A view on the Unscented Kalman Filter and comparison with the Extended Kalman Filter

David Di Ruscio

Telemark Institute of technology
N-3914 Porsgrunn, Norway
Fax: +47 35 57 52 50
Tel: +47 35 57 51 68
E-mail: David.Di.Ruscio@hit.no

1 Introduction

The Unscented Kalman Filter (UKF) algorithm is a variant of the Extended Kalman Filter (EKF) algorithm for non-linear systems. The UKF algorithm is a derivative free Kalman filter algorithm which is based on the Unscented Transform (UT) for computing the mean and covariance matrix of a variable, say y, which is given by a non-linear function $y = g(x)$ where the mean and covariance matrix of the argument x are given. The EKF is a "direct" extension of the Kalman filter for linear systems to on non-linear system models where the linear system matrix A and the linear measurements matrix C are replaced by the Jacobian matrices, i.e., $A := \frac{\partial f}{\partial x_k}$ and $C := \frac{\partial h}{\partial x_k}$, respectively. These Jacobian matrices may be calculated analytically but it is well known that the Jacobians may be calculated numerically, e.g. by using the explicit Euler method ore by its central difference approximations. Hence, the EKF may be implemented with evaluations of the non-linear functions only. hence, it may be of interest to compare the UKF and EKF algorithms, both with respect to accuracy on some particular examples and to computational aspects.

It is reported that the UKF algorithm gives the same result as the Kalman filter for linear systems and that the UKF may give better results than the EKF for non-linear systems, but this is not the case in general.

This work is inspired by the work by Sarkka (2007) and some of the material is taken and translated to the notation used in this note.

2 Model description

Given a possibly nonlinear state space model

$$x_{k+1} = f(x_k) + v_k, \tag{1}$$
$$y_k = h(x_k) + w_k, \tag{2}$$

where v_k and w_k are white Gaussian distributed process and measurements noise, respectively. The covariance matrices are $V = \mathrm{E}(v_k v_k^T)$ and $W = \mathrm{E}(w_k w_k^T)$, which are assumed

to be known or specified. In some circumstances the covariance matrices may be time varying, e.g., $V_k := V$ and $W_k := W$. We will assume that the process noise, v_k, and the measurements noise w_k are un-correlated, i.e., $\mathrm{E}(v_k w_k^T) = 0$.

3 On the filter matrix equations

We will in this section take a view on some central covariance matrix equations which are used in the EKF and derive the corresponding covariance matrix equations used in the UKF.

3.1 Covariance matrix equations

For the Kalman filter we may formulate the following covariance matrix equations. The a-posteriori covariance matrix, $\hat{X}_k = \mathrm{E}((x_k - \hat{x}_k)(x_k - \hat{x}_k)^T)$, may be formulated as a function of the a-priori covariance matrix, $\bar{X}_k = \mathrm{E}((x_k - \bar{x}_k)(x_k - \bar{x}_k)^T)$, and the Kalman filter gain matrix, K_k, as

$$\hat{X}_k = \bar{X}_k - K_k D \bar{X}_k. \tag{3}$$

The Kalman filter gain matrix is given by

$$K_k = \bar{X}_k D^T (D \bar{X}_k D^T + W)^{-1}. \tag{4}$$

From Eq. (4) we have that the term $D\bar{X}_k$ in Eq. (3) may be expressed as

$$D\bar{X} = (\bar{X}_k D^T)^T = (D \bar{X}_k D^T + W) K_k^T, \tag{5}$$

where $\bar{X}_k D^T = K_k (D \bar{X}_k D^T + W)$ is found from Eq. (4).

Combining Eqs. (3) and (5) gives the alternative formulation

$$\hat{X}_k = \bar{X}_k - K_k (D \bar{X}_k D^T + W) K_k^T. \tag{6}$$

Furthermore, notice, and define the following. Define the cross covariance matrix between x_k and y_k as

$$C_k = \bar{X}_k D^T = \mathrm{E}((x_k - \bar{x}_k)(x_k - \bar{x}_k)^T) D^T = \mathrm{E}((x_k - \bar{x}_k)(y_k - \bar{y}_k)^T). \tag{7}$$

Define the matrix

$$S_k = D \bar{X}_k D^T + W = \tilde{S}_k + W, \tag{8}$$

where $\tilde{S}_k$ is the covariance matrix of the output y_k, which may be expressed as

$$\tilde{S}_k = D \bar{X}_k D^T = \mathrm{E}((y_k - \bar{y}_k)(y_k - \bar{y}_k)^T). \tag{9}$$

Using the definitions given by Eqs. (7)-(9) shows that the covariance matrix Eqs. (4) and (5) may be written as

$$K_k = C_k S_k^{-1}. \tag{10}$$

and

$$\hat{X}_k = \bar{X}_k - K_k S_k K_k^T, \tag{11}$$

where $\tilde{S}_k$ is the covariance matrix of the output y_k and given by Eq. (9).

Eq. (10) for the Kalman filter gain matrix, and Eq. (11) for the a-posteriori state covariance matrix are of central importance and used in the UKF algorithm.

The remaining part in the UKF algorithm is to compute the covariance matrices $\bar{X}_k$, $\tilde{S}_k$ and the cross covariance matrix C_k. This is done by using the so called Unscented Transform (UT). The covariance matrix $\bar{X}_k$ and the a-priori state estimate $\bar{x}_k$ are found by using the UT algorithm on the non-linear function $f(\cdot)$. The covariance matrix $\tilde{S}_k$, the cross covariance matrix C_k and the predicted (mean) output $\bar{y}_k$ are found by using the UT algorithm on the non-linear measurement function $h(\cdot)$.

The UKF algorithm is presented in Section 5, but first we give a view of the EKF in the next Section 4.

3.2 On the a-posteriori covariance matrix equation

Consider the a-posteriori state estimate

$$\hat{x}_k = \bar{x}_{k+1} + K_k(y_k - \bar{y}_k). \tag{12}$$

This gives

$$x_k - \hat{x}_k = x_k - \bar{x}_k - K_k(Dx_k + w_k - D\bar{x}_k). \tag{13}$$

An expression for the a-posteriori state covariance matrix may be found from

$$\begin{gathered}(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T = \\ (x_k - \bar{x}_k - K_k(Dx_k + w_k - D\bar{x}_k))(x_k - \bar{x}_k - K_k(Dx_k + w_k - D\bar{x}_k))^T\end{gathered} \tag{14}$$

Evaluating the right hand side of Eq. (14) and taking the mean gives

$$\hat{X}_k = (I - K_k D)\bar{X}_k(I - K_k D)^T + K_k W K_k^T \tag{15}$$

Eq. (15) is frequently used in the KF and EKF algorithm due to its numerical stability properties.

Some alternative formulations of Eq. (15) may be deduced as follows. From Eq. (15) we have

$$\hat{X}_k = \bar{X}_k - \bar{X}_k D^T K_k^T - K_k D^T \bar{X}_k + K_k D^T \bar{X}_k D^T K_k^T + K_k W K_k^T, \tag{16}$$

which gives

$$\hat{X}_k = \bar{X}_k - \bar{X}_k D^T K_k^T - K_k D^T \bar{X}_k + K_k(D\bar{X}_k D^T + W)K_k^T. \tag{17}$$

from the expression for the Kalman filter gain matrix $K_k = \bar{X}_k D^T (D\bar{X}_k D^T + W)^{-1}$ we find that $\bar{X}_k D^T = K_k(D\bar{X}_k D^T + W)$ and this may be used to show that the sum of the 2nd and 3rd terms on the right hand side of Eq. (17) may be expressed as

$$\bar{X}_k D^T K_k^T + K_k D^T \bar{X}_k = 2K_k(D\bar{X}_k D^T + W)K_k^T. \tag{18}$$

Using Eq. (18) in Eq. (17) gives an alternative formulation also used in the UKF algorithm

$$\hat{X}_k = \bar{X}_k - K_k(D^T \bar{X}_k D^T + W)K_k^T. \tag{19}$$

Using that $K_k(D\bar{X}_k D^T + W) = \bar{X}_k D^T$ we find the alternative

$$\hat{X}_k = \bar{X}_k - \bar{X}_k D^T K_k^T = \bar{X}_k - K_k D \bar{X}_k. \tag{20}$$

4 The Extended Kalman Filter

The Extended kalman Filter (EKF) algorithm, Jazwinski (1970), may be formulated in different ways. For the sake of comparison with the UKF, we will in this section give a presentation of the EKF which is as close as the UKF algorithm as possible, in order to point out the main differences. The EKF algorithm may be formulated as follows

1. Given initial values for the predicted (a-priori) state estimate, $\bar{x}_k$, and its covariance matrix $\bar{X}_k$ at startup, e.g. at time $k = 0$.

2. Compute the predicted measurement/output

$$\bar{y}_k = h(\bar{x}_k), \tag{21}$$

   where the covariance matrix of the output, y_k, is $\tilde{S}_k = \mathrm{E}((y_k - \bar{y}_k)(y_k - \bar{y}_k)^T)$, and the cross-covariance matrix is $C_k = \mathrm{E}((x_k - \bar{x}_k)(y_k - \bar{y}_k)^T)$.

   Compute the Jacobian of the non-linear measurements function $h(\cdot)$, i.e.,

$$D_k = \left. \frac{\partial h(x_k)}{\partial x_k^T} \right|_{\bar{x}_k}. \tag{22}$$

   Then we have that the covariance matrix of the output, $\tilde{S}_k$, and the cross covariance matrix between x_k and y_k, i.e.,C_k are given by

$$\tilde{S}_k = D_k \bar{X}_k D_k^T, \tag{23}$$
$$C_k = \bar{X}_k D_k^T. \tag{24}$$

   Furthermore, the intermediate matrix $D\bar{X}_k D^T + W = S_k$ is then given by

$$S_k = \tilde{S}_k + W, \tag{25}$$

   which is used for computing the Kalman filter gain matrix in the next step of the EKF algorithm.

3. Compute the EKF gain matrix

$$K_k = C_k S_k^{-1}. \tag{26}$$

4. Compute the a-posteriori state estimate

$$\hat{x}_k = \bar{x}_k + K_k(y_k - \bar{y}_k), \tag{27}$$

   and the a-posteriori state covariance matrix

$$\hat{X}_k = \bar{X}_k - K_k S_k K_k^T. \tag{28}$$

5. Compute the updated state prediction (a-priori state estimate) using the model

$$\bar{x}_{k+1} = f(\hat{x}_k). \tag{29}$$

Compute the Jacobian (transition) matrix of the non-linear measurements function $f(\cdot)$, i.e.,

$$A_k = \left. \frac{\partial f(x_k)}{\partial x_k^T} \right|_{\bar{x}_k}, \tag{30}$$

and the intermediate expression

$$\tilde{X}_{k+1} = A_k \bar{X}_k A_k^T, \tag{31}$$

and the a-priori state covariance matrix update

$$\bar{X}_{k+1} = \tilde{X}_{k+1} + V. \tag{32}$$

We will in the next Section 5 give a presentation of the UKF algorithm. As we will see, the main difference between the EKF and the UKF lies in steps 2 and 5 of the above EKF algorithm.

5 The Unscented Kalman Filter

In terms of the Unscented Transform, i.e., the function $\mathrm{UT}(\cdot)$, the UKF algorithm may be formulated as follows

1. Given initial values for the predicted (a-priori) state estimate, $\bar{x}_k$, and its covariance matrix $\bar{X}_k$ at startup, e.g. at time $k = 0$.

2. Compute the predicted measurement/output

$$\left[\begin{array}{ccc} \bar{y}_k, & \tilde{S}_k, & C_k \end{array} \right] = UT(h(\cdot), \bar{x}_k, \quad \bar{X}_k), \tag{33}$$

where the covariance matrix is $\tilde{S}_k = \mathrm{E}((y_k - \bar{y}_k)(y_k - \bar{y}_k)^T)$, and the cross-covariance matrix is $C_k = \mathrm{E}((x_k - \bar{x}_k)(y_k - \bar{y}_k)^T)$. Then we have the term $D\bar{X}_k D^T + W$ as

$$S_k = \tilde{S}_k + W. \tag{34}$$

3. Compute the UKF gain matrix

$$K_k = C_k S_k^{-1}. \tag{35}$$

4. Compute the a-posteriori state estimate

$$\hat{x}_k = \bar{x}_k + K_k(y_k - \bar{y}_k), \tag{36}$$

and the a-posteriori state covariance matrix

$$\hat{X}_k = \bar{X}_k - K_k S_k K_k^T. \tag{37}$$

5. Compute the updated state prediction (a-priori state estimate) using the model and the UT

$$\left[\begin{array}{cc} \bar{x}_{k+1}, & \tilde{X}_{k+1} \end{array} \right] = UT(f(\cdot), \hat{x}_k, \hat{X}_k) \tag{38}$$

and the a-priori state covariance matrix update

$$\bar{X}_{k+1} = \tilde{X}_{k+1} + V. \tag{39}$$

6 The unscented transform

The UKF is based on a statement "that it is easier to approximate a Gausian probability distribution than it is to approximate an arbitrarily non-linear function". Instead of linearizing and calculating the Jacobian matrices $D_k = \frac{\partial h}{\partial x_k^T}|_{\bar{x}_k}$ and $A_k = \frac{\partial f}{\partial x_k^T}|_{\hat{x}_k}$, the UKF is using a deterministic approach to calculate the mean and covariance of a variable. This deterministic approach is denoted the Unscented Transform (UT). In the UT usually $2n+1$ discrete points, denoted sigma points, are used to approximate the continuous normal Gaussian probability distribution (Gaussian function). A more crude approximation are using $n+1$ sigma points.

Consider that the a-priori state estimate, $\bar{x}_k$, is known. Then a set of $2n+1$ points, on and around $\bar{x}_k$ is chosen. These $2n+1$ perturbations ore vectors around the mean (a-priori estimate) are denoted sigma points. These $2n+1$ deterministically chosen sigma points is then propagated through the non-linear functions $f(\cdot)$ and $h(\cdot)$ and the mean and covariance of the transformed variable are estimated from them.

Consider a random Gaussian normal distributed variable $x_k \in \mathbb{R}^n$ with mean $\bar{x}_k$ and covariance matrix $\bar{X}_k \in \mathbb{R}^{n\times n}$ formally defined as $x_k \sim \mathrm{N}(\bar{x}_k, \bar{X}_k)$ where $\mathrm{N}(\cdot)$ means a Gaussian normal distribution. Assume that a new random variable y_k is obtained from a non-linear transformation of the random variable x_k as

$$y_k = g(x_k), \tag{40}$$

where $g(\cdot) \in \mathbb{R}^m$ is a non-linear function, i.e., a non-linear mapping of the n-dimensional vector x_k to the transformed random vector $y_k \in \mathbb{R}^m$. The transformed variable y_k is also normal distributed and the joint probability distribution of x_k and y_k is described by the normal distribution

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} \sim N(\begin{bmatrix} \bar{x}_k \\ \bar{y}_k \end{bmatrix}, \begin{bmatrix} \bar{X}_k & C_k \\ C_k^T & \tilde{S}_k \end{bmatrix}). \tag{41}$$

Here $\bar{y}_k$ and $\tilde{S}_k = \mathrm{E}(y_k - \bar{y}_k)((y_k - \bar{y}_k)^T$ are the mean and the covariance matrix of y_k, respectively. Furthermore, $C_k = \mathrm{E}(x_k - \bar{x}_k)((y_k - \bar{y}_k)^T$ is the cross covariance matrix between x_k and y_k.

The UT is an algorithm to estimate the mean $\bar{y}_k$ and the covariance matrices C_k and $\tilde{S}_k$ in the joint normal distribution Eq. (41). Based on the above definitions we present the UT as a function of the form

$$\begin{bmatrix} \bar{y}_k, & \tilde{S}_k, & C_k \end{bmatrix} = UT(g(\cdot), \bar{x}_k, \quad \bar{X}_k), \tag{42}$$

where the mean $\bar{y}_k$ and the covariance matrices C_k and $\tilde{S}_k$ are calculated by the UT from given mean $\bar{x}_k$ and covariance matrix $\bar{X}_k$ and the non-linear function Eq. (40).

There are initially three parameters, α,β and κ, which are chosen in the UT. Common choices for these parameters are

$$\alpha = 0.001, \ \beta = 2, \ \kappa = 0. \tag{43}$$

Define a scaling parameter λ as

$$\lambda = \alpha^2(n+\kappa) - n, \tag{44}$$

and the square root of the scaled covariance matrix $(n+\lambda)\bar{X}_k$, i.e.,

$$L = ((n+\lambda)\bar{X}_k)^{\frac{1}{2}}, \tag{45}$$

where L is the square root of $(n+\lambda)\bar{X}_k$ satisfying $LL^T = (n+\lambda)\bar{X}_k$. The square root matrix L may be computed by the transpose of the Cholesky decomposition as

$$L^T = \text{chol}((n+\lambda)\bar{X}_k)). \tag{46}$$

If the Cholesky decomposition is used then the matrix L is the lower triangular square root of the scaled covariance matrix $(n+\lambda)\bar{X}_k$. The Cholesky decomposition is to be preferred for the calculation of the square root matrix L. Notice that it is possible to scale with $\sqrt{n+\lambda}$ after the Cholesky decomposition

From the square root covariance matrix L we form the $2n+1$ sigma points (vectors) and store these vectors in a matrix $\tilde{X} \in \mathbb{R}^{n \times 2n+1}$, i.e.,

$$\tilde{X} = \left[\begin{array}{cccc} \tilde{x}_1, & \tilde{x}_2 & \cdots & \tilde{x}_{2n+1} \end{array} \right] \tag{47}$$

where the sigma points (vectors) are distributed around the mean $\bar{x}_k$ as follows

$$\begin{aligned} \tilde{x}_1 &= \bar{x}_k & (48)\\ \tilde{x}_i &= \bar{x}_k + l_i \ \forall \ i = 1, \ldots, n & (49)\\ \tilde{x}_1 &= \bar{x}_k - l_{i-n} \ \forall \ i = n+1, \ldots, 2n & (50) \end{aligned}$$

where l_i is the ith column in the square root matrix, $L = \left[\begin{array}{cccc} l_1 & \cdots & l_i & \cdots & l_n \end{array} \right]$. The matrix of sigma points (vectors) may be formulated directly as the matrix

$$\tilde{X} = \left[\begin{array}{ccc} \bar{x}_k, & \bar{x}_k 1_{1\times n} + L & \bar{x}_k 1_{1\times n} - L \end{array} \right], \tag{51}$$

where $1_{1\times n}$ is the $1 \times n$ vector of ones, i.e., $1_{1\times n} = \left[\begin{array}{cccc} 1 & 1 & \cdots & 1 \end{array} \right]$ and then hence $\bar{x}_k 1_{1\times n} = \left[\begin{array}{cccc} \bar{x}_k & \bar{x}_k & \cdots & \bar{x}_k \end{array} \right]$.

from this we see that first sigma vector is the mean, i.e., $\tilde{x}_1 = \bar{x}_k$ and n sigma points (vectors) distributed above (positive) the mean and n sigma points (vectors) below (negative) the mean.

Each of the $2n+1$ sigma points (vectors) are then transformed through the non-linear function os

$$y_i = g(\tilde{x}_i) \ \forall \ i = 1, \ldots, 2n+1 \tag{52}$$

The mean $\bar{y}_k$ are then computed as

$$\bar{y}_k \approx \sum_{i=1}^{2n+1} w_i^m y_i \tag{53}$$

and the covariance matrices $\tilde{S}_k$ and C_k as

$$\begin{aligned} \tilde{S}_k &= \text{E}((y_k - \bar{y}_k)(y_k - \bar{y}_k)^T) \approx \sum_{i=1}^{2n+1} w_i^c (y_i - \bar{y}_k)(y_i - \bar{y}_k)^T & (54)\\ C_k &= \text{E}((x_k - \bar{x}_k)(y_k - \bar{y}_k)^T) \approx \sum_{i=1}^{2n+1} w_i^c (x_i - \bar{x}_k)(y_i - \bar{y}_k)^T & (55) \end{aligned}$$

In Eq. (53) the parameters $w_i^m \ \forall \ i = 1, \ldots, 2n+1$ is the weights for the mean. Similarly, $w_i^c \ \forall \ i = 1, \ldots, 2n+1$ in Eqs. (54)-(55) are the weights for the covariance matrices. They are defined as follows

$$w_1^m = \frac{\lambda}{n+\lambda}, \ w_i^m = \frac{1}{2(n+\lambda)} \ \forall \ i = 2, \ldots, 2n+1 \tag{56}$$

and

$$w_1^c = \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta), \tag{57}$$

$$w_i^m = \frac{1}{2(n+\lambda)} \ \forall \ i = 2, \ldots, 2n+1 \tag{58}$$

We notice that the weights for the mean, w_i^m, and the weights for the covariance matrices, w_i^c, are equal, except for $i = 1$.

References

Jazwinski, A.-H. (1970), *Stochastic Proceses and Filtering Theory*, Academic Press.

Sarkka, S. (2007), "On Unscented Kalman Filtering for State Estimation of Continuous-Time Nonlinear Systems", *IEEE Trans. on Automatic Control* , Vol. 21, pp. 1–11.