Continuous-Time and Continuous-Discrete-Time Unscented Rauch-Tung-Striebel Smoothers

Simo Särkkä [a,1]

[a] *Helsinki University of Technology, Laboratory of Computational Engineering, P.O. Box 9203. FIN-02015 HUT, Finland*

Abstract

This article considers the application of the unscented transformation to approximate fixed-interval optimal smoothing of continuous-time non-linear stochastic systems. The proposed methodology can be applied to systems, where the dynamics can be modeled with non-linear stochastic differential equations and the noise corrupted measurements are obtained continuously or at discrete times. The smoothing algorithm is based on computing the continuous-time limit of the recently proposed unscented Rauch-Tung-Striebel smoother, which is an approximate optimal smoothing algorithm for discrete-time stochastic systems.

Key words: Unscented transformation, Unscented Kalman smoother, Unscented Rauch-Tung-Striebel smoother, Continuous-time smoother

1 Introduction

In this article, we present a new approximation algorithm to continuous-time fixed-interval optimal smoothing of non-linear stochastic state space models. Optimal smoothing in the context of state space models refers to statistical methodology that can be used for computing estimates of the past state history of a time varying system based on the history of noisy measurements obtained from it. By continuous-time we mean that the dynamics of the system are modeled with continuous-time state space models, that is, with stochastic differential equations. The derived continuous-time smoothers can be equally

Email address: simo.sarkka@hut.fi (Simo Särkkä).

[1] *Phone number:* +358 40 757 0730

applied to models, where the measurements are obtained continuously or to models where the measurements are obtained at discrete time instants.

1.1 Optimal Smoothing in Continuous-Time

Optimal smoothing problems are classically divided into fixed-point, fixed-lag and fixed-interval smoothing problems:

- *Fixed-point smoothing* refers to methodology, which can be used for efficiently computing the optimal estimate of the *initial state* or some other fixed-time state of a state space model, given an increasing number of measurements after it. This kind of estimation problem arises, for example, in estimation of the state of a space craft at a given point of time in the past [1] or in alignment and calibration of inertial navigation systems [2].
- *Fixed-lag smoothing* is methodology for computing delayed estimates of state space models given measurements up to the current time plus up to a constant horizon in the future. Fixed-lag smoothing can be considered as optimal filtering, where a constant delay is tolerated in the estimates. Potential applications of fixed-lag smoothing are all the problems where optimal filters are typically applied, and where the small delay is tolerated. An example of such application is digital demodulation [3].
- *Fixed-interval smoothing* can be used for computing estimates of a fixed time interval of states given the measurements on the same interval. These kind of batch estimates of the whole history of a state trajectory are often useful, for example, in target tracking, navigation, audio signal processing, machine learning, time series analysis and time series prediction [4–8].

In this article, we shall specifically concentrate on fixed-interval smoothing problems associated to continuous-time models. The corresponding fixed-point and fixed-lag algorithms could be derived from the fixed-interval smoothers in a quite straight-forward manner (cf., [1,9]), but they are not considered here.

Although, at the first glance *continuous-time* algorithms would not seem to be very useful in the current world of digital computers, in modeling point of view continuous-time systems are often more appropriate models of physical systems than discrete-time models. The Nature is – when the current physical theories are considered – still a continuous-time system. And when the models are in continuous time, the algorithms are also most naturally formulated in continuous time.

Another reason for studying continuous-time models is that digital computers are not the only possible devices for implementing estimation algorithms. Although, analog computers as such are not utilized in practical computations anymore, signal processing and control systems are still often implemented

using analog electrical components or are based on mechanical and optical phenomena. In such implementations, continuous-time algorithms are much more suitable than discrete-time algorithms.

1.2 Problem Formulation

In this article, we shall consider the application of the *unscented transformation* [10,11] to *optimal smoothing* of state space models with *continuous-time dynamics*. Discrete-time unscented transformation based (two-filter) smoothers have already been discussed in [12] and the discrete-time *unscented Rauch-Tung-Striebel smoother* (URTSS) has been presented in [13]. In this article we shall derive continuous-time versions of the URTSS using similar procedure as was used for deriving the *continuous-time and continuous-discrete unscented Kalman filters* in the article [14].

The considered models can be pure *continuous-time models*, where both the dynamics and measurements are modeled in continuous time such as

$$\begin{aligned} \frac{dx(t)}{dt} &= f(x(t), t) + L(t)\,\epsilon(t) \\ y(t) &= h(x(t), t) + \nu(t), \end{aligned} \tag{1}$$

or *continuous-discrete* state space models, where the dynamics are modeled in continuous time, but the measurements are obtained at discrete instants of time:

$$\frac{dx(t)}{dt} = f(x(t), t) + L(t)\,\epsilon(t) \tag{2}$$

$$y_k = h(x(t_k), k) + r_k. \tag{3}$$

In the models above $t \in [0, \infty)$ denotes the continuous time variable and $k \in \{0, 1, 2, \ldots\}$ is the discrete time step number occurring at the time instant t_k. The various terms are defined as follows:

- $x(t) = (x_1(t), \ldots, x_n(t)) \in \mathbb{R}^n$ is the continuous-time state process.
- $f(\cdot)$ is the non-linear drift function.
- $\epsilon(t)$ is the process noise, which is a zero mean Gaussian white noise process with spectral density matrix $Q_c(t) \in \mathbb{R}^{s \times s}$. The noise process is assumed to be independent of states and measurement noise.
- $L(t) \in \mathbb{R}^{n \times s}$ is the noise-feedback matrix, independent of states, measurements and noises.
- $\nu(t)$ is the continuous-time measurement noise, which is a zero mean Gaussian white noise process with spectral density matrix $R_c(t) \in \mathbb{R}^{d \times d}$. The noise process is assumed to be independent of states and process noise.
- $y(t) = (y_1(t), \ldots, y_d(t)) \in \mathbb{R}^d$ is the continuous-time measurement process.

- $y_k = (y_{k,1}, \ldots, y_{dk,}) \in \mathbb{R}^d$ is the discrete-time measurement obtained at the time instant t_k.
- $h(\cdot)$ is the non-linear measurement model function,
- r_k is the discrete-time measurement noise at the time instant t_k, and it is a zero mean Gaussian white noise sequence with covariance matrix $R_k \in \mathbb{R}^{m \times m}$.

In statistical estimation sense then purpose of the fixed-interval optimal smoother is to efficiently compute the marginal posterior distributions of the states at all times on the considered interval. That is, if the time interval is $[0, T]$ and the measurements are modeled in continuous time, we want to compute the distributions

$$p(x(t) \,|\, \{y(\tau) \ : \ 0 \leq \tau \leq T\}),$$

for each $t \in [0, T]$. If the measurements are modeled in discrete time, the corresponding posterior distribution is

$$p(x(t) \,|\, y_1, \ldots, y_T),$$

which we want to compute for each $t \in [0, t_T]$. In practical computations we often compute the point estimate of the state as the posterior mean of the distribution, which is the minimum mean squared estimate (MMSE) of the state. Furthermore, the uncertainty of the mean can be estimated by computing the posterior covariance of the state.

The continuous-time filtering model (1) and the dynamic model (2) fall into the class of models called *stochastic differential equations*. The theory of stochastic differential equations is well known, and it is commonly formulated in terms of *Itô calculus*, which is the theory of differential calculus of stochastic processes (see, e.g., [15,16]). In rigorous mathematical sense, for example, the dynamic model (2) should be actually interpreted as a *stochastic integral equation* of the form

$$x(t) - x(s) = \int_s^t f(x(t), t)\, dt + \int_s^t L(t)\, d\beta(t), \tag{4}$$

which can be written more compactly as

$$dx(t) = f(x(t), t)\, dt + L(t)\, d\beta(t), \tag{5}$$

where $\beta(t)$ is a Brownian motion with diffusion matrix $Q_c(t)$. If we define the *white noise* $\epsilon(t)$ as the formal derivative of the Brownian motion $\epsilon(t) = d\beta(t)/dt$, the equation (5) can be formally written in form (2).

In this article, we shall interpret the stochastic differential equations in the sense of Stratonovich [17], because in that interpretation the mathematical treatment of white noise can be done as if it was a non-random function. This does not affect the interpretation of the models (1) and (2) as such, because noises in these models are purely additive. However, the derived continuous-

time mean, covariance and sigma point differential equations, when they contain stochastic terms, should be interpreted in Stratonovich sense.

The difference between Itô and Stratonovich equations actually arises from how the stochastic integral of the form

$$\int_s^t L(x(t),t)\, d\beta(t), \tag{6}$$

where $x(t)$ is a β-adapted stochastic process, is defined. In the Itô interpretation the definition is chosen such that the integral becomes a martingale. The disadvantage of this is that the conventional chain rule of calculus no longer applies and the Itô formula has to be used instead. In Stratonovich interpretation the integral is defined such that the normal chain rule as well as other normal calculus rules apply, but the integral ceases to be a martingale. These properties make formal manipulation of Stratonovich stochastic differential equations easier than of Itô stochastic differential equations, but direct theoretical analysis of Stratonovich equations is more complicated than that of Itô equations. For discussion about the applicability and physical significance of the Stratonovich integral the reader is referred to [16,17]. Stratonovich stochastic differential equations can always be converted into equivalent Itô equations by using simple transformation formulas [16,17].

Because in the stochastic differential equation (4) the noise-feedback matrix does not depend on the state, that is, we have $L(x(t),t) = L(t)$, the definitions coincide. In this sense it does not matter whether the Itô or Stratonovich interpretation of the model equation is used. However, in the derivation of the continuous-time smoother equations the formulas do contain more general stochastic integral terms, where the different interpretations potentially do matter, and because the derivation is based on normal rules of calculus, the resulting equations should be interpreted in Stratonovich sense.

1.3 Approaches to Approximate Optimal Smoothing

Kalman filtering [18,19] and more general *optimal filtering* [20–24] are closely related to optimal smoothing, and for this reason most of the existing approximate non-linear optimal smoothing methods are based on similar approximations as is used in optimal filters. These kind of approximations are, for example, Taylor series approximations of extended Kalman filters [22], sigma-point approximations used in unscented Kalman filters [11,12,25,26] and Monte Carlo approximations used in particle filters [27,28].

The solution to the fixed-interval *linear Gaussian* smoothing problem in both continuous and discrete time settings is given by the Rauch-Tung-Striebel

smoother [29,30]. Alternatively the solution can be represented as a combination of two Kalman filters, one running forwards and another backwards in time [31]. All the solutions to *linear Gaussian* smoothing problems are well known and well documented in many text books [1,9,24,32,33]. Taylor series based approximations the more general formalism is also well documented in the literature [9,23,32,34–36].

Unscented transformation based two-filter type optimal smoothing in discrete-time model context has already been discussed in [12] and the discrete-time unscented Rauch-Tung-Striebel smoother was presented in the recent article [13]. Discrete-time Monte Carlo based particle smoothing methods are described in the articles [5,7].

Continuous-time optimal smoothing has recently received less attention in applied literature, and the continuous-time smoothing methods can be more easily found in older books [1,23,32]. However, in the recent book [9] also the continuous-time linear and non-linear smoothing is considered. In this article we shall present a new continuous-time smoothing algorithm, which is based on the unscented transformation, or more specifically, is the continuous-time limit of the unscented Rauch-Tung-Striebel smoother [13].

1.4 Contributions and Organization of the Paper

The main contribution of the paper is to derive equations of the *continuous-time* version of the *unscented Rauch-Tung-Striebel smoother* [13]. The derivation is based on the similar limiting procedure as was used for deriving the continuous-time unscented Kalman filter from the discrete-time filter in [14]. Some of the results have already appeared in the doctoral thesis [37], but here the results are significantly extended and more throughly analyzed.

The organization of the paper is the following:

- The unscented transformation, discrete-time unscented Kalman filter and continuous-time unscented Kalman filter are reviewed in Sections 2.1, 2.2, and 2.3, respectively.
- The additive form of the discrete-time unscented Rauch-Tung Striebel is presented in Section 3.1. The derivation of the basic continuous-time unscented Rauch-Tung-Striebel smoother is presented in Section 3.2 and the sigma-point form is derived in Section 3.3
- The computational complexity of the algorithms is analyzed in Section 4.
- In Section 5 a numerical example is presented, where the proposed algorithm is applied to a simulated non-linear filtering and smoothing problem.

2 Unscented Kalman Filter

The *unscented transformation* (UT) [11] is a relatively recent method for approximating non-linear transformations of random variables. Instead of Taylor series approximations, it is based on forming a set of sigma points, which are propagated through the non-linearity. The *unscented Kalman filter* (UKF) [11,12,14,26] is an alternative to the extended Kalman filter (EKF) [22], which utilizes the unscented transformation in the filter computations. In this section, we shall shortly review the unscented transformation, unscented Kalman and the continuous-time unscented Kalman filter, that is, the unscented Kalman-Bucy filter.

2.1 Unscented Transformation

The *unscented transformation* (UT) [10,11] is a numerical method that 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 as 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 $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $y = (y_1, \ldots, y_d) \in \mathbb{R}^d$, and $g(\cdot)$ is a general non-linear function from $\mathbb{R}^n$ to $\mathbb{R}^d$. The *unscented transformation* forms the Gaussian approximation

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} m \\ \mu \end{bmatrix}, \begin{bmatrix} P & C \\ C^T & S \end{bmatrix} \right),$$

to the joint distribution of $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ by computing the parameters μ, C and S with the following procedure:

(1) Form the matrix of sigma points X as

$$X = \begin{bmatrix} m & \cdots & m \end{bmatrix} + \sqrt{c} \begin{bmatrix} 0 & \sqrt{P} & -\sqrt{P} \end{bmatrix},$$

where $c = n + \lambda$ and λ is a scaling parameter, which is defined in terms of algorithm parameters α and κ as follows:

$$\lambda = \alpha^2 (n + \kappa) - n. \tag{8}$$

The parameters α and κ determine the spread of the sigma points around

the mean [12]. The matrix square root denotes a matrix such that $\sqrt{P}\,\sqrt{P}^T = P$. The sigma points are the columns of the sigma point matrix.

(2) Propagate the sigma points through the non-linear function $g(\cdot)$:

$$Y_i = g(X_i), \quad i = 1 \dots 2n+1,$$

where X_i and Y_i denote the ith columns of matrices X and Y, respectively.

(3) Estimates of the mean μ and covariance S of the transformed variable, as well as estimate of the cross-covariance C of the non-transformed and transformed variables can be computed from the sigma points as follows:

$$\begin{aligned} \mu &= \sum_i W_{i-1}^{(m)}\, Y_i \\ S &= \sum_i W_{i-1}^{(c)}\, (Y_i - \mu)\, (Y_i - \mu)^T \\ C &= \sum_i W_{i-1}^{(c)}\, (X_i - m)\, (Y_i - \mu)^T, \end{aligned} \tag{9}$$

where the constant weights $W_i^{(m)}$ and $W_i^{(c)}$ are defined as follows:

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

Here β is an additional algorithm parameter, which can be used for incorporating prior information on the form of the distribution of x [12].

The mean, covariance and cross-covariance equations (9) of the unscented transformation can also be written in convenient *matrix form* as follows [14]:

$$\begin{aligned} \mu &= Y\, w_m \\ S &= Y\, W\, Y^T \\ C &= X\, W\, Y^T, \end{aligned} \tag{11}$$

where the vector w_m and matrix W can be easily computed from the weights $W_i^{(m)}$ and $W_i^{(c)}$ as follows:

$$w_m = \begin{bmatrix} W_0^{(m)} & \cdots & W_{2n}^{(m)} \end{bmatrix}^T$$

$$W = \left(I - \begin{bmatrix} w_m & \cdots & w_m \end{bmatrix} \right) \begin{bmatrix} W_0^{(c)} & & \\ & \ddots & \\ & & W_{2n}^{(c)} \end{bmatrix} \left(I - \begin{bmatrix} w_m & \cdots & w_m \end{bmatrix} \right)^T.$$

The matrix form (11) is particularly convenient in mathematical treatment of the unscented transformation and for this reason it is used in the derivation of the continuous-time smoother in this article. However, from the computational point of view it is more efficient to use the summation form (9) of the transformation [14].

2.2 Discrete-Time Unscented Kalman Filter

In Bayesian sense, optimal non-linear discrete-time filtering considers estimation of the state in generic state space models of the form

$$\begin{aligned} x_k &\sim p(x_k \mid x_{k-1}) \\ y_k &\sim p(y_k \mid x_k), \end{aligned} \tag{12}$$

where x_k is the unknown hidden state and y_k is the measurement on time step k. The dynamic model $p(x_k \mid x_{k-1})$ defines a Markov model for state transitions. The measurement model $p(y_k \mid x_k)$ defines the distribution of measurements for given state configurations. On the initial time step $k = 0$, the state is assumed to have a predefined prior distribution $p(x_0)$.

The purpose of the optimal filter is to compute the filtering distribution, which is the posterior distribution of the current state x_k given the history of measurements up to the current time step k:

$$p(x_k \mid y_1, \ldots, y_k) = p(x_k \mid y_{1:k}). \tag{13}$$

The recursive equations for calculating the posterior state distribution above are called the *optimal filtering equations* or the *Bayesian filtering equations*:

$$p(x_k \mid y_{1:k-1}) = \int p(x_k \mid x_{k-1})\; p(x_{k-1} \mid y_{1:k-1})\; dx_{k-1} \tag{14}$$

$$p(x_k \mid y_{1:k}) = \frac{p(y_k \mid x_k)\; p(x_k \mid y_{1:k-1})}{\int p(y_k \mid x_k)\; p(x_k \mid y_{1:k-1})\; dx_k}, \tag{15}$$

where Equation (14) is called the Chapman-Kolmogorov equation (or the prediction step) and Equation (15) is the Bayes' rule (or the update step).

The discrete-time *unscented Kalman filter* (UKF) [11,12,25,26] is an optimal filtering algorithm that utilizes the unscented transformation and can be used for approximating the filtering distributions of models, which have the following state space form:

$$\begin{aligned} x_k &= f(x_{k-1}, k-1) + q_{k-1} \\ y_k &= h(x_k, k) + r_k, \end{aligned} \tag{16}$$

where $x_k \in \mathbb{R}^n$ is the state, $y_k \in \mathbb{R}^m$ is the measurement, $q_{k-1} \sim \mathrm{N}(0, Q_{k-1})$ is the Gaussian process noise, $r_k \sim \mathrm{N}(0, R_k)$ is the Gaussian measurement noise, $f(\cdot)$ is the dynamic model function and $h(\cdot)$ is the measurement model function. On the initial time step the state is assumed to have a Gaussian prior distribution $x_0 \sim \mathrm{N}(m_0, P_0)$ with a known mean m_0 and covariance P_0.

Clearly the model (16) can be seen to be a special case of the model (12). Note that UKF can be applied to slightly more general models than (16), in particular, the noises do not necessarily need to appear in additive manner as they do here. However, the continuous-time UKF is based on this additive form of the filter and for this reason it is used here.

The UKF forms a Gaussian approximation to the filtering distribution as follows:

$$p(x_k \mid y_1, \ldots, y_k) \approx \mathrm{N}(x_k \mid m_k, P_k), \tag{17}$$

where m_k and P_k are the mean and covariance computed by the algorithm.

In the *unscented Kalman filter* (UKF) algorithm, computations are started from the prior mean m_0 and covariance P_0, and the following operations are performed on each measurement step $k = 1, 2, 3, \ldots$:

(1) *Prediction step:*

(a) Form the matrix of sigma points:

$$X_{k-1} = \begin{bmatrix} m_{k-1} & \cdots & m_{k-1} \end{bmatrix} + \sqrt{c} \begin{bmatrix} 0 & \sqrt{P_{k-1}} & -\sqrt{P_{k-1}} \end{bmatrix}. \tag{18}$$

(b) Propagate the sigma points through the dynamic model:

$$\hat{X}_{k,i} = f(X_{k-1,i}, k-1), \quad i = 1 \ldots 2n+1. \tag{19}$$

(c) Compute the predicted mean m_k^- and the predicted covariance P_k^-:

$$\begin{aligned} m_k^- &= \sum_i W_{i-1}^{(m)} \, \hat{X}_{k,i} \\ P_k^- &= \sum_i W_{i-1}^{(c)} \, (\hat{X}_{k,i} - m_k^-) \, (\hat{X}_{k,i} - m_k^-)^T + Q_{k-1}. \end{aligned} \tag{20}$$

(2) *Update step:*

(a) Form the matrix of sigma points:

$$X_k^- = \begin{bmatrix} m_k^- & \cdots & m_k^- \end{bmatrix} + \sqrt{c} \begin{bmatrix} 0 & \sqrt{P_k^-} & -\sqrt{P_k^-} \end{bmatrix}. \tag{21}$$

(b) Propagate sigma points through the measurement model:

$$\hat{Y}_{k,i} = h(X_{k,i}^-, k), \quad i = 1 \ldots 2n+1. \tag{22}$$

(c) Compute the predicted mean μ_k, the predicted covariance of the measurement S_k, and the cross-covariance of the state and measurement B_k:

$$\begin{aligned} \mu_k &= \sum_i W_{i-1}^{(m)}\, \hat{Y}_{k,i} \\ S_k &= \sum_i W_{i-1}^{(c)}\, (\hat{Y}_{k,i} - \mu_k)\, (\hat{Y}_{k,i} - \mu_k)^T + R_k \\ B_k &= \sum_i W_{i-1}^{(c)}\, (X_{k,i}^- - m_k^-)\, (\hat{Y}_{k,i} - \mu_k)^T. \end{aligned} \tag{23}$$

(d) Compute the filter gain K_k and the filtered state mean m_k and covariance P_k, conditional to the measurement y_k:

$$\begin{aligned} K_k &= B_k\, S_k^{-1} \\ m_k &= m_k^- + K_k\, [y_k - \mu_k] \\ P_k &= P_k^- - K_k\, S_k\, K_k^T. \end{aligned} \tag{24}$$

In terms of the matrix form of unscented transformation, the steps 1b and 1c of the filter, that is, the Equations (19) and (20) can be written in the following compact form [14]:

$$\begin{aligned} \hat{X}_k &= f(X_{k-1}, k-1) \\ m_k^- &= \hat{X}_k\, w_m \\ P_k^- &= \hat{X}_k\, W\, \hat{X}_k^T + Q_{k-1}. \end{aligned} \tag{25}$$

Similarly, the steps 2b and 2c or the Equations (22) and (23) can be written as

$$\begin{aligned} \hat{Y}_k &= h(X_k^-, k) \\ \mu_k &= \hat{Y}_k\, w_m \\ S_k &= \hat{Y}_k\, W\, \hat{Y}_k^T + R_k \\ B_k &= X_k^-\, W\, \hat{Y}_k^T. \end{aligned} \tag{26}$$

In the above equations we have used the shorthand notation introduced in [14] such that the expression

$$Y = g(X), \tag{27}$$

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

$$Y_i = g(X_i).$$

In this section we shall present the *unscented Kalman-Bucy filter* and *continuous-discrete unscented Kalman filter*, which can be derived from the discrete-time unscented Kalman filter by a formal limiting procedure. The derivations of these filters can be found in the article [14].

Theorem 1 (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*

$$\begin{aligned} K(t) &= X(t)\, W\, h^T(X(t),t)\, R_c^{-1}(t) \\ \frac{dm(t)}{dt} &= f(X(t),t)\, w_m + K(t)\, [y(t) - h(X(t),t)\, w_m] \\ \frac{dP(t)}{dt} &= X(t)\, W\, f^T(X(t),t) + f(X(t),t)\, W\, X^T(t) \\ &\quad + L(t)\, Q_c(t)\, L^T(t) - K(t)\, R_c(t)\, K^T(t). \end{aligned} \tag{28}$$

The sigma-point matrix $X(t)$ is defined as

$$X(t) = \begin{bmatrix} m(t) & \cdots & m(t) \end{bmatrix} + \sqrt{c}\, \begin{bmatrix} 0 & \sqrt{P(t)} & -\sqrt{P(t)} \end{bmatrix}. \tag{29}$$

The computations are started from a known initial mean $m(0) = m_0$ and covariance $P(0) = P_0$.

The equations above are conditioned on the measurement signal $y(t)$, but in construction of the continuous-discrete time filter we need the optimal prediction equations which give the optimal state estimate when no measurement signal is available. These prediction equations are valid in both continuous-time and continuous-discrete-time cases.

Theorem 2 (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 instant t_0, that is, $m(t_0)$ and $P(t_0)$ can be computed by integrating the differential equations*

$$\begin{aligned} \frac{dm(t)}{dt} &= f(X(t),t)\, w_m \\ \frac{dP(t)}{dt} &= X(t)\, W\, f^T(X(t),t) + f(X(t),t)\, W\, X^T(t) + L(t)\, Q_c(t)\, L^T(t), \end{aligned} \tag{30}$$

from the initial conditions $m(t_0)$ and $P(t_0)$ to the time instant t. Here $X(t)$ is defined as in Equation (29)*.*

The *continuous-discrete UKF* can be now implemented as follows:

(1) *Prediction.* Integrate the differential equations (30) from the initial conditions $m(t_{k-1}) = m_{k-1}$, $P(t_{k-1}) = P_{k-1}$ to the time instant t_k. The predicted mean and covariance are given as $m_k^- = m(t_k)$ and $P_k^- = P(t_k)$, respectively.
(2) *Update.* The update step is the same as the discrete-time unscented Kalman filter update step (21) - (24) in Section 2.2.

The continuous-time *unscented Kalman-Bucy filter* and the *continuous-discrete unscented Kalman filter* [14] compute the means $m(t)$ and covariances $P(t)$ of Gaussian approximations to the filtering distributions of the continuous-time (1) and continuous-discrete models (2) - (3), respectively. The resulting posterior (filtering) distribution approximations can be then interpreted to be

$$p(x(t) \,|\, \{y(\tau) : 0 \leq \tau \leq t\}) \approx \mathrm{N}(x(t) \,|\, m(t), P(t)).$$
$$p(x(t_k) \,|\, y_1, \ldots, y_k) \approx \mathrm{N}(x(t_k) \,|\, m(t_k), P(t_k)).$$

The continuous-discrete filtering distribution can be generalized to all $t \in [0, \infty)$ by defining that between the measurements, the filtering distribution is the optimal prediction from previous measurement to that time instant. The continuous-discrete Kalman filter can be then interpreted to form the following approximation to the generalized filtering distribution:

$$p(x(t) \,|\, y_1, \ldots, y_k) \approx \mathrm{N}(x(t) \,|\, m(t), P(t)), \tag{31}$$

where k and t are related by $t \in [t_k, t_{k+1})$, where t_k and t_{k+1} are the times of the measurements y_k and y_{k+1}, respectively.

In the article [14], the sigma point versions of the filter equations are also presented, but they are not repeated in this article, because here we only need the mean and covariance form of the filter.

3 Unscented Rauch-Tung-Striebel Smoother

In this section, we shall first review the basic Bayesian theory behind Rauch-Tung-Striebel smoothing and then review the discrete-time unscented Rauch-Tung-Striebel smoother algorithm presented in [13]. After that the mean/covariance and sigma-point forms of the continuous-time unscented Rauch-Tung-Striebel smoothers are derived.

3.1 Discrete-Time Unscented RTS Smoother

In Bayesian estimation perspective, the purpose of (fixed-interval) *optimal smoothing* is to compute the posterior distribution of the state x_k at a time step k after receiving the measurements up to a time step T, where $T > k$

$$p(x_k \mid y_1, \ldots, y_T) = p(x_k \mid y_{1:T}).$$

The difference between filters and smoothers is that *a filter* computes its estimates using only the measurements obtained before and at the time step k, but *a smoother* uses also the future measurements for computing its estimates. After obtaining the filtered posterior state distributions, the following *Bayesian optimal smoothing equations* can be used for calculating the posterior distribution for each time step conditionally to all measurements up to the time step T:

$$p(x_{k+1} \mid y_{1:k}) = \int p(x_{k+1} \mid x_k)\; p(x_k \mid y_{1:k})\; dx_k \tag{32}$$

$$p(x_k \mid y_{1:T}) = \int \frac{p(x_{k+1} \mid x_k)\; p(x_k \mid y_{1:k})}{p(x_{k+1} \mid y_{1:k})} p(x_{k+1} \mid y_{1:T})\; dx_{k+1}. \tag{33}$$

As the noises in the state space model (16) appear in additive manner, it is possible to write the general URTSS equations [13] in a bit simpler additive form in an analogous manner as in the filter case [12]. The resulting *additive form unscented RTS smoother algorithm* is the following:

(1) Form the matrix of sigma points:

$$X_k = \begin{bmatrix} m_k \cdots m_k \end{bmatrix} + \sqrt{c} \begin{bmatrix} 0 & \sqrt{P_k} & -\sqrt{P_k} \end{bmatrix}.$$

(2) Propagate the sigma points through the dynamic model:

$$\hat{X}_{k+1,i} = f(X_{k,i}, k), \quad i = 1 \ldots 2n+1.$$

(3) Compute the predicted mean m^-_{k+1}, predicted covariance P^-_{k+1} and cross-covariance C_{k+1}:

$$\begin{aligned} m^-_{k+1} &= \sum_i W^{(m)}_{i-1}\, \hat{X}_{k+1,i} \\ P^-_{k+1} &= \sum_i W^{(c)}_{i-1}\, (\hat{X}_{k+1,i} - m^-_{k+1})\, (\hat{X}_{k+1,i} - m^-_{k+1})^T + Q_k \\ C_{k+1} &= \sum_i W^{(c)}_{i-1}\, (X_{k,i} - m_k)\, (\hat{X}_{k+1,i} - m^-_{k+1})^T. \end{aligned} \tag{34}$$

(4) Compute the smoother gain D_k, smoothed mean m^s_k and smoothed co-

variance P_k^s as follows:

$$\begin{aligned} D_k &= C_{k+1}\,[P_{k+1}^-]^{-1} \\ m_k^s &= m_k + D_k\,\left[m_{k+1}^s - m_{k+1}^-\right] \\ P_k^s &= P_k + D_k\,\left[P_{k+1}^s - P_{k+1}^-\right]\,D_k^T. \end{aligned} \tag{35}$$

The computations above are started from the filtering result of the last time step $m_T^s = m_T$, $P_T^s = P_T$ and the recursion runs backwards for $k = T-1, \ldots, 0$. Using the matrix form of UT in Equation (11) and with the shorthand notation in Equation (27) the steps 2 and 3 can be written as:

$$\begin{aligned} \hat{X}_{k+1} &= f(X_k, k) \\ m_{k+1}^- &= \hat{X}_{k+1}\,w_m \\ P_{k+1}^- &= \hat{X}_{k+1}\,W\,\hat{X}_{k+1}^T + Q_k \\ C_{k+1} &= X_k\,W\,\hat{X}_{k+1}^T. \end{aligned} \tag{36}$$

The optimal smoothing algorithm discussed above is a *discrete-time* algorithm in the sense that the state evolves in discrete time steps and measurements are obtained on these steps. However, as shown in next section, by using a formal limiting procedure it is possible derive the corresponding smoothing equations for continuous-time and continuous-discrete state space models.

3.2 Continuous-Time Unscented Rauch-Tung-Striebel Smoother

In this section the continuous-time unscented Rauch-Tung-Striebel smoother is derived by taking the formal continuous-time limit of the discrete-time unscented Rauch-Tung-Striebel smoother equations presented in Section 3.1.

In the continuous-time fixed-interval smoothing we are searching for approximations to the following *smoothing distributions*, which can be computed for all times $t \in [0, T]$ if the measurement process up to the time instant T has been observed:

$$p(x(t) \,|\, \{y(\tau) \;:\; 0 \le \tau \le T\}), \qquad 0 \le t \le T. \tag{37}$$

In this article, the approximation is chosen to be Gaussian with mean $m^s(t)$ and covariance $P^s(t)$:

$$p(x(t) \,|\, \{y(\tau) \;:\; 0 \le \tau \le T\}) \approx \mathrm{N}(x(t) \,|\, m^s(t), P^s(t)). \tag{38}$$

In the continuous-discrete-time case, we are looking for approximations to the smoothing distributions, which are conditioned to all the discrete-time

measurements obtained at a given time interval $[0, t_T]$, and these distributions are again approximated with Gaussian distributions:

$$p(x(t) \,|\, y_1, \ldots, y_T) \approx \mathrm{N}(x(t) \,|\, m^s(t), P^s(t)), \qquad 0 \le t \le t_T. \tag{39}$$

Although the filters for the continuous-time and continuous-discrete cases are different, the Rauch-Tung-Striebel form optimal smoother is the same in both the cases. This is because we can define the continuous-discrete filtering distributions as in Equation (31) and thus the mean and covariance are available for all time instants continuously also in the continuous-discrete case. We may then apply the smoother as if the mean and covariance were obtained from a continuous-time filter.

Theorem 3 (continuous-time unscented RTS smoother (CURTSS)) *In the continuous-time limit the discrete-time* unscented Rauch-Tung-Striebel smoother *equations approach the following differential equations:*

$$\begin{aligned}
\frac{dm^s(t)}{dt} &= f(X,t)\, w_m + [f(X,t)\, W\, X^T(t) + L(t)\, Q_c(t)\, L^T(t)] P^{-1}(t)\, [m^s(t) - m(t)] \\
\frac{dP^s(t)}{dt} &= [f(X,t)\, W\, X^T(t) + L(t)\, Q_c(t)\, L^T(t)]\, P^{-1}(t)\, P^s(t) \\
&\quad + P^s(t)\, P^{-1}(t)\, [X(t)\, W\, f^T(X,t) + L(t)\, Q_c(t)\, L^T(t)] - L(t)\, Q_c(t)\, L^T(t),
\end{aligned} \tag{40}$$

where the vector w_m and matrix W are the weights of the matrix form of unscented transformation (11)*. The sigma points $X(t)$ are the sigma points of the continuous-time or continuous-discrete unscented Kalman filter, such that $P(t) = X(t)\, W\, X^T(t)$ and $m(t) = X(t)\, w_m$, where $P(t)$and $m(t)$ are the covariance and mean of the UKF. The integration is performed backwards starting from the terminal conditions $m^s(T) = m(T)$, $P^s(T) = P(T)$.*

PROOF. The stochastic differential equation [2]

$$\frac{dx}{dt} = f(x(t), t) + L(t)\, \epsilon(t),$$

can be approximated up to first order in δt by a discretized dynamic model

$$x(t + \delta t) = x(t) + f(x(t), t)\, \delta t + q(t) + o(\delta t),$$

where $q(t) \sim \mathrm{N}(0, L(t)\, Q_c(t)\, L^T(t)\, \delta t)$. Applying the matrix forms of steps 2 and 3 of URTSS, which are given in Equation (36), we get expressions for the

[2] Strictly speaking, we must consider the stochastic differential equations as Stratonovich equations [16], because the derivation is based on the rules of ordinary calculus.

approximate mean, covariance and cross-covariance as follows:

$$\begin{aligned}
m^-(t+\delta t) &= m(t) + f(X(t),t)\, w_m\, \delta t + o(\delta t) \\
P^-(t+\delta t) &= [X(t) + f(X(t),t)\, \delta t + o(\delta t)]\, W\, [X(t) + f(X(t),t)\, \delta t + o(\delta t)]^T \\
&\quad + L(t)\, Q_c(t)\, L^T(t)\, \delta t + o(\delta t) \\
&= P(t) + \partial P^-(t)\, \delta t + o(\delta t) \\
C^-(t+\delta t) &= X(t)\, W\, [X(t) + f(X(t),t)\, \delta t + o(\delta t)]^T \\
&= P(t) + \partial C^-(t)\, \delta t + o(\delta t),
\end{aligned}$$

where the expressions for the differentials of covariance and cross-covariance are

$$\begin{aligned}
\partial P^-(t) &= f(X(t),t)\, W\, X^T(t) + X(t)\, W\, f^T(X(t),t) + L(t)\, Q_c(t)\, L^T(t) \\
\partial C^-(t) &= X(t)\, W\, f^T(X(t),t).
\end{aligned} \tag{41}$$

The equation for the smoother gain (35) is given as

$$D(t+\delta t) = C^-(t+\delta t)\, \left[P^-(t+\delta t)\right]^{-1}.$$

By the standard differentiation rules for matrices, the differential of the gain can be written as

$$\begin{aligned}
\partial D(t) &= \partial C^-(t)\, P^{-1}(t) - P(t)\, P^{-1}\, \partial P^-(t)\, P^{-1}(t) \\
&= -f(X(t),t)\, W\, X^T(t)\, P^{-1}(t) - L(t)\, Q(t)\, L^T(t)\, P^{-1}(t).
\end{aligned}$$

and hence we have

$$D(t+\delta t) = I + \partial D(t)\, \delta t + o(\delta t).$$

If we substitute the gain expression into the mean expression in (35) and collect the terms up to order δt, we get

$$\begin{aligned}
m^s(t) &= m(t) + D(t+\delta t)\, [m^s(t+\delta t) - m^-(t+\delta t)] \\
&= m^s(t+\delta t) - f(X(t),t)\, w_m\, \delta t \\
&\quad + f(X(t),t)\, W\, X^T(t)\, P^{-1}(t)\, m(t)\, \delta t + L(t)\, Q_c(t)\, L^T(t)\, P^{-1}(t)\, m(t)\, \delta t \\
&\quad - f(X(t),t)\, W\, X^T(t) P^{-1}(t)\, m^s(t+\delta t)\, \delta t \\
&\quad - L(t)\, Q_c(t)\, L^T(t)\, P^{-1}(t)\, m^s(t+\delta t)\, \delta t + o(\delta t).
\end{aligned} \tag{42}$$

For the covariance we similarly get

$$\begin{aligned} P^s(t) &= P(t) + D(t+\delta t)\left(P^s(t+\delta t) - P^-(t+\delta t)\right) D^T(t+\delta t) \\ &= P^s(t+\delta t) - f(X(t),t)\, W\, X^T(t)\, P^{-1}(t)\, P^s(t+\delta t)\, \delta t \\ &\quad - L(t)\, Q_c(t)\, L^T(t)\, P^{-1}(t)\, P^s(t+\delta t)\, \delta t \\ &\quad - P^s(t+\delta t)\, P^{-1}(t)\, X(t)\, W\, f^T(X(t),t)\, \delta t \\ &\quad - P^s(t+\delta t)\, P^{-1}(t)\, L(t)\, Q_c(t)\, L^T(t)\, \delta t \\ &\quad + L(t)\, Q_c(t)\, L^T(t)\, \delta t + o(\delta t). \end{aligned} \tag{43}$$

If we now solve for $(m^s(t+\delta t) - m^s(t))/\delta t$ and $(P^s(t+\delta t) - P^s(t))/\delta t$ in (42) and (43), respectively, and take the limit $\delta t \to 0$, the result follows.

3.3 Sigma Point Form of Continuous-Time URTSS

In this section the differential equations for the smoothed sigma points are derived. The procedure is the same as was used in [38] for deriving the square root version of the extended Kalman-Bucy filter and in [14] for deriving the sigma point form of the unscented Kalman-Bucy filter.

Theorem 4 (Sigma point CURTSS) *The continuous-time unscented Rauch-Tung-Striebel smoother equations* (40) *can be expressed in terms of sigma points as follows:*

$$\begin{aligned} D(t) &= \Big[f(X,t)\, W\, X^T(t) + L(t)\, Q_c(t)\, L^T(t) \Big] P^{-1}(t) \\ M^s(t) &= [A^s(t)]^{-1}\, D(t)\, A^s(t) + [A^s(t)]^T\, D^T(t)\, [A^s(t)]^{-T} \\ &\quad - [A^s(t)]^{-1} L(t)\, Q_c(t)\, L^T(t)\, [A^s(t)]^{-T} \\ \frac{dX_i^s(t)}{dt} &= f(X(t),t)\, w_m + D(t)\left[X^s(t)\, w_m - m(t) \right] \\ &\quad + \sqrt{c}\, \Big[0 \;\; A^s(t)\, \Phi\Big(M^s(t) \Big) \;\; -A^s(t)\, \Phi^s\Big(M^s(t) \Big) \Big]_i , \end{aligned} \tag{44}$$

where the matrix of smoothed sigma-points is defined as

$$X^s(t) = \Big[m^s(t) \;\cdots\; m^s(t) \Big] + \sqrt{c}\, \Big[0 \;\; A^s(t) \;\; -A^s(t) \Big]$$

and $A^s(t)$ is the lower triangular Cholesky factor of $P^s(t)$. Note that in practical computations the terms $A^s(t)$ in Equations (44) *are obtained by "extracting" them from the sigma point matrix above (subtract the mean and divide by $\sqrt{c}$) and for this reason there is no explicit update rule for $A^s(t)$. In Equations* (44) *$\Phi(\cdot)$ is a function returning the lower diagonal part of the argument as*

follows:

$$\Phi_{ij}\Big(M(t)\Big) = \begin{cases} m_{ij}(t) \ , \ \textit{if } i > j \\ \frac{1}{2} m_{ij}(t) \ , \ \textit{if } i = j \\ 0 \ , \ \textit{if } i < j. \end{cases} \tag{45}$$

The backward integration should be started from the set of sigma points $X^s(T) = X(T)$ *formed from the terminal mean* $m(T)$ *and covariance* $P(T)$ *of the UKF.*

PROOF. The mean and covariance differential equations of the continuous-time unscented Kalman smoother are of the form

$$\begin{aligned} \frac{dm^s(t)}{dt} &= f(X,t)\, w_m + D(t)\,[m^s(t) - m(t)] \\ \frac{dP^s(t)}{dt} &= D(t)\, P^s(t) + P^s(t)\, D^T(t) - L(t)\, Q_c(t)\, L^T(t), \end{aligned} \tag{46}$$

where

$$D(t) = [f(X,t)\, W\, X^T(t) + L(t)\, Q_c(t)\, L^T(t)]\, P^{-1}(t).$$

Let $A^s(t)$ be the lower triangular Cholesky factor of $P^s(t)$, that is, $P^s(t) = A^s(t)\,[A^s(t)]^T$. The differential of the covariance can be expanded as

$$\frac{dP^s(t)}{dt} = \frac{dA^s(t)}{dt}\,[A^s(t)]^T + A^s(t)\,\left[\frac{dA^s(t)}{dt}\right]^T.$$

If we substitute this into the covariance equation in (46) and multiply from left and right by $[A^s(t)]^{-1}$ and $[A^s(t)]^{-T}$, respectively, we get

$$\begin{aligned} &[A^s(t)]^{-1}\,\frac{dA^s(t)}{dt} + \left[\frac{dA^s(t)}{dt}\right]^T\,[A^s(t)]^{-T} \\ &\quad = [A^s(t)]^{-1} D(t)\, A^s(t) + [A^s(t)]^T\, D^T(t)\,[A^s(t)]^{-T} \\ &\qquad - [A^s(t)]^{-1} L(t)\, Q_c(t)\, L^T(t)\,[A^s(t)]^{-T}. \end{aligned}$$

Because the left hand side is a sum of lower and upper triangular matrices, we must have

$$\begin{aligned} [A^s(t)]^{-1}\,\frac{dA^s(t)}{dt} = \Phi\Big(&[A^s(t)]^{-1} D(t)\, A^s(t) + [A^s(t)]^T\, D^T(t)\,[A^s(t)]^{-T} \\ &- [A^s(t)]^{-1} L(t)\, Q_c(t)\, L^T(t)\,[A^s(t)]^{-T}\Big), \end{aligned}$$

where $\Phi(\cdot)$ is the function defined in (45). If we multiply from left by $A^s(t)$ and substitute the equations for $dm^s(t)/dt$ and $dA^s(t)/dt$ into the equation

$$\frac{dX^s}{dt} = \Big[dm^s/dt \ \cdots \ dm^s/dt\Big] + \sqrt{c}\,\Big[0 \ \ dA^s/dt \ \ -dA^s/dt\Big],$$

the result follows.

4 Computational Complexity

In this section we shall compare the computational complexity of the proposed smoothing algorithms to the classical linearization based extended Rauch-Tung-Striebel smoother. The equations of the continuous-time extended Rauch-Tung-Striebel smoother can be written in the following form (see, e.g., [9,37]):

$$\begin{aligned}\frac{dm^s(t)}{dt} &= f(m(t),t) + \left[F(m(t),t)\,P(t) + L(t)\,Q_c(t)\,L^T(t)\right] P^{-1}(t)\,\left[m^s(t) - m(t)\right] \\ \frac{dP^s(t)}{dt} &= \left[F(m(t),t)\,P(t) + L(t)\,Q_c(t)\,L^T(t)\right] P^{-1}(t)\,P^s(t) \\ &\quad + P^s(t)\,P^{-1}(t)\,\left[P(t)\,F^T(m,t) + L(t)\,Q_c(t)\,L^T(t)\right] - L(t)\,Q_c(t)\,L^T(t),\end{aligned} \tag{47}$$

where $m(t)$ and $P(t)$ are the mean and covariance of the filtering solution computed by the extended Kalman filter. The integration is performed backwards starting from the terminal conditions $m^s(T) = m(T)$, $P^s(T) = P(T)$. The matrix $F(\cdot)$ is the Jacobian matrix of the function $f(\cdot)$ with the elements

$$F_{ij} = \frac{\partial f_i}{\partial x_j}. \tag{48}$$

By comparing the Equations (47) to (40) it is easy to see that the relationship between the approximations is

$$\begin{aligned} f(X(t),t)\,w_m &\Longleftrightarrow f(m(t),t) \\ f(X(t),t)\,W\,X^T(t) &\Longleftrightarrow F(m(t),t)\,P(t). \end{aligned} \tag{49}$$

The differences between the computational complexities arise from these terms only. The left hand side terms can also be written in terms of the summation form of the unscented transformation (see, Section 2.1) as follows:

$$\begin{aligned} f(X(t),t)\,w_m &= \sum_{i=1}^{2n+1} W_{i-1}^{(m)}\,f(X_i(t),t) \triangleq \mu(t) \\ f(X(t),t)\,W\,X^T(t) &= \sum_{i=1}^{2n+1} W_{i-1}^{(c)}\,\left(f(X_i(t),t) - \mu(t)\right)\left(X_i(t) - m(t)\right)^T, \end{aligned} \tag{50}$$

which are computationally lighter than the matrix forms. The actual number of computations depends on the model, but roughly we have the following:

- The linearization based smoother requires single evaluation of the function f (n scalar evaluations) and single evaluation of the Jacobian F (n^2 scalar evaluations).
- The unscented transformation based smoother requires $2n + 1$ evaluations of the function f (total of $2n^2 + n$ scalar evaluations).

- In the linearization based smoother the mean and covariance computations require one product of two $n \times n$ matrices, which takes roughly n^3 additions and multiplications.
- In the unscented smoother the mean and covariance computations require computation of weighted sum of $2n + 1$ length n-vectors ($2n^2 + n$ operations), computation of weighted sum of $2n + 1$ size $n \times n$ matrices ($2n^3 + n^2$ operations) and a single Cholesky factorization of $n \times n$ matrix (roughly $n^3/2$ operations).

That is, as with the corresponding filters [14], it is quite safe to say that the computational requirements of the unscented transformation based smoother are roughly 2–3 times the requirements of the linearization based filter.

The analysis above only applies to the mean/covariance form of the URTSS and the computational requirements of the sigma point form (44) are a bit higher. The computational requirements depend much on the way that the algorithm is actually implemented. The extra computations are due to the 6 additional matrix products with the lower triangular $A^s(t)$ and its inverse, which require roughly n^3 operations each. That is, the complexity of the sigma point form of URTSS can be even 3–4 times the complexity of the mean/covariance form. But of course, the complexity of the square root version of the linearization based smoother is quite high also.

5 Numerical Example

In this section, the performance of the continuous-time unscented Rauch-Tung-Striebel smoother is tested in the continuous-time version of the nonlinear filtering problem presented as an example in [23], which was also used for demonstrating the performance of UKBF in [14]. Both the state $x(t)$ and the measurements $y(t)$ are one-dimensional continuous-time processes and the state space model is:

$$\begin{aligned} dx(t)/dt &= -\sin x(t) + e(t) \\ y(t) &= \frac{1}{2}\sin x(t) + n(t), \end{aligned} \tag{51}$$

where $e(t)$ and $n(t)$ are continuous-time white noise processes with spectral densities $q_c = 0.01$ and $r_c = 0.004$, respectively. The simulation was performed over time period of 5 seconds using the Euler-Maruyama scheme [39] and with time steps of $\Delta t = 0.01$. In the simulation, the initial states of trajectories were drawn from zero mean Gaussian distribution with unity variance and the initial conditions for filters (and smoothers) were $m_0 = 0$, $P_0 = 1$.

A typical initial period behavior of the filters and smoothers is illustrated in

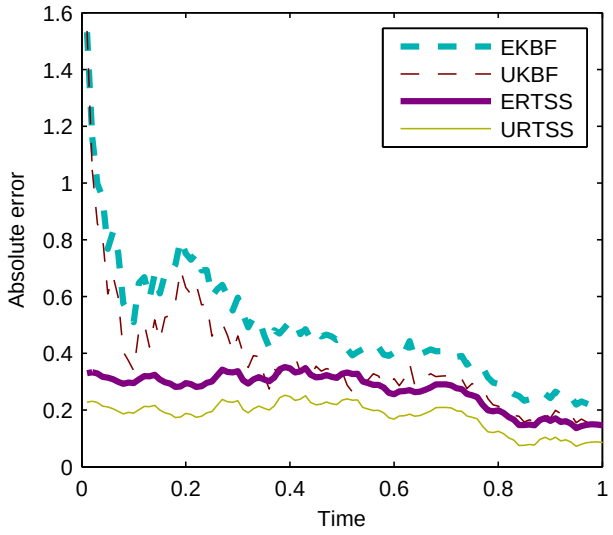

Fig. 1. Illustration of initial period error behavior in continuous-time non-linear filtering and smoothing problem.

Algorithm	*RMSE*	*RMSE*($t = 0$)	*RMSE*($t > 2$)
EKBF	0.26	0.80	0.17
UKBF	0.22	0.80	0.12
ERTSS	0.20	0.29	0.10
URTSS	0.15	0.24	0.10

Table 1
RMSE values of the signal for the whole simulation period $0 < t < 5$, for the initial state $t = 0$ and for the stabilized period $2 < t < 5$ averaged over 1000 Monte Carlo runs.

Figure 1. It can be seen that initially errors of the filters are large, because of the uncertain initial conditions. After approximately 0.6 seconds the filter errors decrease closer to the level of smoothers. In the smoother results the initial uncertainty has been resolved by the later observations and there is only slight initial error transient, which is most likely due to the Gaussian approximations used in the algorithms.

The root mean squared error (RMSE) results of 1000 Monte Carlo runs with extended Kalman-Bucy filter (EKBF), unscented Kalman-Bucy filter (UKBF), extended Rauch-Tung-Striebel smoother (ERTSS) and unscented Rauch-Tung-Striebel smoother (URTSS) are shown in Table 1. The following errors are given in the table:

- *RMSE* is the overall root mean squared error of the algorithm averaged over the whole simulation period $0 < t < 5$.
- *RMSE*($t = 0$) is the error in the initial state estimate. Note that the initial

state estimate of all the filters is just the a priori mean $m_0 = 0$ and thus the theoretical mean RMSE of filters is $\sqrt{2/\pi}$.

- *RMSE($t > 2$)* is the error average over the time period $2 < t < 5$. During this period the filters can be assumed to have already converged and thus this measures the RMSE errors after the initial transient is over.

As expected, all the errors of smoothers are significantly lower than of the filters. The difference is the largest at the initial time step, because at that time step the filter estimates are just the prior means whereas the smoother estimates are based on the whole interval of measurements. In the steady state period $t > 2$ the errors between filters and smoothers are smaller, but still significant. With this particular model, the errors of UKBF and URTS are lower than those of EKBF and ERTS, respectively, which indicates that in this case the unscented transformation seems to provide better approximation to the non-linearities in the model.

In this small dimensional simulation, the computational requirements of the linearization and unscented transformation based methods are almost the same. The amount of computations needed by UT based method might be a bit higher, but not more than 1.5 times the computations needed by the linearization based method.

6 Conclusion

In this article we have derived a new unscented transformation based approximation method to fixed-interval optimal smoothing of continuous-time state space models. The method has been derived by taking the formal continuous-time limit of the unscented Rauch-Tung-Striebel smoother. Both mean-covariance form and sigma-point form of the equations have been presented. The performance of the new smoother has been demonstrated in a simulated continuous-time filtering and smoothing problem.

References

[1] J. Meditch, Stochastic Optimal Linear Estimation and Control, McGraw-Hill, 1969.

[2] M. Grewal, R. Miyasako, J. Smith, Application of fixed point smoothing to the calibration, alignment and navigation data of inertial navigation systems, in: Position Location and Navigation Symposium, 1988, pp. 476–479.

[3] P. Tam, D. Tam, J. Moore, Fixed-lag demodulation of discrete noisy measurements of FM signals, Automatica 9 (1973) 725–729.

[4] Y. Bar-Shalom, X.-R. Li, T. Kirubarajan, Estimation with Applications to Tracking and Navigation, Wiley, New York, 2001.

[5] W. Fong, S. J. Godsill, A. Doucet, M. West, Monte Carlo smoothing with application to audio signal enhancement, IEEE Transactions on Signal Processing 50(2) (2002) 438–449.

[6] Z. Ghahramani, S. T. Roweis, Learning nonlinear dynamical systems using an EM algorithm, in: Neural Information Processing Systems 11 (NIPS'98), 1999, pp. 431–437.

[7] G. Kitagawa, Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, Journal of Computational and Graphical Statistics 5 (1996) 1–25.

[8] S. Särkkä, A. Vehtari, J. Lampinen, Rao-Blackwellized particle filter for multiple target tracking, Information Fusion Journal 8(1) (2007) 2–15.

[9] J. L. Crassidis, J. L. Junkins, Optimal Estimation of Dynamic Systems, Chapman & Hall/CRC, 2004.

[10] S. J. Julier, J. K. Uhlmann, A general method of approximating nonlinear transformations of probability distributions, Tech. rep., Robotics Research Group, Department of Engineering Science, University of Oxford (1995).

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

[12] E. A. Wan, R. Van der Merwe, The unscented Kalman filter, in: S. Haykin (Ed.), Kalman Filtering and Neural Networks, Wiley, 2001, Ch. 7.

[13] S. Särkkä, Unscented Rauch-Tung-Striebel smoother, IEEE Transactions on Automatic Control 53(3) (2008) 845–849.

[14] S. Särkkä, On unscented Kalman filtering for state estimation of continuous-time nonlinear systems, IEEE Transactions on Automatic Control 52(9) (2007) 1631–1641.

[15] I. Karatzas, S. E. Shreve, Brownian Motion and Stochastic Calculus, Springer-Verlag, New York, 1991.

[16] B. Øksendal, Stochastic Differential Equations: An Introduction with Applications, 6th Edition, Springer, New York, 2003.

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

[18] R. E. Kalman, A new approach to linear filtering and prediction problems, Transactions of the ASME, Journal of Basic Engineering 82 (1960) 35–45.

[19] R. E. Kalman, R. S. Bucy, New results in linear filtering and prediction theory, Transactions of the ASME, Journal of Basic Engineering 83 (1961) 95–108.

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

[21] R. S. Bucy, Nonlinear filtering theory, IEEE Transactions on Automatic Control 10 (1965) 198–198.

[22] A. H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.

[23] A. Gelb, Applied Optimal Estimation, The MIT Press, Cambridge, MA, 1974.

[24] B. D. O. Anderson, J. B. Moore, Optimal Filtering, Prentice-Hall, 1979.

[25] S. J. Julier, J. K. Uhlmann, H. F. Durrant-Whyte, A new approach for filtering nonlinear systems, in: Proceedings of the 1995 American Control, Conference, Seattle, Washington, 1995, pp. 1628–1632.

[26] S. J. Julier, J. K. Uhlmann, Unscented filtering and nonlinear estimation, Proceedings of the IEEE 92(3) (2004) 401–422.

[27] A. Doucet, N. de Freitas, N. Gordon, Sequential Monte Carlo Methods in Practice, Springer, New York, 2001.

[28] B. Ristic, S. Arulampalam, N. Gordon, Beyond the Kalman Filter, Artech House, Norwood, MA, 2004.

[29] H. E. Rauch, Solutions to the linear smoothing problem, IEEE Transactions on Automatic Control AC-8 (1963) 371–372.

[30] H. E. Rauch, F. Tung, C. T. Striebel, Maximum likelihood estimates of linear dynamic systems, AIAA Journal 3(8) (1965) 1445–1450.

[31] D. Fraser, J. Potter, The optimum linear smoother as a combination of two optimum linear filters, IEEE Transactions on Automatic Control 14(4) (1969) 387–390.

[32] A. P. Sage, J. L. Melsa, Estimation Theory with Applications to Communications and Control, McGraw-Hill Book Company, 1971.

[33] P. Maybeck, Stochastic Models, Estimation and Control, Volume 2, Academic Press, 1982.

[34] H. Cox, On the estimation of state variables and parameters for noisy dynamic systems, IEEE Transactions on Automatic Control 9(1) (1964) 5–12.

[35] C. T. Leondes, J. B. Peller, E. B. Stear, Nonlinear smoothing theory, IEEE Transactions on Systems Science and Cybernetics 6(1).

[36] J. Meditch, A survey of data smoothing for linear and nonlinear dynamic systems, Automatica 9 (1973) 151–162.

[37] S. Särkkä, Recursive Bayesian inference on stochastic differential equations, Doctoral dissertation, Helsinki University of Technology, Department of Electrical and Communications Engineering (2006).

[38] R. F. Stengel, Optimal Control and Estimation, Dover Publications, New York, 1994.

[39] P. E. Kloeden, E. Platen, Numerical Solution to Stochastic Differential Equations, Springer, New York, 1999.