second, all of the sigma points (apart from the zeroth point) lie on the hypersphere of radius $\sqrt{N_x}/(1 - W_0)$.

Appendix IV
Capturing the Kurtosis

This Appendix shows how a sigma point set can be chosen which matches the first four moments of a Gaussian distribution. It was originally presented in [51].

If the sigma points are to capture the kurtosis of a Gaussian distribution precisely, they must have the property that

$$m_{c_1c_2c_3c_4} = \begin{cases} 1, & \text{when } c_1 = c_2, c_3 = c_4 \text{ and } c_1 \neq c_3 \\ 3, & \text{when } c_1 = c_2 = c_3 = c_4 \\ 0, & \text{otherwise} \end{cases}. \tag{34}$$

Since there are $O(N_x^2)$ constraints, $O(N_x^2)$ sigma points are required. Guided by the structure of the second-order solution, we introduce the minimum number of extra points which retain the symmetry of the distribution. Fig. 12 shows the sigma points in one quadrant for a 2-D case. Three types of points are used. The first type consists of a single point which lies at the origin with weight w_0. The second type of point lies on the coordinate axes a distance s_1 from the origin and with weight w_1. There are four occurrences of the third type of point. These lie at $(\pm s_2, \pm s_2)$ and are assigned a weight w_2. The form of the distribution generalizes in higher dimensions. In three dimensions, there are six type 2 points and 12 type 3 points. The type 3 points lie at $(\pm s_2, \pm s_2, 0), (\pm s_2, 0, \pm s_2)$, and $(0, \pm s_2, \pm s_2)$. Extending

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

Table 1
Points and Weights for $n = 3$ Dimensions for the Fourth-Order UT

i	s_i	W_i
0	0	0.3583
1	3.2530	0.0045
2	1.4862	0.0512

this example to a higher number of dimensions, it can be shown that $2N_x^2+1$ points are required. As a result, the comptutational costs are $O(N_x^4)$, which is the same order as that required for analytically derived fourth-order filters.

The constraints placed on the points are for the covariance, kurtosis, and normalization. The symmetry means that these conditions need only be ensured in the z_1 and z_2 directions

$$\mathbf{g}[\mathcal{S}, p_z(\mathbf{z})] = \begin{pmatrix} 2w_1 s_1^2 + 4(N_x - 1)w_2 s_2^2 - 1 \\ 2w_1 s_1^4 + 4(N_x - 1)w_2 s_2^4 - 3 \\ 4w_2 s_2^2 - 1 \\ w_0 + 2N_x w_1 + 2n(N_x - 1)w_2 - 1 \end{pmatrix}.$$

The solutions to these equations possess a single degree of freedom. However, unlike the second-order transform, there does not appear to be a convenient set of parameters which can be used to capture this freedom. Even so, it is possible to apply a disutility function which adjust the properties of the desired distribution. Extending from the results for the second-order filter, we choose a disutility function which minimizes the discrepancy of the sixth-order moments. Since $m_{111\,111} = 15$, the cost function is

$$c[\mathcal{S}, p_z(\mathbf{z})] = \left|2w_1 s_1^6 + 4(n-1)w_2 s_2^6 - 15\right|.$$

The values of w_0, w_1, w_2, s_1, and s_2 are given in Table 1 for the case when $n = 3$. A total of 19 sigma points are required.

A more formal derivation and generalization of this approach was carried out by Lerner [31]

Appendix V
Scaled Sigma Points

This Appendix describes an alternative method of scaling the sigma points which is equivalent to the approach described in Section V. That section described how a modified form of the nonlinear transformation could be used to influence the effects of the higher order moments. In this Appendix, we show that the same effect can be achieved by modifying the sigma point set rather than the transformation.

Suppose a set of sigma points $\mathcal{S}$ have been constructed with mean $\bar{\mathbf{x}}$ and covariance $\mathbf{\Sigma}_x$.

From (15), the mean is

$$\begin{aligned}\bar{\mathbf{z}} &= \sum_{i=0}^{p} W^{(i)} \tilde{\mathbf{h}}[\mathbf{x}^{(i)}] \\ &= \sum_{i=0}^{p} W^{(i)} \left(\frac{\mathbf{h}\left[\bar{\mathbf{x}} + \alpha(\mathbf{x}^{(i)} - \bar{\mathbf{x}})\right] - \mathbf{h}[\bar{\mathbf{x}}]}{\mu} + \mathbf{h}[\bar{\mathbf{x}}] \right) \\ &= \left(\frac{W^{(0)}}{\mu} + \frac{\mu - 1}{\mu} \right) \mathbf{h}[\bar{\mathbf{x}}] \\ &\quad + \sum_{i=0}^{p} \frac{W^{(i)}}{\mu} \mathbf{h}\left[\bar{\mathbf{x}} + \alpha\left(\mathbf{x}^{(i)} - \bar{\mathbf{x}}\right)\right].\end{aligned}$$

The same mean can be calculated by applying a different sigma point set $\tilde{\mathcal{S}} = \{i = 0, \ldots, p : \tilde{\mathbf{x}}^{(i)}, \tilde{W}^{(i)}\}$ to $\mathbf{h}[\cdot]$. $\tilde{\mathcal{S}}$ and $\mathcal{S}$ are related according to the expression

$$\begin{aligned}\tilde{\mathbf{x}}^{(i)} &= \alpha \mathbf{x}^{(i)} + (1 - \alpha)\mathbf{x}^{(0)} \\ \tilde{W}^{(i)} &= \begin{cases} \left(W^{(0)} + \mu - 1\right)/\mu & i = 0 \\ W^{(i)}/\mu & i \neq 0 \end{cases}.\end{aligned}$$

Given this set of points, the scaled UT calculates its statistics as follows:

$$\begin{aligned}\tilde{\mathbf{z}}^{(i)} &= \mathbf{h}\left[\tilde{\mathbf{x}}^{(i)}\right] \\ \bar{\mathbf{z}} &= \sum_{i=0}^{p} \tilde{W}^{(i)} \tilde{\mathbf{z}}^{(i)} \\ \mathbf{\Sigma}_z &= \sum_{i=0}^{p} \tilde{W}^{(i)} \left\{\tilde{\mathbf{z}}^{(i)} - \bar{\mathbf{z}}\right\} \left\{\tilde{\mathbf{z}}^{(i)} - \bar{\mathbf{z}}\right\}^T \\ &\quad + (1 - \alpha^2) \left\{\tilde{\mathbf{z}}^{(0)} - \bar{\mathbf{z}}\right\} \left\{\tilde{\mathbf{z}}^{(0)} - \bar{\mathbf{z}}\right\}^T. \qquad (35)\end{aligned}$$

Appendix VI
Incorporating Covariance Correction Terms

Although the sigma points only capture the first two moments of the sigma points (and so the first two moments of the Taylor series expansion), the scaled UT can be extended to include partial higher order information of the fourth-order term in the Taylor series expansion of the covariance [52]. The fourth-order term of (30) is

$$\begin{aligned}\mathbf{A} = \frac{1}{4}\mathbf{D}_{\mathbf{e}}^2\mathbf{h}\left(\mathrm{E}[\mathbf{e}^4] - \mathrm{E}[\mathbf{e}^2\mathbf{\Sigma}_x] - \mathrm{E}[\mathbf{\Sigma}_x\mathbf{e}^2] + \mathbf{\Sigma}_x^2\right)\left(\mathbf{D}_{\mathbf{e}}^2\mathbf{h}\right)^T \\ + \frac{1}{3!}\mathbf{D}_{\mathbf{e}}^3\mathbf{h}\mathrm{E}[\mathbf{e}^4](\mathbf{D}_{\mathbf{e}}\mathbf{h})^T. \qquad (36)\end{aligned}$$

The term $(1/2)\mathbf{D}_{\mathbf{e}}^2\mathbf{h}\mathbf{\Sigma}_y^2(\mathbf{D}_{\mathbf{e}}^2\mathbf{h})^T$ can be calculated from the same set of sigma points which match the mean and covariance. From (3) and (28)

$$\mathbf{z}^{(0)} - \bar{\mathbf{z}} = \frac{1}{2}\mathbf{D}_{\mathbf{e}}^2\mathbf{h}\mathbf{\Sigma}_x + \frac{1}{6}\mathbf{D}_{\mathbf{e}}^3\mathbf{h}\,\mathrm{E}[\mathbf{e}^3] + \cdots.$$

Taking outer products

$$\left(\bar{\mathbf{z}} - \mathbf{z}^{(0)}\right)\left(\bar{\mathbf{z}} - \mathbf{z}^{(0)}\right)^T = \frac{1}{4}\mathbf{D}_{\mathbf{e}}^2\mathbf{h}\mathbf{\Sigma}_x^2\left(\mathbf{D}_{\mathbf{e}}^2\mathbf{h}\right)^T + \cdots.$$

Therefore, adding extra weighting to the contribution of the zeroth point, further higher order effects can be incorporated at no additional computational cost by rewriting (35) as

$$\begin{aligned}\mathbf{\Sigma}_z^* = \sum_{i=0}^{p} W^{(i)} \left\{\mathbf{z}^{(i)} - \bar{\mathbf{z}}\right\} \left\{\mathbf{z}^{(i)} - \bar{\mathbf{z}}\right\}^T \\ + (\beta + 1 - \alpha^2) \left\{\mathbf{z}^{(0)} - \bar{\mathbf{z}}\right\} \left\{\mathbf{z}^{(0)} - \bar{\mathbf{z}}\right\}^T.\end{aligned}$$

In this form, the error in the fourth-order term is

$$\begin{aligned}\Delta\mathbf{A} = \frac{1}{4}\mathbf{D}_{\mathbf{e}}^2\mathbf{h}\left(\mathrm{E}[\mathbf{e}^4] - \mathrm{E}[\mathbf{e}^2\mathbf{\Sigma}_x]\right. \\ \left. - \mathrm{E}[\mathbf{\Sigma}_x\mathbf{e}^2] + (1 - \beta)\mathbf{\Sigma}_x^2\right)\left(\mathbf{D}_{\mathbf{e}}^2\mathbf{h}\right)^T \\ + \frac{1}{3!}\mathbf{D}_{\mathbf{e}}^3\mathbf{h}\mathrm{E}[\mathbf{e}^4](\mathbf{D}_{\mathbf{e}}\mathbf{h})^T.\end{aligned}$$

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

In the special case that μ Gaussian-distributed, $\mathrm{E}[\mathbf{e}^4] = 3\boldsymbol{\Sigma}_x^2$ and so the error is

$$\Delta\mathbf{A} = (2-\beta)\mathbf{D}_{\mathbf{e}}^2\mathbf{h}\boldsymbol{\Sigma}_x^2\left(\mathbf{D}_{\mathbf{e}}^2\mathbf{h}\right)^T + \frac{1}{3!}\mathbf{D}_{\mathbf{e}}^3\mathbf{h}\mathrm{E}[\mathbf{e}^4](\mathbf{D}_{\mathbf{e}}\mathbf{h})^T. \quad (37)$$

Under the assumption that no information about $\mathbf{h}[\cdot]$ is used, this term is minimized when $\beta = 2$.

REFERENCES

[1] P. J. Costa, "Adaptive model architecture and extended Kalman–Bucy filters," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, pp. 525–533, Apr. 1994.

[2] G. Prasad, G. W. Irwin, E. Swidenbank, and B. W. Hogg, "Plant-wide predictive control for a thermal power plant based on a physical plant model," in *IEE Proc.—Control Theory Appl.*, vol. 147, Sept. 2000, pp. 523–537.

[3] H. J. Kushner, "Dynamical equations for optimum nonlinear filtering," *J. Different. Equat.*, vol. 3, pp. 179–190, 1967.

[4] D. B. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. Automat. Contr.*, vol. 24, pp. 843–854, Dec. 1979.

[5] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 439–447, Aug. 1972.

[6] K. Murphy and S. Russell, "Rao–Blackwellised particle filters for dynamic Bayesian networks," in *Sequential Monte Carlo Estimation in Practice*. ser. Statistics for Engineering and Information Science, A. Doucet, N. de Freitas, and N. Gordon, Eds. New York: Springer-Verlag, 2000, ch. 24, pp. 499–515.

[7] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. San Diego, CA: Academic, 1970.

[8] H. W. Sorenson, Ed., *Kalman Filtering: Theory and Application*. Piscataway, NJ: IEEE, 1985.

[9] M. Athans, R. P. Wishner, and A. Bertolini, "Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy measurements," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 504–518, Oct. 1968.

[10] J. W. Austin and C. T. Leondes, "Statistically linearized estimation of reentry trajectories," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, pp. 54–61, Jan. 1981.

[11] R. K. Mehra, "A comparison of several nonlinear filters for reentry vehicle tracking," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 307–319, Aug. 1971.

[12] D. Lerro and Y. K. Bar-Shalom, "Tracking with debiased consistent converted measurements vs. EKF," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-29, pp. 1015–1022, July 1993.

[13] T. Viéville and P. Sander, "Using pseudo Kalman-filters in the presence of constraints application to sensing behaviors," INRIA, Sophia Antipolis, France, Tech. Rep. 1669, 1992.

[14] J. K. Tugnait, "Detection and estimation for abruptly changing systems," *Automatica*, vol. 18, no. 5, pp. 607–615, May 1982.

[15] K. Kastella, M. A. Kouritzin, and A. Zatezalo, "A nonlinear filter for altitude tracking," in *Proc. Air Traffic Control Assoc.*, 1996, pp. 1–5.

[16] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[17] J. K. Kuchar and L. C. Yang, "Survey of conflict detection and resolution modeling methods," presented at the AIAA Guidance, Navigation, and Control Conf., New Orleans, LA, 1997.

[18] P. A. Dulimov, "Estimation of ground parameters for the control of a wheeled vehicle," M.S. thesis, Univ. Sydney, Sydney, NSW, Australia, 1997.

[19] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME J. Basic Eng.*, vol. 82, pp. 34–45, Mar. 1960.

[20] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*. New York: Academic, 1988.

[21] J. Leonard and H. F. Durrant-Whyte, *Directed Sonar Sensing for Mobile Robot Navigation*. Boston, MA: Kluwer, 1991.

[22] P. S. Maybeck, *Stochastic Models, Estimation, and Control*. New York: Academic, 1979, vol. 1.

[23] L. Mo, X. Song, Y. Zhou, Z. Sun, and Y. Bar-Shalom, "Unbiased converted measurements in tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, pp. 1023–1027, July 1998.

[24] J. K. Uhlmann, "Simultaneous map building and localization for real time applications," transfer thesis, Univ. Oxford, Oxford, U.K., 1994.

[25] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[26] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*. San Diego, CA: Academic, 1975.

[27] T. Lefebvre, H. Bruyninckx, and J. De Schuller, "Comment on 'A new method for the nonlinear transformation of means and covariances in filters and estimators [and author's reply]'," *IEEE Trans. Automat. Contr.*, vol. 47, pp. 1406–1409, Aug. 2002.

[28] J. Holtzmann, "On using perturbation analysis to do sensitivity analysis: Derivaties vs. differences," in *IEEE Conf. Decision and Control*, 1989, pp. 2018–2023.

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

[30] H. J. Kushner, "Approximations to optimal nonlinear filters," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 546–556, Oct. 1967.

[31] U. N. Lerner, "Hybrid Bayesian networks for reasoning about complex systems," Ph.D. dissertation, Stanford Univ., Stanford, CA, 2002.

[32] D. Tenne and T. Singh, "The higher order unscented filter," in *Proc. Amer. Control Conf.*, vol. 3, 2003, pp. 2441–2446.

[33] S. J. Julier and J. K. Uhlmann, "The scaled unscented transformation," in *Proc. Amer. Control Conf.*, 2002, pp. 4555–4559.

[34] S. F. Schmidt, "Applications of state space methods to navigation problems," in *Advanced Control Systems*, C. T. Leondes, Ed. New York: Academic, 1966, vol. 3, pp. 293–340.

[35] C. B. Chang, R. H. Whiting, and M. Athans, "On the state and parameter estimation for maneuvering reentry vehicles," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 99–105, Feb. 1977.

[36] L. Klein, *Sensor and Data Fusion Concepts and Applications*, 2nd ed. Bellingham, WA: SPIE, 1965.

[37] S. J. Julier, "Comprehensive process models for high-speed navigation," Ph.D. dissertation, Univ. Oxford, Oxford, U.K., Oct. 1997.

[38] S. Clark, "Autonomous land vehicle navigation using millimeter wave radar," Ph.D. dissertation, Univ. Sydney, Sydney, NSW, Australia, 1999.

[39] A. Montobbio, "Sperimentazione ed affinamento di un localizzatore," B.S. thesis, Politecnico di Torino, Torino, Italy, 1998.

[40] X. Wang, C. H. Bishop, and S. J. Julier, "What's better, an ensemble of positive/negative pairs or a centered simplex ensemble?," presented at the EGS-AGU-EUG Meeting, Nice, France, 2003.

[41] R. Smith, "Navigation of an underwater remote operated vehicle," Univ. Oxford, Oxford, U.K., 1995.

[42] T. S. Schei, "A finite difference approach to linearization in nonlinear estimation algorithms," in *Proc. Amer. Control Conf.*, vol. 1, 1995, pp. 114–118.

[43] R. L. Bellaire, E. W. Kamen, and S. M. Zabin, "A new nonlinear iterated filter with applications to target tracking," in *Proc. AeroSense: 8th Int. Symp. Aerospace/Defense Sensing, Simulation and Controls*, vol. 2561, 1995, pp. 240–251.

[44] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," in *Proc. AeroSense: 11th Int. Symp. Aerospace/Defense Sensing, Simulation and Controls*, 1997, pp. 182–193.

[45] S. J. Julier, "A skewed approach to filtering," in *Proc. AeroSense: 12th Int. Symp. Aerospace/Defense Sensing, Simulation and Controls*, vol. 3373, Apr. 1998, pp. 54–65.

[46] R. Merwe and E. Wan, "The square-root unscented Kalman filter for state and parameter-estimation," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 6, 2001, pp. 3461–3464.

[47] ——, "Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 6, 2003, pp. VI-701–VI-704.

[48] R. Todling and S. E. Cohn, "Suboptimal schemes for atmospheric data assimilation based on the Kalman filter," *Montly Weather Rev.*, vol. 122, pp. 2530–2557, Nov. 1994.

[49] S. J. Julier and J. K. Uhlmann, "Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations," in *Proc. Amer. Control Conf.*, 2002, pp. 887–892.

[50] S. Julier, "The spherical simplex unscented transformation," in *The Proc. Amer. Control Conf.*, vol. 3, 2003, pp. 2430–2434.

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.

[51] S. J. Julier and J. K. Uhlmann, "A consistent, debiased method for converting between polar and Cartesian coordinate systems," in *The Proc. AeroSense: Acquisition, Tracking and Pointing XI*, vol. 3086, 1997, pp. 110–121.

[52] J. K. Uhlmann. (1995) A real time algorithm for simultaneous map building and localization.

Simon J. Julier (Member, IEEE) received the M.Eng. degree (first-class honors) in engineering, economics, and management and the D.Phil. degree in robotics from the University of Oxford, Oxford, U.K., in 1993 and 1997, respectively.

His principal research interests include distributed data fusion, vehicle localization and tracking, and user interfaces for mobile augmented reality systems.

Jeffrey K. Uhlmann (Member, IEEE) received the B.A. degree in philosophy and the M.Sc. degree in computer science from the University of Missouri–Columbia, Columbia, in 1986 and 1987, respectively, and the doctorate degree in robotics from Oxford University, Oxford, U.K., in 1995.

From 1987 to 2000, he was with the Naval Research Laboratory, Washington, DC, working in the areas of control, tracking, and virtual reality. He was a leading researcher in the development of multiple-target tracking systems for the Strategic Defense Initiative (SDI). He is currently on the faculty of the Department of Computer Science, University of Missouri, Columbia. His current research interests include data fusion, entertainment engineering, spatial data structures, and graphical rendering and processing.

Prof. Uhlmann has received over a dozen invention and publication awards from the U.S. Navy.

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on July 21,2021 at 21:09:34 UTC from IEEE Xplore. Restrictions apply.