On Unscented Kalman Filtering for State Estimation of Continuous-Time Nonlinear Systems

Simo Särkkä*, *Member, IEEE*

***Abstract*— This article considers the application of the unscented Kalman filter (UKF) to continuous-time filtering problems, where both the state and measurement processes are modeled as stochastic differential equations. The mean and covariance differential equations which result in the continuous-time limit of the UKF are derived. The continuous-discrete unscented Kalman filter is derived as a special case of the continuous-time filter, when the continuous-time prediction equations are combined with the update step of the discrete-time unscented Kalman filter. The filter equations are also transformed into sigma-point differential equations, which can be interpreted as matrix square root versions of the filter equations.**

***Index Terms*— unscented Kalman filter, continuous-time filter, continuous-discrete filter, stochastic differential equation, nonlinear system, continuous-time state space model, nonlinear state space model**

I. Introduction

THE unscented Kalman filter (UKF) [1]–[3] is an efficient derivative free filtering algorithm for computing approximate solutions to discrete-time non-linear optimal filtering problems. It has been successfully applied to numerous practical problems and it has been shown to outperform the extended Kalman filter (EKF) in many cases [4]. However, in its original form, the UKF is a discrete-time algorithm and it cannot be directly applied to continuous-discrete filtering problems, where the state dynamics are modeled as continuous-time stochastic processes, or to continuous-time filtering problems, where both the state and measurement processes are modeled as continuous-time stochastic processes.

Phenomena, which can be modeled as time varying systems, where a continuous-time signal is observed discretely or continuously in time are very common in engineering and physics applications. Examples of such applications are GPS and inertial navigation [5], [6], target tracking [7]–[9], estimation of biological processes [10], telecommunications [11], [12], stochastic optimal control [13], [14] and inverse problems in physics [15].

In this article, the differential equations which result in the continuous-time limit of the UKF are derived. Both the continuous-time and continuous-discrete cases are considered. The derived continuous-time filtering equations, which could be called the unscented Kalman-Bucy filter (UKBF) equations, are similar to the extended Kalman-Bucy filter (EKBF) equations [16]–[18] and consist of a pair of differential equations for the mean and covariance of the posterior state process. A square root version of the filter is also derived and it consists of a matrix differential equation for the sigma points.

The continuous-discrete unscented Kalman filter is derived as a special case of the continuous-time filter, when the continuous-time prediction equations are combined with the update step of the discrete-time UKF. A square root version of this continuous-discrete filter is also presented.

Manuscript received April 26, 2005; revised February 9, 2007.

Simo Särkkä* is with Helsinki University of Technology, P.O. Box 9203. FIN-02015 HUT, Finland. E-mail: simo.sarkka@hut.fi, Tel: +358 40 757 0730, Fax: +358 9 451 4830.

II. Problem Statement

A. Optimal Continuous-Time Filtering

In analog communication systems [11] the measured signals are typically continuous-time processes and analog receivers are devices, which demodulate or estimate the actual transmitted continuous-time signals from the noisy measured signals. Also in many analog (electrical or mechanical) control systems (see, e.g., [13], [14]) operating without digital computers the measured signals are continuous-time, not discrete-time signals. The optimal recursive estimation of this kind of systems is called *optimal continuous-time filtering* [16], [17].

The most general form of the optimal continuous-time filtering models considered in this article is

$$\begin{aligned} \mathrm{d}x(t) &= f(x(t),t)\,\mathrm{d}t + L(t)\,\mathrm{d}\beta(t) \\ \mathrm{d}y(t) &= h(x(t),t)\,\mathrm{d}t + V(t)\,\mathrm{d}\eta(t), \end{aligned} \tag{1}$$

where

- $x(t) \in \mathbb{R}^n$ is the state process,
- $y(t) \in \mathbb{R}^m$ is the (integrated) measurement process,
- f is the drift function,
- h is the measurement model function,
- $L(t)$ and $V(t)$ are arbitrary time varying matrices, independent of $x(t)$ and $y(t)$,
- $\beta(t)$ and $\eta(t)$ are independent Brownian motions with diagonal diffusion matrices $Q_c(t)$ and $R_c(t)$, respectively.

The dynamic and measurement models can be equivalently interpreted as Itô or Stratonovich type *stochastic differential equations* [19], [20].

The filtering model can also be formulated in terms of formal white noises $e(t) = \mathrm{d}\beta(t)/\mathrm{d}t$, $n(t) = \mathrm{d}\eta(t)/\mathrm{d}t$, and differential measurement $z(t) = \mathrm{d}y(t)/\mathrm{d}t$ as follows [16]:

$$\begin{aligned} \frac{\mathrm{d}x(t)}{\mathrm{d}t} &= f(x(t),t) + L(t)\,e(t) \\ z(t) &= h(x(t),t) + V(t)\,n(t), \end{aligned} \tag{2}$$

where the white noise processes $e(t)$ and $n(t)$ have spectral densities $Q_c(t)$ and $R_c(t)$, respectively.

The purpose of the optimal (Bayesian) continuous-time filtering is to recursively compute the posterior distribution

$$p(x(t) \,|\, \{y(\tau) \,:\, 0 \le \tau \le t\}), \tag{3}$$

or at least the relevant moments of the distribution (e.g., mean and covariance).

The formal solution to the filtering problem is well known and it is given by the *Kushner-Stratonovich equation* [21]–[23], which is a measure valued stochastic partial differential equation. The unnormalized version of the equation is called the *Zakai equation* [24], [25]. However, these equations only give the *formal solution*, and the actual computation of the distribution or its expectations (such as mean and covariance) would require an infinite amount of computational resources. In certain special cases, the equations do have finite dimensional solutions, which leads to Kalman-Bucy filters [26] and Beneš filters [27].

Because the exact solution to the continuous-time optimal filtering problem is generally intractable, approximations must be used. The most common approximation method is the extended Kalman-Bucy filter (EKBF) (see, e.g., [16]–[18]), which approximates the exact solution by replacing the non-linear model with a suitably linearized approximate model, which can be solved by the Kalman-Bucy filter. Another general way of forming approximations is Monte Carlo sampling [28], [29], where a set of weighted particles is used for approximating the posterior probability measure.

B. Optimal Continuous-Discrete Filtering

Nowadays, in many signal processing systems, the sensor measurements are obtained at discrete instances of time either due to sampling, due to processing delays in the device or because the sensor operates in scans. Still in the Nature time is continuous, not discrete, and for this reason often a physically more realistic approach than discrete-time filtering or continuous-time filtering is *continuous-discrete filtering* [16], [17]. In continuous-discrete filtering the state dynamics are modeled as continuous-time stochastic processes, and the measurements are obtained at discrete instances of time. This differs from discrete-time filtering, because in that approach both the dynamics and measurements are modeled as discrete-time processes.

The continuous-discrete filtering models considered here have the general form

$$\begin{aligned} \mathrm{d}x(t) &= f(x(t), t)\,\mathrm{d}t + L(t)\,\mathrm{d}\beta(t) \\ y_k &= h_d(x(t_k), t_k) + r_k, \end{aligned} \tag{4}$$

where

- $x(t) \in \mathbb{R}^n$ is the state,
- $y_k \in \mathbb{R}^m$ is the measurement,
- f is the drift function,
- $L(t)$ is the dispersion matrix,
- $\beta(t)$ is Brownian motion with diffusion matrix $Q_c(t)$,
- h_d is the measurement model function,
- r_k is a zero mean Gaussian measurement noise with covariance matrix R_k.

The dynamic model can be equivalently interpreted as a Itô or Stratonovich type *stochastic differential equation*. As in the continuous-time filtering case the dynamic model can also be written in terms of white noise process $e(t)$ as

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = f(x(t), t) + L(t)\,e(t). \tag{5}$$

The purpose of (Bayesian) continuous-discrete filtering is to recursively compute the posterior distribution

$$p(x(t_k) \,|\, y_1, \ldots, y_k), \tag{6}$$

where t_k is the time of measurement y_k. By using optimal prediction the corresponding distribution can also be computed for all time instances before the next measurement $t \in [t_k, t_{k+1})$.

In theory, the solution to the continuous-discrete filtering problem can be computed by the following prediction and update steps [16]:

- *Prediction step:* solves the predicted probability density at time step t_k from the *Kolmogorov forward partial differential equation* using the old posterior probability density at time step t_{k-1} as the boundary condition.
- *Update step:* uses the *Bayes' rule* for computing the posterior probability density of state at time step t_k from the predicted probability density of the prediction step, and the likelihood of the measurement y_k.

As in the continuous-time case, the closed form solutions to these equations or to the equations for the moments (e.g., mean and covariance) of the distributions can only be found in a few special cases [30]–[32] and approximations are generally needed. The continuous-discrete extended Kalman filter (EKF) [16]–[18] uses a Taylor series expansion approximation to the non-linear drift function and forms a Gaussian process approximation to the SDE. Another possible approach is to simulate sample paths of SDEs [33] and use particle filters for estimation [34], [35]. Interacting and branching particle systems [36] are particle based solutions to nonlinear filtering problems also in the continuous-discrete setting.

Other possible approaches are statistical linearization [17], grid based methods [35], [37] and multiple model methods [7], [35], Gaussian sum approximations [38], [39], and numerical solving of the Kolmogorov forward equation [40], [41].

III. Discrete-Time Unscented Kalman Filtering

A. Unscented Transform

The *unscented transform* (UT) [1]–[3] can be used for forming a Gaussian approximation to the joint distribution of random variables x and y, when the random variable y is obtained by a non-linear transformation of the Gaussian random variable x as follows:

$$\begin{aligned} x &\sim \mathrm{N}(m, P) \\ y &= g(x), \end{aligned} \tag{7}$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, and $g : \mathbb{R}^n \mapsto \mathbb{R}^m$ is a general non-linear function. The idea of UT is to form a fixed number of deterministically chosen sigma-points, which capture the mean and covariance of the original distribution of x exactly. These sigma-points are then propagated through the non-linearity and the mean and covariance of the transformed variable are estimated from them. Note that the unscented transform is significantly different from Monte Carlo estimation, because the sigma points are selected deterministically [1], [2].

Algorithm 3.1 (Unscented transform): The unscented transform can be used for forming the Gaussian approximation

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} m \\ \mu_U \end{pmatrix}, \begin{pmatrix} P & C_U \\ C_U^T & S_U \end{pmatrix} \right), \tag{8}$$

to the joint probability density of $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. The unscented transform is the following:

1) Form the set of $2n+1$ sigma points from the columns of the $n \times n$ matrix $\sqrt{(n+\lambda)\,P}$ as follows:

$$\begin{aligned} x^{(0)} &= m_x \\ x^{(i)} &= m_x + \left[\sqrt{(n+\lambda)\,P}\right]_i, \quad i = 1, \ldots, n \\ x^{(i)} &= m_x - \left[\sqrt{(n+\lambda)\,P}\right]_{i-n}, \quad i = n+1, \ldots, 2n \end{aligned} \tag{9}$$

and compute the associated weights:

$$\begin{aligned} W_0^{(m)} &= \lambda/(n+\lambda) \\ W_0^{(c)} &= \lambda/(n+\lambda) + (1 - \alpha^2 + \beta) \\ W_i^{(m)} &= 1/\{2(n+\lambda)\}, \quad i = 1, \ldots, 2n \\ W_i^{(c)} &= 1/\{2(n+\lambda)\}, \quad i = 1, \ldots, 2n. \end{aligned} \tag{10}$$

The parameter λ is a scaling parameter defined as

$$\lambda = \alpha^2\,(n + \kappa) - n. \tag{11}$$

The positive constants α, β and κ are used as parameters of the method.

2) Transform each of the sigma points as

$$y^{(i)} = g(x^{(i)}), \quad i = 0, \ldots, 2n. \tag{12}$$

3) Mean and covariance estimates for y can be computed as

$$\begin{aligned} \mu_U &\approx \sum_{i=0}^{2n} W_i^{(m)}\, y^{(i)} \\ S_U &\approx \sum_{i=0}^{2n} W_i^{(c)}\, (y^{(i)} - \mu_U)\, (y^{(i)} - \mu_U)^T. \end{aligned} \tag{13}$$

4) The cross-covariance of x and y can estimated as

$$C_U \approx \sum_{i=0}^{2n} W_i^{(c)}\, (x^{(i)} - m)\, (y^{(i)} - \mu_U)^T. \tag{14}$$

The matrix square root of positive definite matrix P means a matrix $A = \sqrt{P}$ such that

$$P = A\,A^T. \tag{15}$$

Because the only requirement for A is the definition above, we can, for example, use the lower triangular matrix of the *Cholesky factorization* (see, e.g., [42]).

In this article, the Cholesky factorization is denoted as the function $\mathrm{chol}(\cdot)$:

$$A = \mathrm{chol}(P). \tag{16}$$

The unscented transform can be seen as a function (or functional) from (g, m, P) to (μ_U, S_U, C_U):

$$[\mu_U, S_U, C_U] = \mathrm{UT}(g, m, P). \tag{17}$$

B. Unscented Kalman Filter

The *unscented Kalman filter* (UKF) [1]–[3] is a discrete-time filtering algorithm, which utilizes the unscented transform for computing approximate solutions to the filtering problems of the form[1]

$$\begin{aligned} x_k &= f_d(x_{k-1}, k-1) + q_{k-1} \\ y_k &= h_d(x_k, k) + r_k, \end{aligned} \tag{18}$$

where $x_k \in \mathbb{R}^n$ is the state, $y_k \in \mathbb{R}^m$ is the measurement, $q_{k-1} \in \mathbb{R}^n$ is a Gaussian process noise $q_{k-1} \sim \mathrm{N}(0, Q_{k-1})$, and $r_k \in \mathbb{R}^m$ is a Gaussian measurement noise $r_k \sim \mathrm{N}(0, R_k)$. The mean and covariance of the initial state x_0 are m_0 and P_0, respectively.

In terms of the unscented transform $\mathrm{UT}(\cdot)$ the *unscented Kalman filter* prediction and update steps can be written as follows:

- *Prediction:* Compute the predicted state mean m_k^- and the predicted covariance P_k^- as

$$\begin{aligned} [m_k^-, \tilde{P}_k] &= \mathrm{UT}(f_d, m_{k-1}, P_{k-1}) \\ P_k^- &= \tilde{P}_k + Q_{k-1}. \end{aligned} \tag{19}$$

- *Update:* Compute the predicted mean μ_k and covariance of the measurement S_k, and the cross-covariance of the state and measurement C_k:

$$\begin{aligned} [\mu_k, \tilde{S}_k, C_k] &= \mathrm{UT}(h_d, m_k^-, P_k^-) \\ S_k &= \tilde{S}_k + R_k. \end{aligned} \tag{20}$$

Then compute the filter gain K_k, the state mean m_k and the covariance P_k, conditional to the measurement y_k:

$$\begin{aligned} K_k &= C_k\, S_k^{-1} \\ m_k &= m_k^- + K_k\, [y_k - \mu_k] \\ P_k &= P_k^- - K_k\, S_k\, K_k^T. \end{aligned} \tag{21}$$

The filtering is started from the initial mean m_0 and covariance P_0. A thorough treatment of the unscented Kalman filtering, sigma-point filtering in general and connections to several other filtering algorithms can be found in [4]. Efficient square root versions of the UKF are presented in [4], [43].

IV. Continuous-Time Unscented Kalman Filtering

A. Matrix Form of Unscented Kalman Filter

In this section the unscented Kalman filter is presented in matrix form, where the weighted sums of sigma points are written as equivalent matrix expressions. This eases the derivations of continuous-time forms of the UKF in the next sections.

In order to clean up the notation, we shall use the convention that expression

$$Y = g(X), \tag{22}$$

where $X \in \mathbb{R}^{n \times d}$ and $g : \mathbb{R}^n \mapsto \mathbb{R}^m$ means that the i:th column Y_i of the matrix $Y \in \mathbb{R}^{m \times d}$ is formed as follows:

$$Y_i = g(X_i). \tag{23}$$

[1]In this article we only consider the case of additive noise, but UKF can also be applied to more general filtering problems with non-additive noise.

Lemma 4.1 (The matrix form of UT): The unscented transform can be written in matrix form as follows:

$$\begin{aligned} X &= \begin{bmatrix} m & \cdots & m \end{bmatrix} + \sqrt{c}\begin{bmatrix} 0 & \sqrt{P} & -\sqrt{P} \end{bmatrix} \\ Y &= g(X) \\ \mu_U &= Y\, w_m \\ S_U &= Y\, W\, Y^T \\ C_U &= X\, W\, Y^T, \end{aligned} \tag{24}$$

where X is the matrix of sigma points, $c = \alpha^2\,(n+\kappa)$, and vector w_m and matrix W are defined as follows:

$$\begin{aligned} w_m &= \begin{bmatrix} W_m^{(0)} & \cdots & W_m^{(2n)} \end{bmatrix}^T \\ W &= \left(I - \begin{bmatrix} w_m & \cdots & w_m \end{bmatrix}\right) \\ &\quad \times \operatorname{diag}(W_c^{(0)} \cdots W_c^{(2n)}) \\ &\quad \times \left(I - \begin{bmatrix} w_m & \cdots & w_m \end{bmatrix}\right)^T. \end{aligned} \tag{25}$$

Proof: See Appendix I-A ■

Algorithm 4.1 (Unscented Kalman filter): The UKF prediction and update steps can be written in matrix form as follows:

- *Prediction:* Compute the predicted state mean m_k^- and the predicted covariance P_k^- as

$$\begin{aligned} X_{k-1} &= \begin{bmatrix} m_{k-1} & \cdots & m_{k-1} \end{bmatrix} \\ &\quad + \sqrt{c}\begin{bmatrix} 0 & \sqrt{P_{k-1}} & -\sqrt{P_{k-1}} \end{bmatrix} \\ \hat{X}_k &= f_d(X_{k-1}, k-1) \\ m_k^- &= \hat{X}_k\, w_m \\ P_k^- &= \hat{X}_k\, W\, [\hat{X}_k]^T + Q_{k-1}. \end{aligned} \tag{26}$$

- *Update:* Compute the predicted mean μ_k and covariance of the measurement S_k, and the cross-covariance of the state and measurement C_k:

$$\begin{aligned} X_k^- &= \begin{bmatrix} m_k^- & \cdots & m_k^- \end{bmatrix} \\ &\quad + \sqrt{c}\begin{bmatrix} 0 & \sqrt{P_k^-} & -\sqrt{P_k^-} \end{bmatrix} \\ Y_k^- &= h_d(X_k^-, k) \\ \mu_k &= Y_k^-\, w_m \\ S_k &= Y_k^-\, W\, [Y_k^-]^T + R_k \\ C_k &= X_k^-\, W\, [Y_k^-]^T. \end{aligned} \tag{27}$$

Then compute the filter gain K_k, the state mean m_k and the covariance P_k:

$$\begin{aligned} K_k &= C_k\, S_k^{-1} \\ m_k &= m_k^- + K_k\, [y_k - \mu_k] \\ P_k &= P_k^- - K_k\, S_k\, K_k^T. \end{aligned} \tag{28}$$

B. Continuous-Time Unscented Kalman Filter

By taking the formal limit of the discrete-time unscented Kalman filter equations in Algorithm 4.1, the following novel continuous-time filter can be derived:

Algorithm 4.2 (Unscented Kalman-Bucy filter): The stochastic differential equations corresponding to the UKF in the continuous-time limit of the state and measurement processes, that is, the *unscented Kalman-Bucy filter (UKBF)* equations, are given as

$$\begin{aligned} K(t) &= X(t)\, W\, h^T(X(t),t)\, \left[V(t)\, R_c(t)\, V^T(t)\right]^{-1} \\ \frac{\mathrm{d}m(t)}{\mathrm{d}t} &= f(X(t),t)\, w_m + K(t)\, \left[z(t) - h(X(t),t)\, w_m\right] \\ \frac{\mathrm{d}P(t)}{\mathrm{d}t} &= X(t)\, W\, f^T(X(t),t) + f(X(t),t)\, W\, X^T(t) \\ &\quad + L(t)\, Q_c(t)\, L^T(t) \\ &\quad - K(t)\, V(t)\, R_c(t)\, V^T(t)\, K^T(t), \end{aligned} \tag{29}$$

where we have formally defined the differential measurement $z(t) = \mathrm{d}y(t)/\mathrm{d}t$. The sigma-point matrix $X(t)$ is defined as

$$X(t) = \begin{bmatrix} m(t) & \cdots & m(t) \end{bmatrix} + \sqrt{c}\,\begin{bmatrix} 0 & \sqrt{P(t)} & -\sqrt{P(t)} \end{bmatrix}. \tag{30}$$

Proof: See Appendix I-B. ■

To avoid problems related to the finite numerical precision of computer arithmetic, Kalman filter equations are often implemented such that the matrix square roots of covariance matrices are used in computations instead of their actual values [18]. Because UKF already uses matrix square roots in its sigma-points, the square root version of continuous-time UKF can be obtained by formulating the filter as a differential equation for the sigma points.

Algorithm 4.3 (Square root unscented Kalman-Bucy filter): The unscented Kalman-Bucy filter can be formulated in terms of sigma points as follows:[2]

$$\begin{aligned} K(t) &= X(t)\, W\, h^T(X(t),t)\, \left[V(t)\, R_c(t)\, V^T(t)\right]^{-1} \\ M(t) &= A^{-1}(t)\, \Big[X(t)\, W\, f^T(X(t),t) \\ &\quad + f(X(t),t)\, W\, X^T(t) \\ &\quad + L(t)\, Q_c(t)\, L^T(t) \\ &\quad - K(t)\, V(t)\, R_c(t)\, V^T(t)\, K^T(t)\Big]\, A^{-T}(t) \\ \frac{\mathrm{d}X_i(t)}{\mathrm{d}t} &= f(X(t),t)\, w_m + K(t)\, \left[z(t) - h(X(t),t)\, w_m\right] \\ &\quad + \sqrt{c}\,\Big[0 \quad A(t)\,\Phi\Big(M(t)\Big) \quad -A(t)\,\Phi\Big(M(t)\Big)\Big]_i \end{aligned} \tag{31}$$

where the matrix of sigma-points is defined as

$$X(t) = \begin{bmatrix} m(t) & \cdots & m(t) \end{bmatrix} + \sqrt{c}\,\begin{bmatrix} 0 & A(t) & -A(t) \end{bmatrix}, \tag{32}$$

and $\Phi(\cdot)$ is a function returning the lower diagonal part of the argument as follows:

$$\Phi_{ij}\Big(M(t)\Big) = \begin{cases} M_{ij}(t) & , \text{ if } i > j \\ \frac{1}{2} M_{ij}(t) & , \text{ if } i = j \\ 0 & , \text{ if } i < j. \end{cases} \tag{33}$$

The initial $A(0)$ should be the lower triangular Cholesky factor of the initial covariance matrix $A(0) = \operatorname{chol}(P(0))$.

Proof: See Appendix I-D. ■

Note that in the square root filter, the equations contain the matrix $A(t)$, which is the lower triangular Cholesky factor of the covariance $A(t) = \operatorname{chol}(P(t))$ at all time instances $t \geq 0$.

[2]Here $()_i$ denotes the ith column of the argument matrix.

However, the factorization needs to be explicitly computed only on the initial time step, because $A(t)$ can be extracted from $X(t)$ at any time step by very simple means. This also implies that in the implementation of the algorithm the covariance $P(t)$ of the state never needs to be evaluated.

The unscented Kalman-Bucy filter equations presented in this section are actually Stratonovich type of stochastic differential equations, because they have been derived by taking the continuous-time limit of discrete-time equations by using the rules of conventional calculus. In the beginning of the paper it was stated that the continuous-time dynamic and measurement models can be equivalently interpreted as either Itô or Stratonovich type of stochastic differential equations, which is indeed true, but with both the interpretations the filter equations still are of the Stratonovich type.

C. Unscented Continuous-Time Prediction

In this section the continuous-time prediction equations are derived as special cases of the continuous-time filters in the previous section. The motivation is that using the prediction equations we can construct the continuous-discrete versions of the unscented Kalman filter.

Algorithm 4.4 (Mean and covariance prediction): The predicted mean $m(t)$ and covariance $P(t)$ of the state for times $t \geq t_0$ given the mean and covariance at the time instance t_0, that is, $m(t_0)$ and $P(t_0)$ can be computed by integrating the differential equations

$$\begin{aligned}\frac{\mathrm{d}m(t)}{\mathrm{d}t} &= f(X(t),t)\, w_m \\ \frac{\mathrm{d}P(t)}{\mathrm{d}t} &= X(t)\, W\, f^T(X(t),t) + f(X(t),t)\, W\, X^T(t) \\ &\quad + L(t)\, Q_c(t)\, L^T(t),\end{aligned} \tag{34}$$

from initial conditions $m(t_0)$ and $P(t_0)$ to time instance t. Here $X(t)$ is defined as in Equation (30),

Proof: Formally set $R_c(t) = \infty I$ in Algorithm 4.2, which results in $K(t) = 0$. ∎

Algorithm 4.5 (Square root prediction): The UKF prediction equations can be written in terms of sigma points as

$$\begin{aligned}M(t) &= A^{-1}(t)\,[X(t)\, W\, f^T(X(t),t) \\ &\quad + f(X(t),t)\, W\, X^T(t) \\ &\quad + L(t)\, Q_c(t)\, L^T(t)]\, A^{-T}(t) \\ \frac{\mathrm{d}X_i(t)}{\mathrm{d}t} &= f(X(t),t)\, w_m \\ &\quad + \sqrt{c}\, \Big[0 \quad A(t)\, \Phi\Big(M(t)\Big) \quad -A(t)\, \Phi\Big(M(t)\Big)\Big]_i,\end{aligned} \tag{35}$$

where $A(t)$, $X(t)$ and $\Phi(\cdot)$ are defined as in Algorithm 4.3. The integration is started from the sigma points $X(t_0)$, which are generated from $m(t_0)$ and $P(t_0)$.

Proof: Formally set $R_c(t) = \infty I$ in Algorithm 4.3, which results in $K(t) = 0$. ∎

D. Continuous-Discrete Unscented Kalman Filter

The mean and covariance form of continuous-discrete UKF can be now implemented as follows:

Algorithm 4.6 (Continuous-discrete UKF): The prediction and update steps of the continuous-discrete unscented Kalman filter are the following:

- *Prediction.* Integrate the differential equations (34) in Algorithm 4.4 from the initial conditions $m(t_{k-1}) = m_{k-1}$, $P(t_{k-1}) = P_{k-1}$ to time instance t_k. The predicted mean and covariance are given as $m_k^- = m(t_k)$ and $P_k^- = P(t_k)$, respectively.
- *Update.* The update step is the same as the discrete-time unscented Kalman filter update step (27) of Algorithm 4.1.

The corresponding square root continuous-discrete UKF has the same form:

Algorithm 4.7 (Square root continuous-discrete UKF): The prediction and update steps of the square root continuous-discrete unscented Kalman filter are the following:

- *Prediction.* Integrate the differential equations (35) in Algorithm 4.5 from the initial conditions $X(t_{k-1})$, which are the sigma points generated from $m(t_{k-1}) = m_{k-1}$ and $P(t_{k-1}) = P_{k-1}$. The integration is continued up to the time instance t_k and the predicted sigma points are given as $X_k^- = X(t_k)$.
- *Update.* The update step is the same as the discrete-time unscented Kalman filter update step (27) of Algorithm 4.1, but now the predicted sigma points do not have to be generated from mean and covariance, because they are already available. To enhance the numerical stability of the update equations, square root versions [4], [43] of the UKF update step can also be used.

V. Remarks and Discussion

A. Computational Complexity

Although, the new matrix form of the unscented transform is very useful in derivation of the filtering equations, in numerical computations, the classical form can be computationally more efficient. The matrix expressions appearing in the filtering equations can be equivalently written as summations as follows:

$$\begin{aligned}&X(t)\, W\, f^T(X(t),t) \\ &\quad = \sum_{i=0}^{2n} W_i^{(c)}\, (x^{(i)} - m_x)\, (f(x^{(i)},t) - m_f)^T \\ &f(X(t),t)\, W\, X^T(t) \\ &\quad = \sum_{i=0}^{2n} W_i^{(c)}\, (f(x^{(i)},t) - m_f)\, (x^{(i)} - m_x)^T \\ &X(t)\, W\, h^T(X(t),t) \\ &\quad = \sum_{i=0}^{2n} W_i^{(c)}\, (x^{(i)} - m_x)\, (h(x^{(i)},t) - m_h)^T,\end{aligned} \tag{36}$$

where

$$\begin{aligned} m_x &= \sum_{i=0}^{2n} W_i^{(m)}\, x^{(i)} \\ m_f &= \sum_{i=0}^{2n} W_i^{(m)}\, f(x^{(i)}, t) \\ m_h &= \sum_{i=0}^{2n} W_i^{(m)}\, h(x^{(i)}, t). \end{aligned} \tag{37}$$

By using the summation forms of the covariance terms the full matrix product is replaced by the weighted sum of outer products, which is lighter to compute.

The relationship between the unscented Kalman-Bucy filter and extended Kalman-Bucy filter approximations can be seen to be

$$\begin{aligned} X(t)\, W\, h^T(X(t),t) &\Longleftrightarrow P(t)\, H^T(m(t),t) \\ X(t)\, W\, f^T(X(t),t) &\Longleftrightarrow P(t)\, F^T(m(t),t) \\ f(X(t),t)\, w_m &\Longleftrightarrow f(m(t),t) \\ h(X(t),t)\, w_m &\Longleftrightarrow h(m(t),t), \end{aligned} \tag{38}$$

where the elements of the Jacobian matrices $F(m(t),t)$ and $H(m(t),t)$ are given as

$$\begin{aligned} F_{ij}(m(t),t) &= \left.\frac{\mathrm{d}f_i(x,t)}{\mathrm{d}x_j}\right|_{x=m(t)} \\ H_{ij}(m(t),t) &= \left.\frac{\mathrm{d}h_i(x,t)}{\mathrm{d}x_j}\right|_{x=m(t)}. \end{aligned} \tag{39}$$

The computational complexity of the unscented Kalman-Bucy filter (or continuous-discrete UKF) can be seen to be 2-3 times the computational complexity of the extended Kalman-Bucy filter, when compared in terms of number of multiplications and additions. When the state dimension is n, the UKBF needs $2n+1$ evaluations of f and h, when EKBF needs only one. However, in addition to that, EKBF needs evaluations of the Jacobian matrices (and possibly Hessian matrices) of both the functions.

VI. Illustrative Examples

A. Continuous-Time Non-Linear Filtering

In this section we shall test the performance of the unscented Kalman-Bucy filter in continuous-time version of the non-linear filtering problem, which was used in [17] for demonstrating the performance of the statistical linearization based filter. Both the state $x(t)$ and the measurements $z(t)$ are one-dimensional continuous-time processes and the filtering model is:

$$\begin{aligned} dx(t)/dt &= -\sin x(t) + w(t) \\ z(t) &= \frac{1}{2}\sin(x(t)) + n(t), \end{aligned} \tag{40}$$

where $w(t)$ and $n(t)$ are continuous-time white noise processes with spectral densities $q_c = 0.01$ and $r_c = 0.004$, respectively. Note that in [17] the dynamic model was in continuous time and the measurement model was in discrete time, but here both the models are in continuous time. The simulation was performed over time period of 5 seconds using the Euler-Maruyama scheme [33] and with time steps of $\Delta t = 0.01$.

TABLE I

RMSE values of the parameter and signal averaged over 1000 Monte Carlo runs in the continuous-time non-linear filtering problem.

Filter	*RMSE[x]*
EKBF	0.22
EKBF2	0.20
UKBF	0.16
SLF	0.13

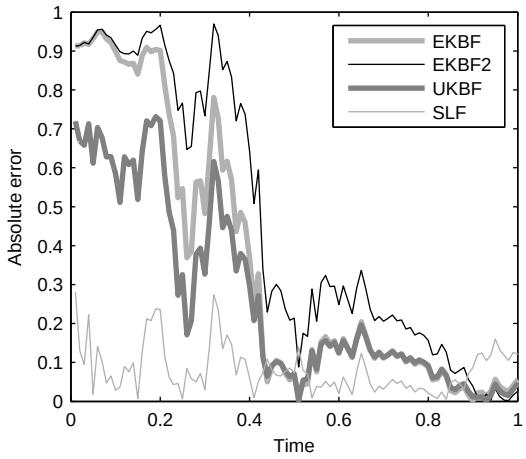

Fig. 1. Illustration of transient period error behavior in continuous-time non-linear filtering problem.

The root mean squared error (RMSE) results of 1000 Monte Carlo runs with extended Kalman-Bucy filter (EKBF), second order extended Kalman-Bucy filter (EKBF2), unscented Kalman-Bucy filter (UKBF) and statistically linearized filter (SLF) are shown in Table I. It can be seen that SLF is the best of the filters in RMSE sense, because it uses the closed form formulas for expectations and covariances of the non-linearities (see [17]), when the other filters can only approximate them. The UKBF gives a significantly better result than EKBF and EKBF2, and the UKBF result is also quite close to the base line result of SLF. The performances of EKBF and EKBF2 are most likely that bad because the Taylor series expansion based approximations do not work well when the estimation error is large.

Even thought the SLF is better than UKBF in RMSE sense, SLF has the serious disadvantage that in order to implement it, one has to be able to compute closed form formulas for certain expected values of non-linear transformations of Gaussian random variables [17]. These expected values can be computed in closed form only in simple special cases and thus UKBF is a very good choice for models with significant non-linearities and uncertainties.

A typical transient behavior of the filters is illustrated in Figure 1. It can be seen that SLF is best because it converges

very quickly near the correct state. The convergence of UKBF is slower, but significantly faster than of EKBF and EKBF2.

B. Re-Entry Vehicle Tracking

Here we consider the reentry tracking problem, where a radar is used for tracking a space vehicle, which enters the atmosphere at a very high speed. The reentry problem was used for demonstrating the performance of UKF in [2] and slight corrections to the equations and simulation parameters were later published in [44].

The stochastic equations of motion for the space vehicle are given as [2], [44]

$$\begin{aligned} R(t) &= \sqrt{x_1^2(t) + x_2^2(t)} \\ V(t) &= \sqrt{x_3^2(t) + x_4^2(t)} \\ b(t) &= b_0 \, \exp\left(x_5(t)\right) \\ D(t) &= b(t) \, \exp\left(\frac{R_0 - R}{H_0}\right) V(t) \\ G(t) &= -\frac{G m_0}{R^3(t)} \\ \mathrm{d}x_1/\mathrm{d}t &= x_3(t) \\ \mathrm{d}x_2/\mathrm{d}t &= x_4(t) \\ \mathrm{d}x_3/\mathrm{d}t &= D(t)\, x_3(t) + G(t)\, x_1(t) + w_1(t) \\ \mathrm{d}x_4/\mathrm{d}t &= D(t)\, x_4(t) + G(t)\, x_2(t) + w_2(t) \\ \mathrm{d}x_5/\mathrm{d}t &= w_3(t), \end{aligned} \tag{41}$$

where $w_1(t)$, $w_2(t)$, $w_3(t)$ are white Gaussian process noises with known joint spectral density. The constants are [2]:

$$\begin{aligned} b_0 &= -0.59783 \\ H_0 &= 13.406 \\ G m_0 &= 3.9860 \times 10^5 \\ R_0 &= 6374. \end{aligned} \tag{42}$$

In the article [2] the radar measured 10 times per second and the simulated discrete process noise covariance was

$$Q(k) = \begin{bmatrix} 2.4064 \times 10^{-5} & 0 & 0 \\ 0 & 2.4064 \times 10^{-5} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{43}$$

If we interpret the dynamic model (41) as originally continuous time model, we may assume that the discretized covariance is actually an approximation to a non-linear continuous-time process driven by continuous-time process noise with spectral density Q_c. In this interpretation it is reasonable to assume that the relationship between the discrete covariance and the continuous-time spectral density is originally the approximation $Q(k) \approx Q_c \, \Delta t$ where $\Delta t = 0.1\,\mathrm{s}$ is the sampling time used in [2]. Thus the true spectral density matrix is the matrix in (43) multiplied by 10.

In the actual filter implementation the lower right corner term in the modeled process noise covariance was set to the value 10^{-6}. This additional small noise term was used for the constant parameter to enhance the filter stability as was done in [44]. This modified discrete process noise covariance is denoted as $Q'(k)$ and the corresponding spectral density is denoted as Q'_c.

The radar is located at $(x_r, y_r) = (R_0, 0)$ and the measurement model is

$$\begin{aligned} r_k &= \sqrt{(x_1(t_k) - x_r)^2 + (x_2(t_k) - y_r)^2} + e_k^r \\ \theta_k &= \tan^{-1}\left(\frac{x_2(t_k) - y_r}{x_1(t_k) - x_r}\right) + e_k^a, \end{aligned} \tag{44}$$

where the $e_k^r \sim \mathrm{N}(0, \sigma_r^2)$ and $e_k^a \sim \mathrm{N}(0, \sigma_a^2)$.

In the simulation, the assumed means and covariances of the state, simulated means and covariances of the state and the standard deviations of measurements were selected to be the same as in [2], [44]. The simulated data were generated by simulating the stochastic differential equation (41) with 100 steps of Euler-Maruyama scheme [33] between each measurement.

The discrete UKF (DUKF) and continuous-discrete UKF (CDUKF) were both implemented using 10 steps of Runge-Kutta integration between measurements, but the difference was in handling of process noise:

- In DUKF, each measurement is processed as follows:
  1) Integrate each of the sigma point through the noise free dynamic model using 10 steps of the Runge-Kutta integration.
  2) Compute the predicted mean and covariance, and model process noise effect by approximating the discrete covariance by $Q'(k) \approx Q'_c \, \Delta t$.
  3) Perform standard UKF update step for the measurement.
- In CDUKF, each measurement is processed as follows:
  1) Integrate the mean and covariance differential equations using 10 steps of the Runge-Kutta integration, and using Q'_c as the diffusion matrix (or spectral density) of the process noise.
  2) Perform standard UKF update step for the measurement.

The amount of computations required by the CDUKF is slightly higher than of DUKF, but the number of evaluations of the dynamic model function is the same for both the models and thus the practical total difference is small. The algorithm parameters in the unscented transforms were selected to be $\alpha = 1/2, \beta = 2, \kappa = -2$.

Simulations were performed using different time steps of $\Delta t = 0.1, 0.2, \ldots, 3$ seconds and the results from 100 Monte Carlo simulations per step size are shown in Figure 2. The discrete UKF (DUKF) and continuous-discrete UKF (CDUKF) have very much the same performance when the time step size is short. However, when the time step grows, DUKF encounters numerical problems and its error grows rapidly. The problem is that when the sigma-points are integrated one at a time, nothing prevents the covariance estimate from becoming non-positive definite. With larger time steps this causes severe numerical difficulties. At the same time, no numerical problems can be seen in the CDUKF and its error grows much slower.

In this simulation scenario the advantage of the continuous-time formulation over the discrete-time formulation is the

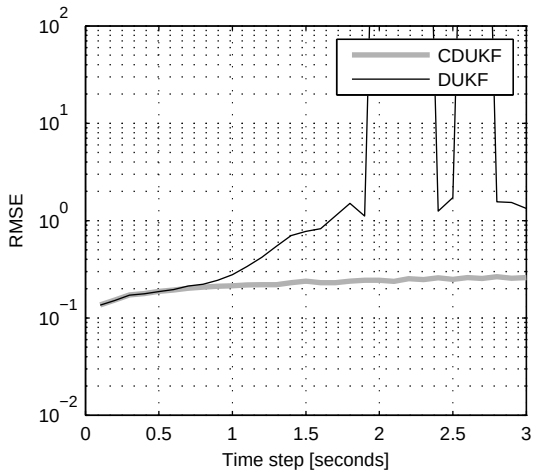

Fig. 2. Root mean squared error (RMSE) versus time step size in the reentry filtering problem. Results are from 100 Monte Carlo runs per time step with the continuous-discrete UKF (CDUKF) and discrete UKF (DUKF). The higher errors of DUKF with longer time step sizes are caused by numerical problems on the prediction step.

numerical stability. This stability is due to that when the noise process is modeled as a continuous-time process the non-linear dynamic model cannot force the covariance to become non-positive definite. In the discrete-time formulation the covariance can become non-positive definite.

In this particular problem it is not essential whether uncertainties are modeled as discrete-time or continuous-time stochastic processes. Since the performance of CDUKF seems to be at least that of DUKF, in cases where the continuous-time stochastic process formulation is more accurate in modeling point of view, the continuous-discrete UKF is likely to perform better than the discrete UKF.

VII. Conclusion

In this article novel continuous-time and continuous-discrete versions of the originally discrete-time unscented Kalman filter (UKF) have been derived and applied to non-linear continuous-time filtering and re-entry vehicle tracking problems. Numerically more stable square-root versions of the new filters have also been derived.

The continuous-time and continuous-discrete unscented filters are good alternatives to the extended Kalman filters in models, where the Jacobian and Hessian matrices of the drift terms are not available or when the Taylor series expansion approximations do not work well. According to the simulations the approximations generated by the new filters are better in the cases where the model non-linearities and estimation uncertainties are significant.

Acknowledgment

The author would like to thank Aki Vehtari, Jouko Lampinen, Ilkka Kalliomäki, Hannes Gruschinski and the anonymous reviewers for helpful comments.

Appendix I
Derivations

A. Derivation of Matrix Form of UT

If we define the matrix of sigma points as

$$X = \begin{bmatrix} x^{(0)} & \cdots & x^{(2n)} \end{bmatrix}, \tag{45}$$

then the sigma point computation in equations (9) can be written in form of the first equation in (24). The second equation is simply the vector form of Equation (12).

If we define the weight vector w_m and matrix W_c as in Equation (25) and denote the matrix of sigma points of y as Y the transformed mean and covariance equations can be written as

$$\begin{aligned} \mu_U &= \sum_i W_m^{(i)}\, y^{(i)} \\ &= Y\, w_m \end{aligned} \tag{46}$$

$$\begin{aligned} S_U &= \sum_i W_c^{(i)}\, (y^{(i)} - \mu_U)\, (y^{(i)} - \mu_U)^T \\ &= \sum_i W_c^{(i)}\, (y^{(i)} - Y w_m)\, (y^{(i)} - Y w_m)^T \\ &= \big(Y - Y \begin{bmatrix} w_m & \cdots & w_m \end{bmatrix}\big) \\ &\quad \times \operatorname{diag}(W_c^{(0)} \cdots W_c^{(2n)}) \\ &\quad \times \big(Y - Y \begin{bmatrix} w_m & \cdots & w_m \end{bmatrix}\big)^T \\ &= Y\, W\, Y^T \end{aligned} \tag{47}$$

$$\begin{aligned} C_U &= \sum_i W_c^{(i)}\, (x^{(i)} - m)\, (y^{(i)} - \mu_U)^T \\ &= \sum_i W_c^{(i)}\, (x^{(i)} - X w_m)\, (y^{(i)} - Y w_m)^T \\ &= \big(X - X \begin{bmatrix} w_m & \cdots & w_m \end{bmatrix}\big) \\ &\quad \times \operatorname{diag}(W_c^{(0)} \cdots W_c^{(2n)}) \\ &\quad \times \big(Y - Y \begin{bmatrix} w_m & \cdots & w_m \end{bmatrix}\big)^T \\ &= X\, W\, Y^T, \end{aligned} \tag{48}$$

which leads to last three equations in (24).

B. Derivation of UKBF

The continuous-time unscented Kalman filter equations can be derived from the discrete-time UKF by the same limiting procedure as is commonly used in derivation of the Kalman-Bucy filter from the discrete-time Kalman filter (see, e.g., [14], [18]). The continuous-time filtering model

$$\begin{aligned} \mathrm{d}x(t) &= f(x(t), t)\, \mathrm{d}t + L(t)\, \mathrm{d}\beta(t) \\ \mathrm{d}y(t) &= h(x(t), t)\, \mathrm{d}t + V(t)\, \mathrm{d}\eta(t), \end{aligned} \tag{49}$$

where $\beta(t)$ and $\eta(t)$ are independent Brownian motions with diagonal diffusion matrices $Q_c(t)$ and $R_c(t)$ can be interpreted such that when Δt is sufficiently close to zero, we have

$$\begin{aligned} x(t + \Delta t) - x(t) &= f(x(t), t)\, \Delta t + L(t)\, \Delta\beta + o(\Delta t) \\ y(t + \Delta t) - y(t) &= h(x(t), t)\, \Delta t + V(t)\, \Delta\eta + o(\Delta t), \end{aligned} \tag{50}$$

where $\Delta\beta \sim \mathrm{N}(0, Q_c(t)\, \Delta t)$, $\Delta\eta \sim \mathrm{N}(0, R_c(t)\, \Delta t)$ and $o(\Delta t)$ is a function such that $o(\Delta t)/\Delta t \to 0$ when $\Delta t \to 0$.

Assuming that $x(t) \sim \mathrm{N}(m(t), P(t))$, the UKF prediction from $x(t)$ to $x(t+\Delta t)$ can be written as

$$\begin{aligned} X(t) &= \begin{bmatrix} m(t) & \cdots & m(t) \end{bmatrix} \\ &\quad + \sqrt{c}\begin{bmatrix} 0 & \sqrt{P(t)} & -\sqrt{P(t)} \end{bmatrix} \\ \hat{X}(t+\Delta t) &= X(t) + f(X(t),t)\,\Delta t + o(\Delta t) \\ m^-(t+\Delta t) &= \hat{X}(t+\Delta t)\, w_m \\ P^-(t+\Delta t) &= \hat{X}(t+\Delta t)\, W\, \hat{X}(t+\Delta t) \\ &\quad + L(t)\, Q_c(t)\, L^T(t)\,\Delta t. \end{aligned} \tag{51}$$

Substituting $\hat{X}(t+\Delta t)$ into the equations of $m^-(t+\Delta t)$ and $P^-(t+\Delta t)$, and using the identities $m(t) = X(t)\, w_m$, $P(t) = X(t)\, W\, X^T(t)$ gives

$$\begin{aligned} m^-(t+\Delta t) &= m(t) + f(X(t),t)\, w_m\,\Delta t + o(\Delta t) \\ P^-(t+\Delta t) &= P(t) + f(X(t),t)\, W\, X^T(t)\,\Delta t \\ &\quad + X(t)\, W\, f^T(X(t),t)\,\Delta t \\ &\quad + L(t)\, Q_c(t)\, L^T(t)\,\Delta t + o(\Delta t). \end{aligned} \tag{52}$$

Assuming that we measure the difference $\Delta y = y(t+\Delta t) - y(t)$ the UKF update step can be written as

$$\begin{aligned} X^-(t+\Delta t) &= \begin{bmatrix} m^-(t+\Delta t) & \cdots & m^-(t+\Delta t) \end{bmatrix} \\ &\quad + \sqrt{c}\begin{bmatrix} 0 & \sqrt{P^-(t+\Delta t)} & -\sqrt{P^-(t+\Delta t)} \end{bmatrix} \\ Y^-(t+\Delta t) &= h(X^-(t+\Delta t),t)\,\Delta t + o(\Delta t) \\ \mu(t+\Delta t) &= Y^-(t+\Delta t)\, w_m \\ S(t+\Delta t) &= Y^-(t+\Delta t)\, W\, [Y^-(t+\Delta t)]^T \\ &\quad + V(t)\, R_c(t)\, V^T(t)\,\Delta t \\ C(t+\Delta t) &= X^-(t+\Delta t)\, W\, [Y^-(t+\Delta t)]^T \\ K(t+\Delta t) &= C(t+\Delta t)\, S^{-1}(t+\Delta t) \\ m(t+\Delta t) &= m^-(t+\Delta t) \\ &\quad + K(t+\Delta t)\, [\Delta y - \mu(t+\Delta t)] \\ P(t+\Delta t) &= P^-(t+\Delta t) \\ &\quad - K(t+\Delta t)\, S(t+\Delta t)\, K^T(t+\Delta t). \end{aligned} \tag{53}$$

Substituting $Y^-(t+\Delta t)$, $\mu(t+\Delta t)$, $C(t+\Delta t)$, $S(t+\Delta t)$, and retaining only first order terms results in

$$\begin{aligned} K(t+\Delta t) &= X^-(t+\Delta t)\, W\, h^T(X^-(t+\Delta t),t) \\ &\quad \times [V(t)\, R_c(t)\, V^T(t)]^{-1} + o(\Delta t)/\Delta t \\ m(t+\Delta t) &= m^-(t+\Delta t) + K(t+\Delta t) \\ &\quad \times \left[\Delta y - h(X^-(t+\Delta t),t)\, w_m\,\Delta t\right] \\ &\quad + o(\Delta t) \\ P(t+\Delta t) &= P^-(t+\Delta t) - K(t+\Delta t) \\ &\quad \times V(t)\, R_c(t)\, V^T(t)\, K^T(t+\Delta t)\,\Delta t \\ &\quad + o(\Delta t). \end{aligned} \tag{54}$$

Substituting $m^-(t+\Delta t)$ and $P^-(t+\Delta t)$ gives

$$\begin{aligned} m(t+\Delta t) &= m(t) + f(X(t),t)\, w_m\,\Delta t + K(t+\Delta t) \\ &\quad \times \left[\Delta y - h(X^-(t+\Delta t),t)\, w_m\,\Delta t\right] \\ &\quad + o(\Delta t) \\ P(t+\Delta t) &= P(t) + f(X(t),t)\, W\, X^T(t)\,\Delta t \\ &\quad + X(t)\, W\, f^T(X(t),t)\,\Delta t \\ &\quad + L(t)\, Q_c(t)\, L^T(t)\,\Delta t - K(t+\Delta t) \\ &\quad \times V(t)\, R_c(t)\, V^T(t)\, K^T(t+\Delta t)\,\Delta t \\ &\quad + o(\Delta t). \end{aligned} \tag{55}$$

Dividing by Δt and taking the limit $\Delta t \to 0$ results in the differential equations in the Algorithm 4.2.

C. Alternative Derivation

In Appendix I-B the equations of the unscented Kalman-Bucy filter were derived by taking the formal limit of the discrete-time equations. However, it is also possible to derive the same equations purely in continuous-time framework. The mean and covariance of the optimal continuous-time filter are, in theory, given by the stochastic differential equations [16]

$$\begin{aligned} \mathrm{d}m &= \mathrm{E}[f(x,t)]\,\mathrm{d}t + \left(\mathrm{E}[x\, h^T(x,t)] - \mathrm{E}[x]\ \mathrm{E}[h^T(x,t)]\right) \\ &\quad \times \left(V(t)\, R_c(t)\, V^T(t)\right)^{-1} [\mathrm{d}y(t) - \mathrm{E}[h(x,t)]\,\mathrm{d}t] \\ \mathrm{d}P_{ij} &= \Big\{ \mathrm{E}[x_i\, f_j(x,t)] - \mathrm{E}[x_i]\ \mathrm{E}[f_j(x,t)] + \mathrm{E}[f_i(x,t)\, x_j] \\ &\quad - \mathrm{E}[f_i(x,t)]\ \mathrm{E}[x_j] + \left[L(t)\, Q_c(t)\, L^T(t)\right]_{ij} \\ &\quad - \left(\mathrm{E}[x_i\, h^T(x,t)] - \mathrm{E}[x_i]\ \mathrm{E}[h^T(x,t)]\right) \\ &\quad \times \left(V(t)\, R_c(t)\, V^T(t)\right)^{-1} \\ &\quad\quad \times \left(\mathrm{E}[h(x,t)\, x_j] - \mathrm{E}[h(x,t)]\ \mathrm{E}[x_j]\right) \Big\}\mathrm{d}t \\ &\quad + \Big\{ \mathrm{E}[x_i\, x_j\, h^T(x,t)] - \mathrm{E}[x_i\, x_j]\ \mathrm{E}[h^T(x,t)] \\ &\quad - \mathrm{E}[x_i]\ \mathrm{E}[x_j\, h^T(x,t)] - \mathrm{E}[x_j]\ \mathrm{E}[x_i\, h^T(x,t)] \\ &\quad + 2\,\mathrm{E}[x_j]\ \mathrm{E}[x_i]\ \mathrm{E}[h^T(x,t)] \Big\} \\ &\quad\quad \times \left(V(t)\, R_c(t)\, V^T(t)\right)^{-1} [\mathrm{d}y(t) - \mathrm{E}[h(x,t)]\,\mathrm{d}t]\,. \end{aligned} \tag{56}$$

The expectations are with respect to the posterior distribution of the state $x(t)$ and thus cannot be in practice computed, because the computation would require the knowledge of all (infinite number of) moments of the distribution. However, an unscented transform based approximation to these equations can be formed as follows.

If we assume that the third order term in the covariance equation above is approximately zero, the equations above can be approximately written as

$$\begin{aligned} \frac{\mathrm{d}m}{\mathrm{d}t} &= \mathrm{E}[f(x,t)] + \mathrm{Cov}[x, h(x,t)] \\ &\quad \times \left(V(t)\, R_c(t)\, V^T(t)\right)^{-1} [z(t) - \mathrm{E}[h(x,t)]] \\ \frac{\mathrm{d}P}{\mathrm{d}t} &= \mathrm{Cov}[x, f(x,t)] + \mathrm{Cov}[f(x,t), x] + L(t)\, Q_c(t)\, L^T(t) \\ &\quad - \mathrm{Cov}[x, h(x,t)] \left(V(t)\, R_c(t)\, V^T(t)\right)^{-1} \mathrm{Cov}[h(x,t), x], \end{aligned} \tag{57}$$

where $\mathrm{Cov}[x,y] = \mathrm{E}[x\,y^T] - \mathrm{E}[x]\,\mathrm{E}[y]^T$ denotes the cross-covariance of x and y, and $z(t) = \mathrm{d}y(t)/\mathrm{d}t$ is the differentiated measurement process. If we assume that the posterior mean and covariance of $x(t)$ are $m(t)$ and $P(t)$, respectively, unscented transform based approximations to the expectations and covariances can be formed as follows:

$$\begin{aligned}
\mathrm{E}[f(x,t)] &\approx f(X(t),t)\,w_m \\
\mathrm{E}[h(x,t)] &\approx h(X(t),t)\,w_m \\
\mathrm{Cov}[x, f(x,t)] &\approx X(t)\,W\,f^T(X(t),t) \\
\mathrm{Cov}[f(x,t), x] &\approx f(X(t),t)\,W\,X^T(t) \\
\mathrm{Cov}[x, h(x,t)] &\approx X(t)\,W\,h^T(X(t),t) \\
\mathrm{Cov}[h(x,t), x] &\approx h^T(X(t),t)\,W\,X^T(t),
\end{aligned} \tag{58}$$

where $X(t)$ are the sigma-points generated from $m(t)$ and $P(t)$. Substituting these approximations into the Equations (57) results in the unscented Kalman-Bucy filter equations in Algorithm 4.2.

Note that the above derivation is not strictly complete as such, because the UKBF equations presented in Section IV-B are actually Stratonovich equations, but the Equations (57) are Itô type of equations. However, we may argue that the extra terms arising from the conversion of Equations (57) from Itô from to Stratonovich form would only introduce terms involving odd moments, which anyway disappear due to the Gaussian approximation.

D. Derivation of SR-UKBF

Assume that the matrix square root used is the Cholesky factorization

$$A(t) = \mathrm{chol}(P(t)), \tag{59}$$

which also assures that the square root is lower triangular. If we define (symmetric matrix)

$$\begin{aligned}
U(X(t),t) = {} & X(t)\,W\,f^T(X(t),t) + f(X(t),t)\,W\,X^T(t) \\
& + L(t)\,Q_c(t)\,L^T(t) \\
& - K(t)\,V(t)\,R_c(t)\,V^T(t)\,K^T(t),
\end{aligned} \tag{60}$$

then the differential equation of the covariance is of the form

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = U(X(t),t). \tag{61}$$

It is now possible to derive differential equations for the Cholesky factor $A(t)$ using the procedure presented in [45] and later in [14]. We first expand the derivative of the covariance matrix as follows:

$$\frac{\mathrm{d}A(t)}{\mathrm{d}t}\,A^T(t) + A\,\frac{\mathrm{d}A^T(t)}{\mathrm{d}t} = U(X(t),t). \tag{62}$$

Multiplying both sides from left by A^{-1} and from right by A^{-T} gives

$$A^{-1}(t)\frac{\mathrm{d}A(t)}{\mathrm{d}t} + \frac{\mathrm{d}A^T(t)}{\mathrm{d}t}A^{-T}(t) = A^{-1}(t)U(X(t),t)A^{-T}(t). \tag{63}$$

Now the left hand side is sum of lower triangular matrix $A^{-1}(t)\,\frac{\mathrm{d}A(t)}{\mathrm{d}t}$ and upper triangular matrix $\frac{\mathrm{d}A^T(t)}{\mathrm{d}t}\,A^{-T}(t)$, and the right hand side is symmetric matrix. Thus it can be concluded that the derivative of $A(t)$ can be written as

$$\frac{\mathrm{d}A(t)}{\mathrm{d}t} = A\,\Phi\Big(A^{-1}(t)\,U(X(t),t)\,A^{-T}(t)\Big), \tag{64}$$

where the function Φ, which is defined in Equation (33) gives the lower diagonal part and half of the diagonal part of its argument. The expression of the sigma points is

$$X(t) = \begin{bmatrix} m(t) & \cdots & m(t) \end{bmatrix} + \sqrt{c}\begin{bmatrix} 0 & A(t) & -A(t) \end{bmatrix}. \tag{65}$$

Taking derivatives from both sides of the equation gives

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = \begin{bmatrix} \frac{\mathrm{d}m(t)}{\mathrm{d}t} & \cdots & \frac{\mathrm{d}m(t)}{\mathrm{d}t} \end{bmatrix} + \sqrt{c}\begin{bmatrix} 0 & \frac{\mathrm{d}A(t)}{\mathrm{d}t} & -\frac{\mathrm{d}A(t)}{\mathrm{d}t} \end{bmatrix}. \tag{66}$$

Substituting the equations for $\mathrm{d}m(t)/\mathrm{d}t$ and $\mathrm{d}A(t)/\mathrm{d}t$ leads to equations in Algorithm 4.3.

REFERENCES

[1] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45(3), pp. 477–482, March 2000.

[2] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92(3), pp. 401–422, March 2004.

[3] E. A. Wan and R. van der Merwe, "The unscented Kalman filter," in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. Wiley, 2001, ch. 7.

[4] R. van der Merwe, "Sigma-point Kalman filters for probabilistic inference in dynamic state-space models," Ph.D. dissertation, OGI School of Science & Engineering, Oregon Health & Science University, Portland, OR, USA, April 2004.

[5] E. D. Kaplan, *Understanding GPS, Principles and Applications*. Artech House, Boston, London, 1996.

[6] M. S. Grewal, L. R. Weill, and A. P. Andrews, *Global Positioning Systems, Inertial Navigation and Integration*. Wiley Interscience, 2001.

[7] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. Wiley Interscience, 2001.

[8] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House Radar Library, 1999.

[9] L. D. Stone, C. A. Barlow, and T. L. Corwin, *Bayesian Multiple Target Tracking*. Artech House, Boston, London, 1999.

[10] J. D. Murray, *Mathematical Biology*. Springer, 1993, vol. 19.

[11] H. L. Van Trees, *Detection, Estimation, and Modulation Theory Part I*. John Wiley & Sons, New York, 1968.

[12] ——, *Detection, Estimation, and Modulation Theory Part II*. John Wiley & Sons, New York, 1971.

[13] P. Maybeck, *Stochastic Models, Estimation and Control, Volume 3*. Academic Press, 1982.

[14] R. F. Stengel, *Optimal Control and Estimation*. Dover Publications, Inc., 1994.

[15] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, ser. Applied mathematical Sciences. Springer, 2005, no. 160.

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

[17] A. Gelb, Ed., *Applied Optimal Estimation*. The MIT Press, 1974.

[18] M. S. Grewal and A. P. Andrews, *Kalman Filtering, Theory and Practice Using MATLAB*. Wiley Interscience, 2001.

[19] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*. Springer, 1991.

[20] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*. Springer, 2003.

[21] H. J. Kushner, "On the differential equations satisfied by conditional probability densities of Markov processes, with applications," *J. SIAM Control Ser. A*, vol. 2(1), 1964.

[22] R. L. Stratonovich, *Conditional Markov Processes and Their Application to the Theory of Optimal Control*. American Elsevier Publishing Company, Inc., 1968.

[23] R. S. Bucy, "Nonlinear filtering theory," *IEEE Transactions on Automatic Control*, vol. 10, pp. 198–198, 1965.
[24] M. Zakai, "On the optimal filtering of diffusion processes," *Zeit. Wahrsch.*, vol. 11, pp. 230–243, 1969.
[25] G. Kallianpur, *Stochastic Filtering Theory*. Springer-Verlag, 1980.
[26] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Transactions of the ASME, Journal of Basic Engineering*, vol. 83, pp. 95–108, March 1961.
[27] V. E. Beneš, "Exact finite-dimensional filters for certain diffusions with nonlinear drift," *Stochastics*, vol. 5, pp. 65–92, 1981.
[28] D. Crisan and T. Lyons, "Nonlinear filtering and measure-valued processes," *Probab. Theory Relat. Fields*, vol. 109, pp. 217–244, 1997.
[29] ——, "A particle approximation of the solution of the Kushner-Stratonovitch equation," *Probab. Theory Relat. Fields*, vol. 115, pp. 549–578, 1999.
[30] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME, Journal of Basic Engineering*, vol. 82, pp. 35–45, March 1960.
[31] F. E. Daum, "Exact finite-dimensional nonlinear filters for continuous time processes with discrete time measurements," in *Proceedings of 23rd Conference on Decision and Control*, 1984, pp. 16–22.
[32] ——, "Exact finite-dimensional nonlinear filters," *IEEE Transactions on Automatic Control*, vol. AC-31(7), pp. 616–622, 1986.
[33] P. E. Kloeden and E. Platen, *Numerical Solution to Stochastic Differential Equations*. Springer, 1999.
[34] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Springer, 2001.
[35] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter*. Artech House, 2004.
[36] P. D. Moral and L. Miclo, "Branching and interacting particle systems. Approximations of Feynman-Kac formulae with applications to nonlinear filtering," in *Séminaire de Probabilités XXXIV*. Springer-Verlag, 2000.
[37] M. Šimandl, J. Královec, and T. Söderström, "Anticipative grid design in point-mass approach to nonlinear state estimation," *IEEE Transactions on Automatic Control*, vol. 47(4), pp. 699–702, April 2002.
[38] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. AC-17(4), August 1972.
[39] T. Söderström, *Discrete-time stochastic systems : estimation and control*. New York, Prentice Hall, 1994.
[40] J. Gunther, R. Beard, J. Wilson, T. Oliphant, and W. Stirling, "Fast nonlinear filtering via Galerkin's method," in *American Control Conference. American Automatic Control Counsil.*, 1997.
[41] S. Challa, Y. Bar-Shalom, and V. Krishnamurthy, "Nonlinear filtering via generalized Edgeworth series and Gauss-Hermite quadrature," *IEEE Transactions on Signal Processing*, vol. 48(6), June 2000.
[42] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 1996.
[43] R. van der Merwe and E. Wan, "The square-root unscented Kalman filter for state and parameter-estimation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP*, vol. 6, 2001, pp. 3461–3464.
[44] S. J. Julier and J. K. Uhlmann, "Corrections to unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92(12), pp. 1958–1958, December 2004.
[45] M. Morf, B. Lévy, and T. Kailath, "Square-root algorithms for the continuous-time linear least-square estimation problem," *IEEE Transactions on Automatic Control*, vol. 23(5), pp. 907–911, Oct 1978.