

# Reduced Sigma Point Filters for the Propagation of Means and Covariances Through Nonlinear Transformations

Simon J. Julier

IDAK Industries  
901 Missouri Blvd., #179

Jefferson City, MO 65109 USA  
sjulier@idak.com

Jeffrey K. Uhlmann

Dept. of Computer Engineering  
and Computer Science  
University of Missouri-Columbia  
Columbia, MO 65211 USA  
uhlmann@cecs.missouri.edu

*Abstract*— The Unscented Transform (UT) approximates the result of applying a specified nonlinear transformation to a given mean and covariance estimate. The UT works by constructing a set of points, referred to as *sigma points*, which has the same known statistics, e.g., first and second and possibly higher moments, as the given estimate. The given nonlinear transformation is applied to the set, and the unscented estimate is obtained by computing the statistics of the transformed set of sigma points. For example, the mean and covariance of the transformed set approximates the nonlinear transformation of the original mean and covariance estimate. The computational efficiency of the UT therefore depends on the number of sigma points required to capture the known statistics of the original estimate. In this paper we examine methods for minimizing the number of sigma points for real-time control, estimation, and filtering applications. We demonstrate results in a 3D localization example.

**Keywords:** Kalman filter, non-linear estimation, unscented filtering.

## I. INTRODUCTION

Many filtering and estimation algorithms need to be run at high sample rates. Head trackers, for example, often need to be run at about 1kHz, and some proposed laser-missile intercept systems require on the order of 1MHz to attain effective focal dwell times. Because of these stringent requirements, rapid computational algorithms are required. Probably the best-known and most widely implemented nonlinear estimator is the extended Kalman filter (EKF) [1, 2]. 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. In so doing, the EKF inherits many of the computational efficiency advantages associated with the linear Kalman filter. Unfortunately, the EKF suffers two well-known problems: First, the required Jacobian matrices, i.e., the matrix/linear approximations of nonlinear functions, can be extremely difficult and error-prone to derive. Second, the EKF linearized approximations can be extremely inaccurate and lead to filter instability [3].

Many different algorithms have been developed to compensate for one or both of the EKF's major limitations. A general class of algorithms, commonly referred to as *Particle filters*, has recently received increased attention. The most common type of particle filter, the *Monte Carlo* filter, uses a finite set of samples from an assumed probability distribution and then transforms each sample independently with an expectation that the collective statistics of the ensemble set of samples will accurately reflect the statistics of the evolving system of interest. Particle filters partially surmount the two major disadvantages of the EKF in that they can apply a "black box" nonlinear transformation to each sample without any need for Jacobian derivations,

and they are provably more accurate than linearization as the number of samples approaches infinity (at least under certain assumptions).

Although sampling methods avoid some difficulties, they introduce others. For example, the number of samples required to achieve reliable estimates can be many thousands in a two or three dimensional problem [4], and many millions as the dimensionality increases, e.g., for a 6-dimensional position and velocity filtering problem. This incurs several orders of magnitude more computational effort than that required by an EKF. In addition to computational expense, it is well-known that the accuracy of Monte Carlo estimates degrades rapidly as a sample set of points is successively transformed. This degradation is due to the fact that even a relatively large number of sample points does not perfectly approximate the underlying distribution, and deviations in the approximation tend to become magnified after successive nonlinear transformations. Additional mechanisms, such as periodic re-sampling, have been used to reduce this problem [4–6], but there still remains a large trade-off between computational expense (number of sample points) and stability.

A radical departure from random sampling came with the advent of the Unscented Transform (UT) [7–10]. The UT works by constructing a set of points, referred to as *sigma points*, which is deterministically constrained to have the same known statistics, e.g., first and second moments, as a given measurement or state estimate. A specified nonlinear transformation can be applied to each sigma point, and the unscented estimate can be obtained by computing the statistics of the transformed set. For example, the mean and covariance of the transformed set approximates the nonlinear transformation of the original mean and covariance estimate. The deterministic component of the UT avoids the random sampling errors introduced by Monte Carlo and other sampling methods and therefore dramatically reduces the number of points required to achieve the same transformation accuracy.

As was discussed in [9], given only an  $n$ -dimensional mean and covariance estimate, with no other error distribution information, a set of  $n + 1$  sigma points can be constructed that fully captures all of the known statistics of the error distribution, i.e., its mean and covariance. This set of sigma points can be regarded as one element of the set of all possible distributions that could possibly underly the given mean and covariance estimate. Given that there is no additional information on which to select one candidate distribution over another, the set of  $n + 1$  sigma points can be used to generate a transformed estimate that cannot in general be improved: There is no other available information to be captured by the inclusion of additional points, and there is no "distribution of distributions" (e.g., as

resulting from additional assumptions such as scale-invariance of the limiting distribution) from which samples can be drawn to further improve the estimate.

Although the  $n + 1$  sigma point UT can be used to optimally approximate the nonlinear transformation of a given mean and nonsingular covariance when no other other assumptions can be made, there are many practical circumstances in which other information is available. For example, virtually all filtering and control applications involve the use of a measuring process that introduces errors that can be empirically characterized to some extent. Thus, the mean, the covariance, and some information about the third central moment (the skew) and the fourth central moment (the kurtosis) may be known. Incorporating this information may require more sigma points, but it can also significantly improve the accuracy of the transformed estimate.

A large class of error processes associated with calibrated measuring devices exhibit symmetries about a set of principal measurement axes, e.g., range and bearing. Such symmetries provide information about the third central moment of the unknown distribution: Specifically, the skew is zero. The minimal set of  $n + 1$  sigma points does not capture this information (its skew is not generally zero), but a symmetric set of  $2n$  sigma points can be trivially generated from the square root of the given covariance matrix to match the first three moments of the assumed symmetric error distribution. The inclusion of another (weighted) sigma point can then be used to capture information about the kurtosis of the distribution, e.g., resulting from an assumption of Gaussianity. The improvement in accuracy of symmetric unscented transformations over linearization is reported in [7–12].

Very recently, methods analogous to the UT approach have been derived from a central difference perspective [13]. These methods have been used to increase the amount of assumed kurtosis information that can be exploited, but with some increase in computational expense. However, even the marginally more efficient  $2n + 1$  symmetric UT is unable to satisfy real time constraints for a variety of extremely high data rate applications. Therefore, in this paper we pursue reduced sigma point UT parameterizations that capture distribution information comparable to that of the  $2n + 1$  symmetric UT, but use computational resources comparable to the minimal  $n + 1$  UT. These parameterizations (which should have analogs for central difference methods) offer high accuracy for high data rate control, estimation, and filtering problems. The structure of this paper is as follows. Section II reviews the unscented transform and describes the symmetric point selection method. Section III describes the simplex set and derives the *minimal skew set*. The performance of this point selection algorithm is studied in Section IV. Conclusions are drawn in Section V.

## II. BACKGROUND

### A. Problem Statement

Let  $\mathbf{x}$  be an  $n$ -dimensional random variable with (a not necessarily Gaussian) probability density function  $p_x(\mathbf{x})$ , mean  $\bar{\mathbf{x}}$  and covariance  $\mathbf{P}_{xx}$ . A second random variable  $\mathbf{y}$  is related to  $\mathbf{x}$  through the nonlinear transformation

$$\mathbf{y} = \mathbf{f}[\mathbf{x}] \quad (1)$$

The objective is to calculate the mean  $\bar{\mathbf{y}}$  and covariance  $\mathbf{P}_{yy}$  of  $\mathbf{y}$ .

### B. The Unscented Transform

The Unscented Transform (UT) builds on the principle that it is easier to approximate a probability distribution than it is

to approximate an arbitrary nonlinear function. A set of  $p + 1$  weighted points  $\mathcal{S} = \{W_i, \mathcal{X}_i\}$  are deterministically so that they obey a condition of the form

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

where  $\mathbf{g}[\cdot, \cdot]$  determines *what* information should be captured about  $\mathbf{x}$ . It is possible to meet this condition and still have some degree of freedom in the choice of the points. This ambiguity can be resolved by assigning a penalty function  $c[\mathcal{S}, p(\mathbf{x})]$  to the different solutions. The purpose of this function is to incorporate features which are desirable, but do not necessarily have to be met. As the value of the penalty function increases, the solution becomes less desirable.

The sigma point set which is used is that which is most desirable and confirms to the necessary conditions. In other words, the sigma points are given by the solution to the equation

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

Given the set of points, each point is instantiated through the nonlinear function,  $\mathcal{Y}_i = \mathbf{f}[\mathcal{X}_i]$ . The appropriate statistics for  $\mathbf{P}_{yy}$  are then calculated from the set  $\{W_i, \mathcal{Y}_i\}$ .

In this paper, we are concerned with the propagation of the first two moments through the nonlinear transformation. Therefore, the constraint equation is

$$\mathbf{g}[\mathcal{S}, p_x(\mathbf{x})] = \begin{bmatrix} \sum_{i=0}^p -1 \\ \sum_{i=0}^p W_i \mathcal{X}_i - \bar{\mathbf{x}} \\ \sum_{i=0}^p W_i \{\mathcal{X}_i - \bar{\mathbf{x}}\} \{\mathcal{X}_i - \bar{\mathbf{x}}\}^T - \mathbf{P}_{xx} \end{bmatrix}$$

and the mean and covariance of the transformed set are

$$\bar{\mathbf{y}} = \sum_{i=0}^p W_i \mathcal{Y}_i \quad (2)$$

$$\mathbf{P}_{yy} = \sum_{i=0}^p W_i \{\mathcal{Y}_i - \bar{\mathbf{y}}\} \{\mathcal{Y}_i - \bar{\mathbf{y}}\}^T. \quad (3)$$

In [7] we examined the following symmetrically-distributed set of points which match the mean and covariance:

$$\begin{aligned} \mathcal{X}_0(k|k) &= \hat{\mathbf{x}}(k|k) \\ \mathcal{X}_i(k|k) &= \hat{\mathbf{x}}(k|k) + \left( \sqrt{(n+\kappa)\mathbf{P}(k|k)} \right)_i \\ \mathcal{X}_{i+n}(k|k) &= \hat{\mathbf{x}}(k|k) - \left( \sqrt{(n+\kappa)\mathbf{P}(k|k)} \right)_i \\ W_0 &= \kappa/(n+\kappa) \\ W_i &= 1/\{2(n+\kappa)\} \\ W_{i+n} &= 1/\{2(n+\kappa)\} \end{aligned} \quad (4)$$

where  $\kappa \in \mathfrak{R}$ ,  $\left( \sqrt{(n+\kappa)\mathbf{P}(k|k)} \right)_i$  is the  $i$ th row or column of the matrix square root of  $(n+\kappa)\mathbf{P}(k|k)$  and  $W_i$  is the weight that is associated with the  $i$ th point.  $\kappa$  scales the third and higher order terms of this set and, if  $(n+\kappa) = 3$ , it is possible to match some of the fourth order terms when  $\mathbf{x}$  is Gaussian [10].

This algorithm has favorable computational properties compared to the EKF<sup>1</sup>. The most significant difference in the computational costs arises from the fact that the EKF requires the calculation of the Jacobian matrix whereas the unscented transform requires  $2n + 1$  function evaluations. Indeed, the computational costs of the UT are directly proportional to the number of sigma points which are used. Therefore, minimizing the number of sigma points minimizes the computational costs. Such considerations are crucial if the process model is expensive to calculate or if real-time performance is required.

<sup>1</sup>Equation 3, for example, uses about the same number as calculations ( $2n^3$ ) as pre- and post-multiplying  $\mathbf{P}_{xx}$  by the Jacobian of  $\mathbf{f}[\cdot]$ . Calculating the matrix square root with the Cholesky Decomposition takes about  $n^3/3$  calculations [14].

### III. THE SIMPLEX SIGMA POINTS

In two dimensions the largest possible affinely independent set of points that can be constructed forms a triangle. The generalization of the triangle to  $n$  dimensions is a *simplex* with  $n+1$  vertices. For example, a 3-dimensional simplex is a tetrahedron. A set of fewer than  $n+1$  points in  $n$  dimensions will always lie on a subspace of less than  $n$  dimensions, so the covariance of the set is singular. For example, any two points in two dimensions are co-linear and therefore have a covariance of rank  $< 2$ , and any three points in three dimensions are co-planar and have a covariance of rank  $< 3$ . This implies that  $n+1$  affinely independent sigma points is necessary and sufficient to form a simplex with a given mean and nonsingular covariance. This can be illustrated by noting that, from the properties of outer products of vectors,

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

When  $\mathbf{P}_{xx}$  is full-rank, at least  $n$  sigma points are required. However, this result explicitly factors out the role of the mean and an additional sigma point is required, giving a total of  $n+1$  points. Any set of  $n+1$  affinely independent points can be translated and linearly transformed to have a given mean and covariance. For example, a canonical simplex with zero mean and identity covariance can be transformed by a desired covariance and then translated by a desired mean. However, the resulting set of points can be reflected across any one of the principal covariance axes without affecting the mean or covariance (the outer produce of each point is unaffected by a sign change on any of its elements), so this provides some degrees of freedom for capturing additional statistics about a distribution. In this paper we examine the use of these degrees of freedom to encode information about the third central moment leading to a *minimal skew* set of simplex points.

#### A. The Minimal Skew Simplex Points

The minimal skew sigma points are chosen to match the first two moments of  $\mathbf{x}$  and to minimize the third order moments (skew). The reason for minimising the skew arises from the observation that since only the first two moments are distributed, no information about the (a)symmetry of the distribution is maintained. Because the distribution can be skewed in any direction the average error will be minimized if the distribution is assumed to be symmetric.

Let  $\mathcal{X}_i^j$  be the  $i$ th sigma point in the set for the  $j$ th dimensional space. It is assumed, without loss of generality<sup>2</sup>, that  $\bar{\mathbf{x}} = \mathbf{0}$  and  $\mathbf{P}_{xx} = \mathbf{I}$  (the  $n \times n$  identity matrix).

First consider the problem of choosing a set of points which capture mean and covariance in a single dimension,  $\mathbf{e}_1$ . Three points are used:  $\mathcal{X}_0^1 = [0]$ ,  $\mathcal{X}_1^1 = [-x_1]$  and  $\mathcal{X}_2^1 = [x_2]$ . The weights for these points are  $W_0, W_1$  and  $W_2$ . From the condition that the means and covariances must be  $\bar{\mathbf{x}}$  and  $\mathbf{P}_{xx}$ ,

$$W_0 + W_1 + W_2 = 1 \quad (5)$$

$$-W_1 x_1 + W_2 x_2 = 0 \quad (6)$$

$$W_1 x_1^2 + W_2 x_2^2 = 1 \quad (7)$$

These conditions are not sufficient to uniquely define a set of points and so the further condition that the skew or third order

<sup>2</sup>A random variable  $\mathbf{z}$  with mean  $\mathbf{0}$  and covariance  $\mathbf{I}$  can be transformed to a random variable  $\mathbf{z}$  with mean  $\bar{\mathbf{z}}$  and covariance  $\mathbf{P}_{zz}$  through the linear transformation  $\mathbf{z} = \bar{\mathbf{z}} + (\sqrt{\mathbf{P}_{zz}})\mathbf{z}$  where  $(\sqrt{\mathbf{P}_{zz}})$  is a matrix square root of  $\mathbf{P}_{zz}$ .

moments are zero is added,

$$S_{111} = -W_1 x_1^3 + W_2 x_2^3 = 0. \quad (8)$$

The solution of Equations 5 to 8 is:

$$x_1 = \frac{1}{2\sqrt{W_1}}, \quad W_1 = \frac{1-W_0}{2}, \quad x_2 = x_1, \quad W_2 = W_1$$

where  $W_0$  is a free parameter whose value affects the fourth and higher moments of the sigma point set. The primary reason for introducing this degree of freedom is that it is exploited in the scaled unscented transform. To extend the set to 2D, points 1 and 2 are translated in the  $\mathbf{e}_2$ -direction by  $-x_3$  and a new point — labelled 3 — is added at  $(0, x_3)$  with weight  $W_3$ . Points 1 and 2 still obey the appropriate conditions in the  $\mathbf{e}_1$ -direction, and it is only necessary to ensure that the normalization as well as the mean and covariance constraints in the  $\mathbf{e}_2$ -direction are upheld:

$$W_0 + 2W_1 + W_3 = 1, \quad (9)$$

$$-2W_1 x_3 + W_3 x_3 = 0, \quad (10)$$

$$2W_1 x_3^2 + W_3 x_3^2 = 1. \quad (11)$$

This set has two further skew terms:

$$S_{222} = -2W_1 x_3^3 + W_3 x_3^3, \quad S_{112} = -2W_1 x_1^2 x_3.$$

The minimal skew sigma point selection algorithm minimizes the skew terms of the extra dimension by an appropriate choice of  $W_3$  and  $x_3$ . Let  $S_{222} = 0$ . The values of  $x_3$  and  $W_3$  are  $W_3 = 2W_1$  and  $x_3 = 1/\sqrt{4W_1}$ . However, with this solution  $S_{112} = -1/4\sqrt{W_1}$  and is clearly non-zero. This reflects the fundamental fact that the sigma points are inherently *asymmetric*.

This principle is applied to extend the sigma points to an arbitrary number of dimensions. Suppose a set of points have been calculated for  $n$  dimensions. The sigma points for  $n+1$  dimensions are

$$\mathcal{X}_i^{n+1} = \begin{cases} \begin{bmatrix} \mathcal{X}_0^n \\ 0 \end{bmatrix} & \text{for } i = 0 \\ \begin{bmatrix} \mathcal{X}_i^n \\ -x_{n+1} \end{bmatrix} & \text{for } i = 1, \dots, n+1 \\ \begin{bmatrix} \mathbf{0}_n \\ x_{n+1} \end{bmatrix} & \text{for } i = n+2 \end{cases}$$

where  $\mathbf{0}_i$  is a vector of  $i$  zeros.

The value of  $x_{n+1}$  and the sequence of weights are calculated from the normalization and moment conditions in the  $(n+1)$ th dimension. These constraints are

$$2W_1 + \sum_{i=3}^{n+1} W_i = 1 - W_0 \quad (12)$$

$$W_{n+1} x_{n+1} - x_{n+1} \left( 2W_1 + \sum_{i=3}^n W_i \right) = 0 \quad (13)$$

$$W_{n+1} x_{n+1}^2 + x_{n+1}^2 \left( 2W_1 + \sum_{i=3}^n W_i \right) = 1 \quad (14)$$

$$W_{n+1} x_{n+1}^3 - x_{n+1}^3 \left( 2W_1 + \sum_{i=3}^n W_i \right) = 0 \quad (15)$$

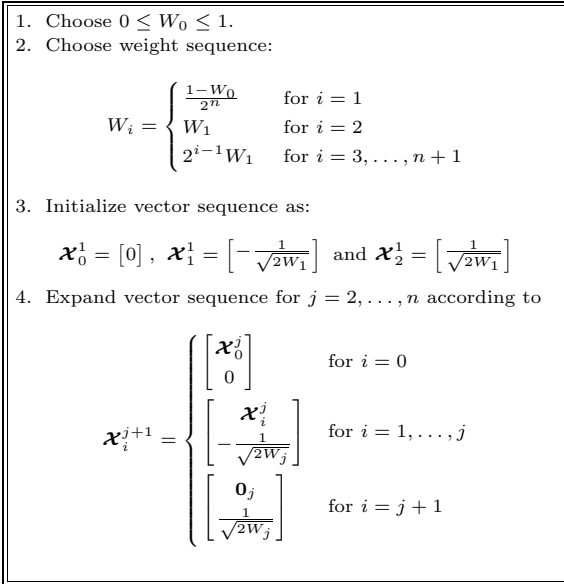


Fig. 1. The Point Selection Algorithm for the Simplex Unscented Transform.

From Equation 13,

$$W_{n+1} = 2W_1 + \sum_{i=3}^n W_i = 2^n W_1.$$

Substituting into Equation 14,

$$x_{n+1} = \frac{1}{\sqrt{2W_{n+1}}}.$$

From Equation 12,

$$W_0 = 1 - 2W_1 - \sum_{i=3}^{n+1} W_i = 1 - 2^n W_1.$$

Therefore,

$$W_1 = \frac{1 - W_0}{2^n}.$$

The algorithm is summarized in Figure 1.

It should be noted that, as the number of dimensions  $n$  increase, the effects of the third order moments become more significant. A companion paper presents a general scaling algorithm which preserves the first two moments of the sigma points but allows third and higher order moments to be scaled by an arbitrary amount.

We illustrate the performance on the simplex sigma points in a 3D vehicle localization example.

#### IV. EXAMPLE

The position and orientation of a vehicle operating in an uneven 3D terrain is to be estimated. The vehicle contains onboard inertial sensors which measure angular velocity about body-fixed axes ( $\omega_x, \omega_y, \omega_z$ ) and odometric sensors (which measure nominal linear speed  $v$ ). Periodically, the vehicle receives external measurements of its position and orientation (from, for example, a 3D map-based localization algorithm). Because the ground is uneven, the bandwidth of the filter must be extremely high. Therefore, a filter design similar to the one proposed by Kelly [15] is used: the gyro measurements are also treated as control inputs to the platform orientation. As a result, the high frequency content of the inertial data is preserved.

The vehicle state is:

$$\mathbf{x}(k) = [x \ y \ z \ q_0 \ q_1 \ q_2 \ q_3 \ b_x \ b_y \ b_z \ s]^T.$$

The position of the center of the vehicle is  $(x, y, z)$ . The vehicle orientation is expressed as the unit quaternion  $q = (q_0, q_1, q_2, q_3)$ .  $(b_x, b_y, b_z)$  are the gyro biases and  $s$  is a slip parameter which takes account of the slip between the wheels and the ground. Specifically, if  $v(k)$  is the measured speed of the vehicle at time step  $k$ , the actual translational speed of the vehicle is  $s(k)v(k)$ . It is assumed that the vehicle only travels along its body-fixed  $x$ -axis.

The control inputs are

$$\mathbf{u}(k) = [v \ \omega_x \ \omega_y \ \omega_z]^T.$$

There are eight process noise terms. These arise from the measurements of the control inputs as well as the assumption that the gyro bias and slip values vary according to a random walk:

$$\mathbf{v}(k) = [\delta v \ \delta \omega_x \ \delta \omega_y \ \delta \omega_z \ \delta \dot{b}_x \ \delta \dot{b}_y \ \delta \dot{b}_z \ \delta \dot{s}]^T.$$

Suppressing the time index  $k$ , let

$$\begin{aligned} p &= \Delta T(\omega_x + \delta \omega_x - b_x)/2 \\ q &= \Delta T(\omega_y + \delta \omega_y - b_y)/2 \\ r &= \Delta T(\omega_z + \delta \omega_z - b_z)/2 \\ n &= \sqrt{p^2 + q^2 + r^2} \end{aligned}$$

$$\mathbf{M} = \mathbf{I} \cos n + \begin{bmatrix} 0 & r & -q & p \\ -r & 0 & p & q \\ q & -p & 0 & r \\ -p & -q & -r & 0 \end{bmatrix} \frac{\sin n}{n}$$

Therefore the vehicle process model is

$$\begin{aligned} x(k+1) &= x + s(v + \delta v)(q_0^2 + q_3^2 - q_1^2 - q_2^2)\Delta T \\ y(k+1) &= y + 2s(v + \delta v)(q_0 q_1 + q_2 q_3)\Delta T \\ z(k+1) &= z + 2s(v + \delta v)(q_0 q_2 - q_1 q_3)\Delta T \\ q_0(k+1) &= M_{11}q_0 + M_{12}q_1 + M_{13}q_2 + M_{14}q_3 \\ q_1(k+1) &= M_{21}q_0 + M_{22}q_1 + M_{23}q_2 + M_{24}q_3 \\ q_2(k+1) &= M_{31}q_0 + M_{32}q_1 + M_{33}q_2 + M_{34}q_3 \\ q_3(k+1) &= M_{41}q_0 + M_{42}q_1 + M_{43}q_2 + M_{44}q_3 \\ b_x(k+1) &= b_x + \delta \dot{b}_x \Delta T \\ b_y(k+1) &= b_y + \delta \dot{b}_y \Delta T \\ b_z(k+1) &= b_z + \delta \dot{b}_z \Delta T \\ s(k+1) &= s + \delta \dot{s} \Delta T \end{aligned}$$

The length of each time step  $\Delta T = 0.002s$ . A measurement of the attitude becomes available once in every 500 time steps (once per second). The measurement consists of the position and orientation of the vehicle (expressed as a quaternion).

The measurement noise on the control inputs<sup>3</sup> are

$$\mathbf{R}_u(k) = \begin{bmatrix} 1.0 \times 10^{-4} & 0 & 0 & 0 \\ 0 & 9.7 \times 10^{-7} & 4.4 \times 10^{-9} & 2.5 \times 10^{-8} \\ 0 & 4.4 \times 10^{-9} & 1.0 \times 10^{-6} & 3.0 \times 10^{-8} \\ 0 & 2.5 \times 10^{-8} & 3.0 \times 10^{-8} & 1.1 \times 10^{-6} \end{bmatrix}$$

The process noise which acts on the simulated vehicle results from the drift parameters in the bias states and the slip:

$$\mathbf{Q}_T(k) = \text{diag}(0, 0, 0, 0, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6})$$

<sup>3</sup>The noises on the gyros were empirically calculated from the measurements of a Crossbow IMU.

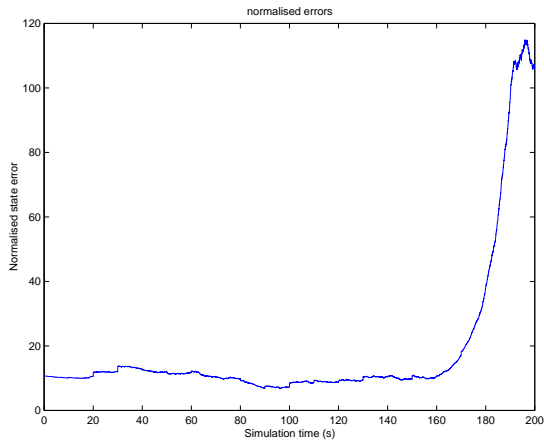


Fig. 2. The normalized state error for the EKF.

The covariance matrix used in the filter consists of the drift terms plus the covariance due to the noise in the control input measurement terms,

$$\mathbf{Q}(k) = \mathbf{Q}_T(k) + \begin{bmatrix} \mathbf{R}_u(k) & \mathbf{0}_{4 \times 4} \\ \mathbf{0}_{4 \times 4} & \mathbf{0}_{4 \times 4} \end{bmatrix}$$

The attitude measurement noise is

$$\mathbf{R}(k) = \text{diag}(0.25, 0.25, 0.25, 0.001, 0.001, 0.001, 0.001).$$

When an attitude measurement is received, it is updated using the Kalman filter and the gyro states are renormalized to ensure that the norm of the quaternion states is 1.

The results from the EKF are shown in Figure 2. This plots the normalized state error, which is defined to be:

$$\tilde{\mathbf{x}}^T(k|k) \mathbf{P}^{-1}(k|k) \tilde{\mathbf{x}}(k|k).$$

This is an average calculated over 10 Monte Carlo runs. If the filter is consistent, the mean value should be the same as the dimensions of the state (10). Although the results initially appear consistent, at simulation time of about 160s the normalized error sharply increases, showing that the filter has become inconsistent. This is a result of accumulated linearization errors. Figure 3 plots the symmetric sigma point result (using the points drawn by Equation 4). As can be seen, the filter is consistent and is, in fact, slightly conservative. Figure 4 plots the results from the simplex transform. Although the general shape of the curve is different (due to the different set of sigma points), the mean of this curve still indicates that the estimate is consistent.

## V. CONCLUSIONS

This paper has investigated the use of minimal sigma point distributions to propagate the first two moments of a random variable through a nonlinear transformation. We have derived an explicit sigma point solution which, for an  $n$ -dimensional space, requires only  $n+2$  sigma points. The resulting algorithm has the same predictive capability as the truncated second order filter but without the need to use derivatives or any other kinds of approximations. This has been demonstrated through a simulation study which has also demonstrated the utility of the scaled unscented transformation.

The simulation study also shows that the performance of the symmetric and simplex sets can be markedly different. In the

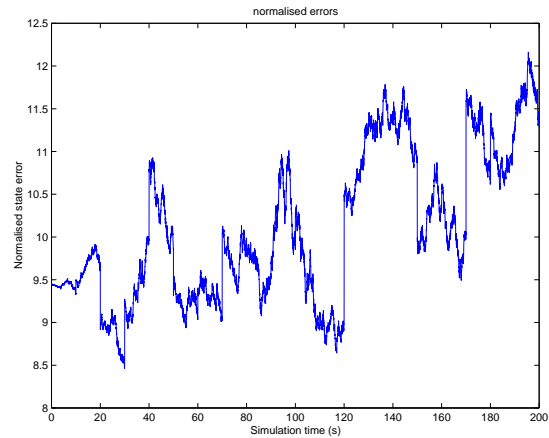


Fig. 3. The results from the symmetric transformation.

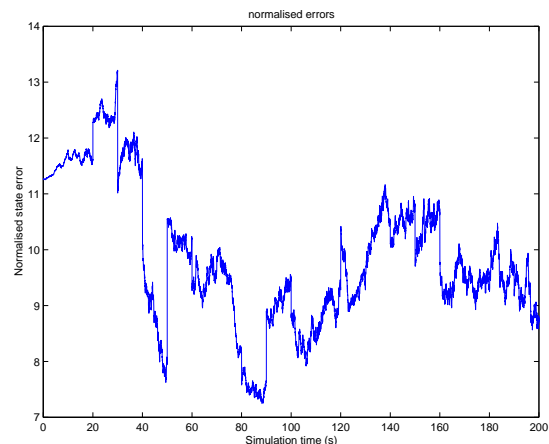


Fig. 4. The results from the simplex transformation.

localization scenario described here, the shape of the normalized error curves were different but both filters were consistent. However, this is not always the case and, in a related paper, we describe a method for minimising unwanted effects of higher order moments [16].

## REFERENCES

- [1] J. K. Uhlmann, "Algorithms for multiple target tracking," *American Scientist*, vol. 80, no. 2, pp. 128–141, 1992.
- [2] H. W. Sorenson, Ed., *Kalman filtering: theory and application*, IEEE Press, Piscataway NJ, USA, 1985.
- [3] D. Lerro and Y. K. Bar-Shalom, "Tracking with Debiased Consistent Converted Measurements vs. EKF," *IEEE Transactions on Aerospace and Electronics Systems*, vol. AES-29, no. 3, pp. 1015–1022, July 1993.
- [4] S. Thrun D. Fox, W. Burgard and F. Dellaert, "Robust Monte Carlo Localization for Mobile Robots," *Artificial Intelligence*, , no. 101, pp. 99–141, 2000.
- [5] N. J. Gordon, D. J. Salmond and A. F. M. Smith, "Novel Approach to Nonlinear/non-Gaussian Bayesian State Estimation," *IEE Proceedings-F*, vol. 140, no. 2, pp. 107–113, April 1993.
- [6] M. Isard and A. Blake, "condensation – conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.
- [7] S. J. Julier and H. F. Durrant-Whyte, "Navigation and Parameter Estimation of High Speed Road Vehicles," in *Robotics and Automation Conference, Nagoya, Japan.*, 1995, pp. 101–105.
- [8] S. J. Julier, *Comprehensive Process Models for High-Speed Navigation*, Ph.D. thesis, University of Oxford, October 1997.
- [9] J. K. Uhlmann, *Dynamic Map Building and Localization for Autonomous Vehicles*, Ph.D. thesis, University of Oxford, 1995.
- [10] 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*, vol. 5, no. 3, pp. 477–482, March 2000.

- [11] S. J. Julier and J. K. Uhlmann, "A Consistent, Debiased Method for Converting Between Polar and Cartesian Coordinate Systems," in *The Proceedings of AeroSense: Acquisition, Tracking and Pointing XI*. 1997, vol. 3086, pp. 110–121, SPIE.
- [12] S. J. Julier, "A Skewed Approach to Filtering," in *The Proceedings of AeroSense: The 12th International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, Orlando FL, USA, April 1998, vol. 3373, pp. 54–65, SPIE, Signal and Data Processing of Small Targets.
- [13] M. Noorgaard, N. K. Poulsen and O. Ravn, "New developments in state estimation for nonlinear systems," *Automatica*, vol. 36, pp. 1627–1638, 2000.
- [14] 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.
- [15] A. Kelly, "A 3D State Space Formulation of a Navigation Kalman Filter for Autonomous Vehicles," Tech. Rep. CMU-RI-TR-94-19, Carnegie Mellon University, May 1994.
- [16] S. Julier, "The Scaled Unscented Transformation," Submitted to the 2002 American Control Conference, September 2001.