A New Extension of the Kalman Filter to Nonlinear Systems

Simon J. Julier siju@robots.ox.ac.uk
Jeffrey K. Uhlmann uhlmann@robots.ox.ac.uk
The Robotics Research Group, Department of Engineering Science, The University of Oxford
Oxford, OX1 3PJ, UK, Phone: +44-1865-282180, Fax: +44-1865-273908

ABSTRACT

The Kalman filter(KF) is one of the most widely used methods for tracking and estimation due to its simplicity, optimality, tractability and robustness. However, the application of the KF to nonlinear systems can be difficult. The most common approach is to use the Extended Kalman Filter (EKF) which simply linearises all nonlinear models so that the traditional linear Kalman filter can be applied. Although the EKF (in its many forms) is a widely used filtering strategy, over thirty years of experience with it has led to a general consensus within the tracking and control community that it is difficult to implement, difficult to tune, and only reliable for systems which are almost linear on the time scale of the update intervals.

In this paper a new linear estimator is developed and demonstrated. Using the principle that a set of discretely sampled points can be used to parameterise mean and covariance, the estimator yields performance equivalent to the KF for linear systems yet generalises elegantly to nonlinear systems without the linearisation steps required by the EKF. We show analytically that the expected performance of the new approach is superior to that of the EKF and, in fact, is directly comparable to that of the second order Gauss filter. The method is not restricted to assuming that the distributions of noise sources are Gaussian. We argue that the ease of implementation and more accurate estimation features of the new filter recommend its use over the EKF in virtually all applications.

Keywords: Navigation, estimation, non-linear systems, Kalman filtering, sampling.

1 INTRODUCTION

Filtering and estimation are two of the most pervasive tools of engineering. Whenever the state of a system must be estimated from noisy sensor information, some kind of state estimator is employed to fuse the data from different sensors together to produce an accurate estimate of the true system state. When the system dynamics and observation models are linear, the minimum mean squared error (MMSE) estimate may be computed using the *Kalman filter*. However, in most applications of interest the system dynamics and observation equations are nonlinear and suitable extensions to the Kalman filter have been sought. It is well-known that the optimal solution to the nonlinear filtering problem requires that a complete description of the conditional probability density is maintained[14]. Unfortunately this exact description requires a potentially unbounded number of parameters and a number of suboptimal approximations have been proposed[6–8, 13, 16, 21].

Probably the most widely used estimator for nonlinear systems is the extended Kalman filter (EKF)[20, 22]. The EKF applies the Kalman filter to nonlinear systems by simply linearising all the nonlinear models so that the traditional linear Kalman filter equations can be applied. However, in practice, the use of the EKF has two well-known drawbacks:

1. Linearisation can produce highly unstable filters if the assumptions of local linearity is violated.
2. The derivation of the Jacobian matrices are nontrivial in most applications and often lead to significant implementation difficulties.

In this paper we derive a new linear estimator which yields performance equivalent to the Kalman filter for linear systems, yet generalises elegantly to nonlinear systems without the linearisation steps required by the EKF. The fundamental component of this filter is the *unscented transformation* which uses a set of appropriately chosen weighted points to parameterise the means and covariances of probability distributions. We argue that the expected performance of the new approach is superior to that of the EKF and, in fact, is directly comparable to that of the second order Gauss filter. Further, the nature of the transform is such that the process and observation models can be treated as "black boxes". It is not necessary to calculate Jacobians and so the algorithm has superior implementation properties to the EKF. We demonstrate the differences in performance in an example application, and we argue that the ease of implementation and more accurate estimation features of the new filter recommend its use over the EKF in virtually all applications.

The structure of this paper is as follows. In Section 2 we describe the problem statement for applying a Kalman filter to nonlinear systems. We argue that the principle problem is the ability to predict the state of the system. Section 3 introduces the unscented transformation. Its properties are analysed and a full filtering algorithm, which includes the effects of process noise, is developed. In Section 4 an example is presented. Using realistic data, the comparison of the unscented filter and EKF for the tracking of a reentry body is considered. Conclusions are drawn in Section 5. A companion paper[10], extends the basic method and shows that judiciously selecting additional points can lead to any desired level of accuracy for any given prior distribution.

2 ESTIMATION IN NONLINEAR SYSTEMS

2.1 Problem Statement

We wish to apply a Kalman filter to a nonlinear discrete time system of the form

$$\mathbf{x}(k+1) = \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), \mathbf{v}(k), k], \tag{1}$$

$$\mathbf{z}(k) = \mathbf{h}[\mathbf{x}(k), \mathbf{u}(k), k] + \mathbf{w}(k), \tag{2}$$

where $\mathbf{x}(k)$ is the n-dimensional state of the system at timestep k, $\mathbf{u}(k)$ is the input vector, $\mathbf{v}(k)$ is the q-dimensional state noise process vector due to disturbances and modelling errors, $\mathbf{z}(k)$ is the observation vector and $\mathbf{w}(k)$ is the measurement noise. It is assumed that the noise vectors $\mathbf{v}(k)$ and $\mathbf{w}(k)$, are zero-mean and

$$\mathrm{E}\left[\mathbf{v}(i)\mathbf{v}^T(j)\right] = \delta_{ij}\mathbf{Q}(i), \quad \mathrm{E}\left[\mathbf{w}(i)\mathbf{w}^T(j)\right] = \delta_{ij}\mathbf{R}(i), \quad \mathrm{E}\left[\mathbf{v}(i)\mathbf{w}^T(j)\right] = \mathbf{0}, \quad \forall i, j.$$

The Kalman filter propagates the first two moments of the distribution of $\mathbf{x}(k)$ recursively and has a distinctive "predictor-corrector" structure. Let $\hat{\mathbf{x}}(i \mid j)$ be the estimate of $\mathbf{x}(i)$ using the observation information information up to and including time j, $\mathbf{Z}^j = [\mathbf{z}(1), \ldots, \mathbf{z}(j)]$. The covariance of this estimate is $\mathbf{P}(i \mid j)$. Given an estimate $\hat{\mathbf{x}}(k \mid k)$, the filter first predicts what the future state of the system will be using the process model. Ideally, the predicted quantities are given by the expectations

$$\hat{\mathbf{x}}(k+1 \mid k) = \mathrm{E}\left[\mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), \mathbf{v}(k), k] \mid \mathbf{Z}^k\right] \tag{3}$$

$$\mathbf{P}(k+1 \mid k) = \mathrm{E}\left[\{\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1 \mid k)\}\{\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1 \mid k)\}^T \mid \mathbf{Z}^k\right]. \tag{4}$$

When $\mathbf{f}[\cdot]$ and $\mathbf{h}[\cdot]$ are nonlinear, the precise values of these statistics can only be calculated if the distribution of $\mathbf{x}(k)$, condition on $\mathbf{Z}^k$, is known. However, this distribution has no general form and a potentially unbounded number of parameters are required. In many applications, the distribution of $\mathbf{x}(k)$ is approximated so that only a finite and tractable number of parameters need be propagated. It is conventionally assumed that the distribution of $\mathbf{x}(k)$ is Gaussian for two reasons. First, the distribution is completely parameterised by just the mean and covariance. Second, given that only the first two moments are known, the Gaussian distribution is the least informative[3].

The estimate $\hat{\mathbf{x}}(k+1 \mid k+1)$ is given by updating the prediction with the current sensor measurement. In the Kalman filter a linear update rule is specified and the weights are chosen to minimise the mean squared error of the estimate. The update rule is

$$\begin{aligned}
\hat{\mathbf{x}}(k+1 \mid k+1) &= \hat{\mathbf{x}}(k+1 \mid k) + \mathbf{W}(k+1)\,\nu(k+1), \\
\mathbf{P}(k+1 \mid k+1) &= \mathbf{P}(k+1 \mid k) - \mathbf{W}(k+1)\,\mathbf{P}_{\nu\nu}(k+1 \mid k)\,\mathbf{W}^T(k+1) \\
\nu(k+1) &= \mathbf{z}(k+1) - \hat{\mathbf{z}}(k+1 \mid k) \\
\mathbf{W}(k+1) &= \mathbf{P}_{x\nu}(k+1 \mid k)\,\mathbf{P}_{\nu\nu}^{-1}(k+1 \mid k).
\end{aligned}$$

It is important to note that these equations are only a function of the predicted values of the first two moments of $\mathbf{x}(k)$ and $\mathbf{z}(k)$. Therefore, the problem of applying the Kalman filter to a nonlinear system is the ability to predict the first two moments of $\mathbf{x}(k)$ and $\mathbf{z}(k)$. This problem is a specific case of a general problem — to be able to calculate the statistics of a random variable which has undergone a nonlinear transformation.

2.2 The Transformation of Uncertainty

The problem of predicting the future state or observation of the system can be expressed in the following form. Suppose that $\mathbf{x}$ is a random variable with mean $\bar{\mathbf{x}}$ and covariance $\mathbf{P}_{xx}$. A second random variable, $\mathbf{y}$ is related to $\mathbf{x}$ through the nonlinear function

$$\mathbf{y} = \mathbf{f}[\mathbf{x}]. \tag{5}$$

We wish to calculate the mean $\bar{\mathbf{y}}$ and covariance $\mathbf{P}_{yy}$ of $\mathbf{y}$.

The statistics of $\mathbf{y}$ are calculated by (i) determining the density function of the transformed distribution and (ii) evaluating the statistics from that distribution. In some special cases (for example when $\mathbf{f}[\cdot]$ is linear) exact, closed form solutions exist. However, such solutions do not exist in general and approximate methods must be used. In this paper we advocate that the method should yield *consistent* statistics. Ideally, these should be *efficient* and *unbiased*.

The transformed statistics are consistent if the inequality

$$\mathbf{P}_{yy} - \mathrm{E}\left[\{\mathbf{y} - \bar{\mathbf{y}}\}\{\mathbf{y} - \bar{\mathbf{y}}\}^T\right] \geq \mathbf{0} \tag{6}$$

holds. This condition is extremely important for the validity of the transformation method. If the statistics are not consistent, the value of $\mathbf{P}_{yy}$ is *under*-estimated. If a Kalman filter uses the inconsistent set of statistics, it will place too much weight on the information and under estimate the covariance, raising the possibility that the filter will diverge. By ensuring that the transformation is consistent, the filter is guaranteed to be consistent as well. However, consistency does not necessary imply usefulness because the calculated value of $\mathbf{P}_{yy}$ might be greatly in excess of the actual mean squared error. It is desirable that the transformation is efficient — the value of the left hand side of Equation 6 should be minimised. Finally, it is desirable that the estimate is *unbiased* or $\bar{\mathbf{y}} \approx \mathrm{E}[\mathbf{y}]$.

The problem of developing a consistent, efficient and unbiased transformation procedure can be examined by considering the Taylor series expansion of Equation 5 about $\bar{\mathbf{x}}$. This series can be expressed (using rather informal

notation) as:

$$\begin{aligned}\mathbf{f}\left[\mathbf{x}\right] &= \mathbf{f}\left[\bar{\mathbf{x}} + \boldsymbol{\delta}\mathbf{x}\right] \\ &= \mathbf{f}\left[\bar{\mathbf{x}}\right] + \boldsymbol{\nabla}\mathbf{f}\boldsymbol{\delta}\mathbf{x} + \frac{1}{2}\boldsymbol{\nabla}^2\mathbf{f}\boldsymbol{\delta}\mathbf{x}^2 + \frac{1}{3!}\boldsymbol{\nabla}^3\mathbf{f}\boldsymbol{\delta}\mathbf{x}^3 + \frac{1}{4!}\boldsymbol{\nabla}^4\mathbf{f}\boldsymbol{\delta}\mathbf{x}^4 + \cdots\end{aligned} \tag{7}$$

where $\boldsymbol{\delta}\mathbf{x}$ is a zero mean Gaussian variable with covariance $\mathbf{P}_{xx}$, and $\boldsymbol{\nabla}^n\mathbf{f}\boldsymbol{\delta}\mathbf{x}^n$ is the appropriate nth order term in the multidimensional Taylor Series. Taking expectations, it can be shown that the transformed mean and covariance are

$$\bar{\mathbf{y}} = \mathbf{f}\left[\bar{\mathbf{x}}\right] + \frac{1}{2}\boldsymbol{\nabla}^2\mathbf{f}\,\mathbf{P}_{xx} + \frac{1}{2}\boldsymbol{\nabla}^4\mathbf{f}\,\mathrm{E}\left[\boldsymbol{\delta}\mathbf{x}^4\right] + \cdots \tag{8}$$

$$\begin{aligned}\mathbf{P}_{yy} = \boldsymbol{\nabla}\mathbf{f}\,\mathbf{P}_{xx}(\boldsymbol{\nabla}\mathbf{f})^T + \frac{1}{2\times 4!}\boldsymbol{\nabla}^2\mathbf{f}\left(\mathrm{E}\left[\boldsymbol{\delta}\mathbf{x}^4\right] - \mathrm{E}\left[\boldsymbol{\delta}\mathbf{x}^2\mathbf{P}_{yy}\right] - \mathrm{E}\left[\mathbf{P}_{yy}\boldsymbol{\delta}\mathbf{x}^2\right] + \mathbf{P}_{yy}^2\right)(\boldsymbol{\nabla}^2\mathbf{f})^T + \\ \frac{1}{3!}\boldsymbol{\nabla}^3\mathbf{f}\mathrm{E}\left[\boldsymbol{\delta}\mathbf{x}^4\right](\boldsymbol{\nabla}\mathbf{f})^T + \cdots.\end{aligned} \tag{9}$$

In other words, the nth order term in the series for $\bar{\mathbf{x}}$ is a function of the nth order moments of $\mathbf{x}$ multiplied by the nth order derivatives of $\mathbf{f}\left[\cdot\right]$ evaluated at $\mathbf{x} = \bar{\mathbf{x}}$. If the moments and derivatives can be evaluated correctly up to the nth order, the mean is correct up to the nth order as well. Similar comments hold for the covariance equation as well, although the structure of each term is more complicated. Since each term in the series is scaled by a progressively smaller and smaller term, the lowest order terms in the series are likely to have the greatest impact. Therefore, the prediction procedure should be concentrated on evaluating the lower order terms.

Linearisation assumes that the second and higher order terms of $\boldsymbol{\delta}\mathbf{x}$ in Equation 7 can be neglected. Under this assumption,

$$\bar{\mathbf{y}} = \mathbf{f}\left[\bar{\mathbf{x}}\right], \tag{10}$$

$$\mathbf{P}_{yy} = \boldsymbol{\nabla}\mathbf{f}\;\mathbf{P}_{xx}\left(\boldsymbol{\nabla}\mathbf{f}\right)^T. \tag{11}$$

Comparing these expressions with Equations 8 and 9, it is clear that these approximations are accurate only if the second and higher order terms in the mean and fourth and higher order terms in the covariance are negligible. However, in many practical situations linearisation introduces significant biases or errors. An extremely common and important problem is the transformation of information between polar and Cartesian coordinate systems[10, 15]. This is demonstrated by the simple example given in the next subsection.

2.3 Example

Suppose a mobile robot detects beacons in its environment using a range-optimised sonar sensor. The sensor returns polar information (range r and bearing θ) and this is to be converted to estimate to Cartesian coordinates. The transformation is:

$$\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}r\cos\theta\\r\sin\theta\end{pmatrix} \text{ with } \boldsymbol{\nabla}\mathbf{f} = \begin{bmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{bmatrix}.$$

The real location of the target is $(0, 1)$. The difficulty with this transformation arises from the physical properties of the sonar. Fairly good range accuracy (with 2cm standard deviation) is traded off to give a very poor bearing measurement (standard deviation of 15°). The large bearing uncertainty causes the assumption of local linearity to be violated.

To appreciate the errors which can be caused by linearisation, its values of the statistics of (x, y) were compared with those calculated by the true statistics which are calculated by Monte Carlo simulation. Due to the slow convergence of random sampling methods, an extremely large number of samples (3.5×10^6) were used to ensure that accurate estimates of the true statistics were obtained. The results are shown in Figure 1. This figure shows the mean and 1σ contours for which are 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 $\mathbf{P}_{yy}$. As can be seen, the linearised transformation is biased and inconsistent. This is most pronounced in the range direction, where linearisation estimates that the position is 1m whereas in reality it is 96.7cm. This is extremely substantial. Linearisation errors effectively introduce an error which is over 1.5 times the standard deviation of the range measurement. Since it is a bias which arises from the transformation process itself, the same error with the same sign will be committed each time a coordinate transformation takes place. Even if there were no bias, the transformation is inconsistent. Its ellipse is not long enough in the r direction. In fact, the nature of the inconsistency compounds the problem of the biased-ness: not only is the estimate or r in error, but also its estimated mean squared error is much smaller than the true value.

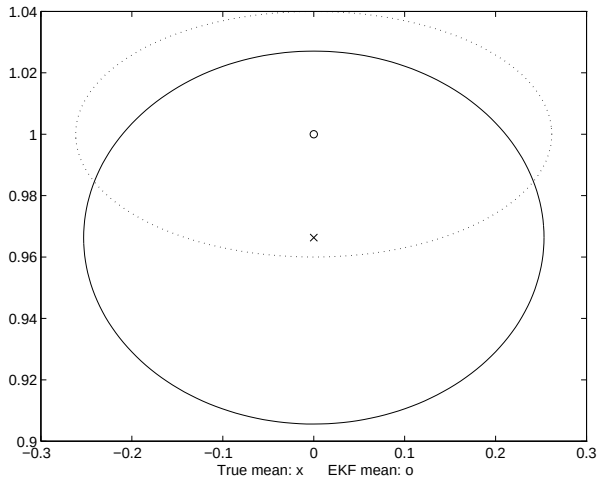


Figure 1: The mean and standard deviation ellipses for the actual and linearised form of the transformation. The true mean is at $\times$ and the uncertainty ellipse is solid. Linearisation calculates the mean at $\circ$ and the uncertainty ellipse is dashed.

In practice the inconsistency can be resolved by introducing additional stabilising noise which increases the size of the transformed covariance. This is one possible of why EKFs are so difficult to tune — sufficient noise must be introduced to offset the defects of linearisation. However, introducing stabilising noise is an undesirable solution since the estimate remains biased and there is no general guarantee that the transformed estimate remains consistent or efficient. A more accurate prediction algorithm is required.

3 THE UNSCENTED TRANSFORM

3.1 The Basic Idea

The unscented transformation is a new, novel method for calculating the statistics of a random variable which undergoes a nonlinear transformation. It is founded on the intuition that *it is easier to approximate a Gaussian distribution than it is to approximate an arbitrary nonlinear function or transformation*[23]. The approach is illustrated in Figure 2. A set of points (or *sigma points*) are chosen so that their sample mean and sample covariance are $\bar{\mathbf{x}}$ and $\mathbf{P}_{xx}$. The nonlinear function is applied to each point in turn to yield a cloud of transformed points and $\bar{\mathbf{y}}$ and $\mathbf{P}_{yy}$ are the statistics of the transformed points. Although this method bares a superficial resemblance to Monte

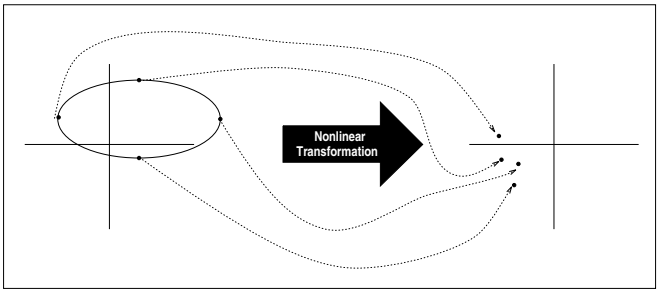


Figure 2: The principle of the unscented transform.

Carlo-type methods, there is an extremely important and fundamental difference. The samples are *not* drawn at random but rather according to a specific, deterministic algorithm. Since the problems of statistical convergence are not an issue, high order information about the distribution can be captured using only a very small number of points.

The n-dimensional random variable $\mathbf{x}$ with mean $\bar{\mathbf{x}}$ and covariance $\mathbf{P}_{xx}$ is approximated by $2n+1$ weighted points given by

$$\begin{array}{lll} \boldsymbol{\mathcal{X}}_0 & = \bar{\mathbf{x}} & W_0 = \kappa/(n+\kappa) \\ \boldsymbol{\mathcal{X}}_i & = \bar{\mathbf{x}} + \left(\sqrt{(n+\kappa)\mathbf{P}_{xx}}\right)_i & W_i = 1/2(n+\kappa) \\ \boldsymbol{\mathcal{X}}_{i+n} & = \bar{\mathbf{x}} - \left(\sqrt{(n+\kappa)\mathbf{P}_{xx}}\right)_i & W_{i+n} = 1/2(n+\kappa) \end{array} \tag{12}$$

where $\kappa \in \Re$, $\left(\sqrt{(n+\kappa)\mathbf{P}_{xx}}\right)_i$ is the ith row or column of the matrix square root of $(n+\kappa)\mathbf{P}_{xx}$ and W_i is the weight which is associated with the ith point. The transformation procedure is as follows:

1. Instantiate each point through the function to yield the set of transformed sigma points,
$$\boldsymbol{\mathcal{Y}}_i = \mathbf{f}\left[\boldsymbol{\mathcal{X}}_i\right].$$
2. The mean is given by the weighted average of the transformed points,
$$\bar{\mathbf{y}} = \sum_{i=0}^{2n} W_i \boldsymbol{\mathcal{Y}}_i. \tag{13}$$
3. The covariance is the weighted outer product of the transformed points,
$$\mathbf{P}_{yy} = \sum_{i=0}^{2n} W_i \left\{\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}}\right\}\left\{\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}}\right\}^T. \tag{14}$$

The properties of this algorithm have been studied in detail elsewhere[9, 12] and we present a summary of the results here:

1. Since the mean and covariance of $\mathbf{x}$ are captured precisely up to the second order, the calculated values of the mean and covariance of $\mathbf{y}$ are correct to the second order as well. This means that the mean is calculated to a higher order of accuracy than the EKF, whereas the covariance is calculated to the same order of accuracy. However, there are further performance benefits. Since the distribution of $\mathbf{x}$ is being approximated rather than $\mathbf{f}[\cdot]$, its series expansion is not truncated at a particular order. It can be shown that the unscented algorithm is able to partially incorporate information from the higher orders, leading to even greater accuracy.
2. The sigma points capture the same mean and covariance irrespective of the choice of matrix square root which is used. Numerically efficient and stable methods such as the Cholesky decomposition[18] can be used.
3. The mean and covariance are calculated using standard vector and matrix operations. This means that the algorithm is suitable for *any* choice of process model, and implementation is extremely rapid because it is not necessary to evaluate the Jacobians which are needed in an EKF.
4. κ provides an extra degree of freedom to "fine tune" the higher order moments of the approximation, and can be used to reduce the overall prediction errors. When $\mathbf{x}(k)$ is assumed Gaussian, a useful heuristic is to select $n+\kappa = 3$. If a different distribution is assumed for $\mathbf{x}(k)$ then a different choice of κ might be more appropriate.

5. Although κ can be positive or negative, a negative choice of κ can lead to a non-positive semidefinite estimate of $\mathbf{P}_{yy}$. This problem is not uncommon for methods which approximate higher order moments or probability density distributions[8, 16, 21]. In this situation, it is possible to use a *modified* form of the prediction algorithm. The mean is still calculated as before, but the "covariance" is evaluated about $\boldsymbol{\mathcal{X}}_0 (k+1 \mid k)$. It can be shown that the modified form ensures positive semi-definiteness and, in the limit as $(n+\kappa) \to 0$,

$$\lim_{(n+\kappa)\to 0} \bar{\mathbf{y}} = \mathbf{f}[\bar{\mathbf{x}}] + \frac{1}{2}\boldsymbol{\nabla}^2\mathbf{f}\,\mathbf{P}_{xx}, \lim_{(n+\kappa)\to 0} \mathbf{P}_{yy} = \boldsymbol{\nabla}\mathbf{f}\,\mathbf{P}_{xx}\,(\boldsymbol{\nabla}\mathbf{f})^T.$$

In other words, the algorithm can be made to perform exactly like the second Order Gauss Filter, but without the need to calculate Jacobians or Hessians.

The performance benefits of using the unscented transform can be seen in Figure 3 which shows the means and 1σ contours determined by the different methods. The true mean lies at $\times$ with a dotted covariance contour. The position of the unscented mean is indicated by a $\star$ and its contour is solid. The linearised mean is at $\circ$ and used a dashed contour. As can be seen the unscented mean value is the same as the true value — on the scale of the graph, the two points lie on top of one another. Further, the unscented transform is consistent — in fact, its contour is slightly larger than the true contour in the r direction.

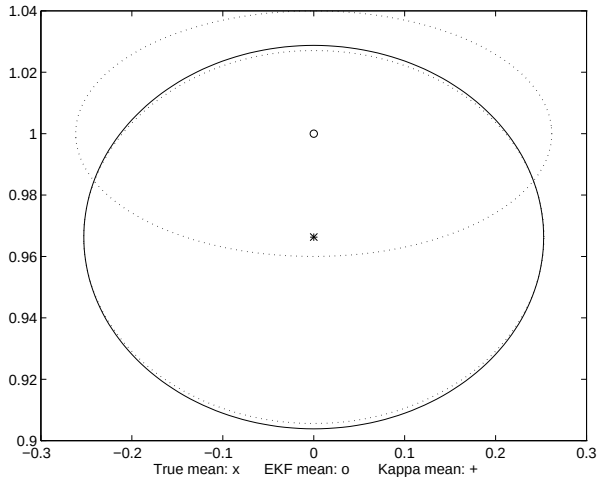


Figure 3: The unscented transform as applied to the measurement example.

Given its properties of superior estimation accuracy and ease of implementation, the unscented transform is better suited than linearisation for filtering applications. Indeed, since it can predict the mean and covariance with second order accuracy, any filter which uses the unscented transform will have the same performance as the Truncated Second Order Gauss Filter[1] but does not require the derivation of Jacobians or Hessians. The next subsection examines the application of the unscented transform to the filtering problem and develops the *unscented filter*.

3.2 The Unscented Filter

The transformation processes which occur in a Kalman filter consist of the following steps:

- Predict the new state of the system $\hat{\mathbf{x}}(k+1 \mid k)$ and its associated covariance $\mathbf{P}(k+1 \mid k)$. This prediction must take account of the effects of process noise.
- Predict the expected observation $\hat{\mathbf{z}}(k+1 \mid k)$ and the innovation covariance $\mathbf{P}_{\nu\nu}(k+1 \mid k)$. This prediction should include the effects of observation noise.
- Finally, predict the cross-correlation matrix $\mathbf{P}_{xz}(k+1 \mid k)$.

These steps can be easily accommodated by slightly restructuring the state vector and process and observation models. First, the state vector is augmented with the process and noise terms to give an $n^a = n + q$ dimensional

1. The set of sigma points are created by applying Equation 12 to the augmented system given by Equation 15.
2. The transformed set is given by instantiating each point through the process model,

$$\boldsymbol{\mathcal{X}}_i\left(k+1 \mid k\right)=\mathbf{f}\left[\boldsymbol{\mathcal{X}}_i^a\left(k \mid k\right), \mathbf{u}\left(k\right), k\right].$$

3. The predicted mean is computed as

$$\hat{\mathbf{x}}\left(k+1 \mid k\right)=\sum_{i=0}^{2n^a} W_i \boldsymbol{\mathcal{X}}_i^a\left(k+1 \mid k\right).$$

4. And the predicted covariance is computed as

$$\mathbf{P}\left(k+1 \mid k\right)\sum_{i=0}^{2n^a} W_i\left\{\boldsymbol{\mathcal{X}}_i\left(k+1 \mid k\right)-\hat{\mathbf{x}}\left(k+1 \mid k\right)\right\}\left\{\boldsymbol{\mathcal{X}}_i\left(k+1 \mid k\right)-\hat{\mathbf{x}}\left(k+1 \mid k\right)\right\}^T$$

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

$$\boldsymbol{\mathcal{Z}}_i\left(k+1 \mid k\right)=\mathbf{h}\left[\boldsymbol{\mathcal{X}}_i\left(k+1 \mid k\right), \mathbf{u}\left(k\right), k\right]$$

6. The predicted observation is calculated by

$$\hat{\mathbf{z}}\left(k+1 \mid k\right)=\sum_{i=1}^{2n^a} W_i \boldsymbol{\mathcal{Z}}_i\left(k+1 \mid k\right).$$

7. Since the observation noise is additive and independent, the innovation covariance is

$$\mathbf{P}_{\nu\nu}\left(k+1 \mid k\right)=\mathbf{R}\left(k+1\right)+\sum_{i=0}^{2n^a} W_i\left\{\boldsymbol{\mathcal{Z}}_i\left(k \mid k-1\right)-\hat{\mathbf{z}}\left(k+1 \mid k\right)\right\}\left\{\boldsymbol{\mathcal{Z}}_i\left(k \mid k-1\right)-\hat{\mathbf{z}}\left(k+1 \mid k\right)\right\}^T$$

8. Finally the cross correlation matrix is determined by

$$\mathbf{P}_{xz}\left(k+1 \mid k\right)=\sum_{i=0}^{2n^a} W_i\left\{\boldsymbol{\mathcal{X}}_i\left(k \mid k-1\right)-\hat{\mathbf{x}}\left(k+1 \mid k\right)\right\}\left\{\boldsymbol{\mathcal{Z}}_i\left(k \mid k-1\right)-\hat{\mathbf{z}}\left(k+1 \mid k\right)\right\}^T$$

Box 3.1: The prediction algorithm using the unscented transform.

vector,

$$\mathbf{x}^a\left(k\right)=\begin{bmatrix}\mathbf{x}\left(k\right)\\ \mathbf{v}(k)\end{bmatrix}.$$

The process model is rewritten as a function of $\mathbf{x}^a\left(k\right)$,

$$\mathbf{x}\left(k+1\right)=\mathbf{f}\left[\mathbf{x}^a\left(k\right), \mathbf{u}\left(k\right), k\right]$$

and the unscented transform uses $2n^a+1$ sigma points which are drawn from

$$\hat{\mathbf{x}}^a\left(k \mid k\right)=\begin{pmatrix}\hat{\mathbf{x}}\left(k \mid k\right)\\ \mathbf{0}_{q\times 1}\end{pmatrix} \text{ and } \mathbf{P}^a\left(k \mid k\right)=\begin{bmatrix}\mathbf{P}\left(k \mid k\right) & \mathbf{P}_{xv}\left(k \mid k\right)\\ \mathbf{P}_{xv}\left(k \mid k\right) & \mathbf{Q}\left(k\right)\end{bmatrix}. \tag{15}$$

The matrices on the leading diagonal are the covariances and off-diagonal sub-blocks are the correlations between the state errors and the process noises. Although this method requires the use of additional sigma points, it means that the effects of the process noise (in terms of its impact on the mean and covariance) are

introduced with the same order of accuracy as the uncertainty in the state. The formulation also means that correlated noise sources (which can arise in Schmidt-Kalman filters[19]) can be implemented extremely easily. The expression for the unscented transform is given by the equations in Box 3.1.

Various extensions and modifications can be made to this basic method to take account of specific details of a given application. For example, if the observation noise is introduced in a nonlinear fashion, or is correlated with process and/or observation noise, then the augmented vector is expanded to include the observation terms.

This section has developed the unscented transform so that it can be used in filtering and tracking applications. The next section demonstrates its benefits over the EKF for a sample application.

4 EXAMPLE APPLICATION

In this section we consider the problem which is illustrated in Figure 4: a vehicle enters the atmosphere at high altitude and at a very high speed. The position of the body is to be tracked by a radar which accurately measures range and bearing. This type of problem has been identified by a number of authors[1, 2, 5, 17] 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 which act. 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 towards the centre of the earth. The final forces are random buffeting terms. The effect of these forces gives a trajectory of the form shown in Figure 4: 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.

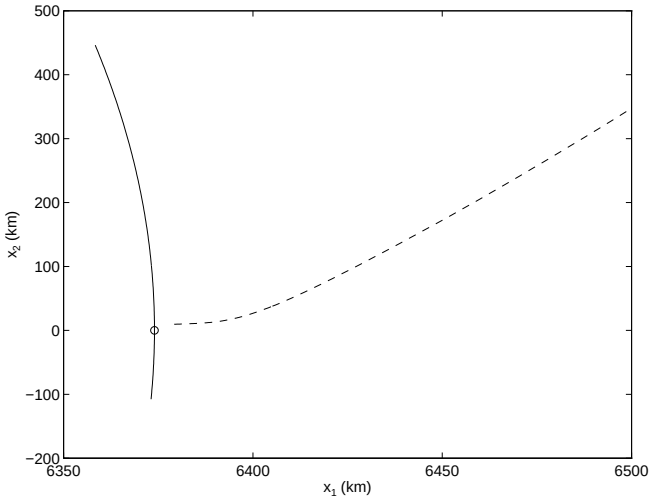


Figure 4: The reentry problem. The dashed line is the sample vehicle trajectory and the solid line is a portion of the Earth's surface. The position of the radar is marked by a ∘.

In summary, the tracking system should be able to track an object which 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

$$\begin{aligned}
\dot{x}_1(k) &= x_3(k) \\
\dot{x}_2(k) &= x_4(k) \\
\dot{x}_3(k) &= D(k)x_3(k) + G(k)x_1(k) + v_1(k) \\
\dot{x}_4(k) &= D(k)x_4(k) + G(k)x_2(k) + v_2(k) \\
\dot{x}_5(k) &= v_3(k)
\end{aligned} \tag{16}$$

where $D(k)$ is the drag-related force term, $G(k)$ is the gravity-related force term and $v.(k)$ are the process noise terms. Defining $R(k) = \sqrt{x_1^2(k) + x_2^2(k)}$ as the distance from the centre of the Earth and $V(k) = \sqrt{x_3^2(k) + x_4^2(k)}$

as absolute vehicle speed then the drag and gravitational terms are

$$D(k) = -\beta(k) \exp\left\{\frac{[R_0 - R(k)]}{H_0}\right\} V(k),\ G(k) = -\frac{Gm_0}{r^3(k)}$$
$$\text{and } \beta(k) = \beta_0 \exp x_5(k).$$

For this example the parameter values are $\beta_0 = -0.59783$, $H_0 = 13.406$, $Gm_0 = 3.9860 \times 10^5$ and $R_0 = 6374$ and reflect typical environmental and vehicle characteristics[2]. The parameterisation of the ballistic coefficient, $\beta(k)$, reflects the uncertainty in vehicle characteristics[5]. 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 which is located at (x_r, y_r). It is able to measure range r and bearing θ at a frequency of 10Hz, 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 1m and 17mrad respectively[4]. 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 that the EKF is able to predict the sensor readings accurately with very little bias.

The true initial conditions for the vehicle are

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

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

The vehicle is buffeted by random accelerations,

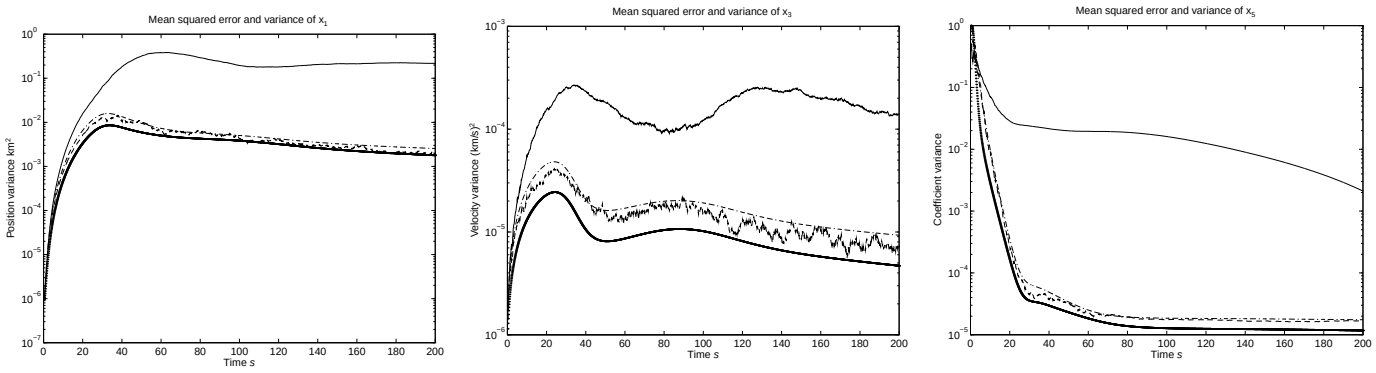
$$\mathbf{Q}(k) = \begin{bmatrix} 2.4064 \times 10^{-5} & 0 & 0 \\ 0 & 2.4064 \times 10^{-5} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The initial conditions assumed by the filter are,

$$\hat{\mathbf{x}}(0 \mid 0) = \begin{pmatrix} 6500.4 \\ 349.14 \\ -1.8093 \\ -6.7967 \\ 0 \end{pmatrix} \text{ and } \mathbf{P}(0 \mid 0) = \begin{bmatrix} 10^{-6} & 0 & 0 & 0 & 0 \\ 0 & 10^{-6} & 0 & 0 & 0 \\ 0 & 0 & 10^{-6} & 0 & 0 \\ 0 & 0 & 0 & 10^{-6} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The filter uses the nominal initial condition and, to offset for the uncertainty, the variance on this initial estimate is 1.

Both filters were implemented in discrete time and observations were taken at a frequency of 10Hz. However, due to the intense nonlinearities of the vehicle dynamics equations, the Euler approximation of Equation 16 was only valid for small time steps. The integration step was set to be 50ms which meant that two predictions were



(a) Results for x_1. **(b) Results for x_3.** **(c) Results for x_5.**

Figure 5: The 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.

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 re-linearise before the second step.

The performance of each filter is shown in Figure 5. This figure plots the estimated mean squared estimation error (the diagonal elements of $\mathbf{P}(k \mid k)$) 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 be confident with the filter estimates. The EKF, however, is highly inconsistent: the peak mean squared error in x_1 is 0.4km^2, whereas its estimated covariance is over one hundred times smaller. Similarly, the peak mean squared velocity error is $3.4 \times 10^{-4}\text{km}^2\text{s}^{-2}$ which is over 5 times the true mean squared error. Finally, it can be seen that x_5 is highly biased, and this bias only slowly decreases over time. This poor performance is the direct result of linearisation errors.

5 CONCLUSIONS

In this paper we have argued that the principle difficulty for applying the Kalman filter to nonlinear systems is the need to consistently predict the new state and observation of the system. We have introduced a new filtering algorithm, called the unscented filter. By virtue of the unscented transformation, this algorithm has two great advantages over the EKF. First, it is able to predict the state of the system more accurately. Second, it is much less difficult to implement. The benefits of the algorithm were demonstrated in a realistic example.

This paper has considered one specific form of the unscented transform for one particular set of assumptions. In a companion paper[11], we extend the development of the unscented transform and yield a general framework for its derivation and application. It is shown that the number of sigma points can be extended to yield a filter which matches moments up to the fourth order. This higher order extension effectively de-biases almost all common nonlinear coordinate transformations.

REFERENCES

[1] M. Athans, R. P. Wishner and A. Bertolini. Suboptimal State Estimation For Continuous-Time Nonlinear Systems From Discrete Noisy Measurements. *IEEE Transactions on Automatic Control*, TAC-13(6):504–518, October 1968.

[2] J. W. Austin and C. T. Leondes. Statistically Linearized Estimation of Reentry Trajectories. *IEEE Transactions on Aerospace and Electronic Systems*, AES-17(1):54–61, January 1981.

[3] D. E. Catlin. Estimation, Control and the Discrete Kalman Filter. In *Applied Mathematical Sciences 71*, page 84. Springer-Verlag, 1989.

[4] C. B. Chang, R. H. Whiting and M. Athans. On the State and Parameter Estimation for Maneuvering Reentry Vehicles. *IEEE Transactions on Automatic Control*, AC-22:99–105, February 1977.

[5] P. J. Costa. Adaptive Model Architecture and Extended Kalman-Bucy Filters. *IEEE Transactions on Aerospace and Electronic Systems*, AES-30(2):525–533, April 1994.

[6] F. E. Daum. New Exact Nonlinear Filters. In J. C. Spall, editor, *Bayesian Analysis of Time Series and Dynamic Models*, chapter 8, pages 199–226. Marcel Drekker, Inc., 1988.

[7] N. J. Gordon, D. J. Salmond and A. F. M. Smith. Novel Approach to Nonlinear/non-Gaussian Bayesian State Estimation. *IEE Proceedings-F*, 140(2):107–113, April 1993.

[8] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[9] S. J. Julier and J. K. Uhlmann. A General Method for Approximating Nonlinear Transformations of Probability Distributions. Can be downloaded from `http://www.robots.ox.ac.uk/~siju`, August 1994.

[10] S. J. Julier and J. K. Uhlmann. A Consistent, Debiased Method for Converting Between Polar and Cartesian Coordinate Systems. In *The Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls, Orlando, Florida*. SPIE, 1997. Acquisition, Tracking and Pointing XI.

[11] S. J. Julier, J. K. Uhlmann and H. F. Durrant-Whyte. A New Approach for Filtering Nonlinear Systems. In *The Proceedings of the American Control Conference, Seattle, Washington.*, pages 1628–1632, 1995.

[12] S. J. Julier, J. K. Uhlmann and H. F. Durrant-Whyte. A New Approach for the Nonlinear Transformation of Means and Covariances in Linear Filters. *IEEE Transactions on Automatic Control*, Accepted for publication as a Technical Note. 1996.

[13] H. J. Kushner. Approximations to Optimal Nonlinear Filters. *IEEE Transactions on Automatic Control*, AC-12(5):546–556, October 1967.

[14] H. J. Kushner. Dynamical Equations For Optimum Non-linear Filtering. *Journal of Differential Equations*, 3:179–190, 1967.

[15] D. Lerro and Y. K. Bar-Shalom. Tracking with Debiased Consistent Converted Measurements vs. EKF. *IEEE Transactions on Aerospace and Electronics Systems*, AES-29(3):1015–1022, July 1993.

[16] P. S. Maybeck. *Stochastic Models, Estimation, and Control*, volume 2. Academic Press, 1982.

[17] R. K. Mehra. A Comparison of Several Nonlinear Filters for Reentry Vehicle Tracking. *IEEE Transactions on Automatic Control*, AC-16(4):307–319, August 1971.

[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2 edition, 1992.

[19] S. F. Schmidt. Applications of state space methods to navigation problems. In C. T. Leondes, editor, *Advanced Control Systems*, volume 3, pages 293–340. Academic Press, 1966.

[20] H. W. Sorenson, editor. *Kalman filtering: theory and application*. IEEE Press, 1985.

[21] H. W. Sorenson and A. R. Stubberud. Non-linear Filtering by Approximation of the *a posteriori* Density. *International Journal of Control*, 8(1):33–51, 1968.

[22] J. K. Uhlmann. Algorithms for multiple target tracking. *American Scientist*, 80(2):128–141, 1992.

[23] J. K. Uhlmann. Simultaneous map building and localization for real time applications. Technical report, University of Oxford, 1994. Transfer thesis.