Gaussian Filtering and Smoothing for Continuous-Discrete Dynamic Systems

Simo Särkkä[a,*], Juha Sarmavuori[b]

[a] *Department of Biomedical Engineering and Computational Science, Aalto University, Espoo, Finland*
[b] *Nokia Siemens Networks, Espoo, Finland*

Abstract

This article is concerned with Bayesian optimal filtering and smoothing of non-linear continuous-discrete state space models, where the state dynamics are modeled with non-linear Itô-type stochastic differential equations, and measurements are obtained at discrete time instants from a non-linear measurement model with Gaussian noise. We first show how the recently developed sigma-point approximations as well as the multi-dimensional Gauss–Hermite quadrature and cubature approximations can be applied to classical continuous-discrete Gaussian filtering. We then derive two types of new Gaussian approximation based smoothers for continuous-discrete models and apply the numerical methods to the smoothers. We also show how the latter smoother can be efficiently implemented by including one additional cross-covariance differential equation to the filter prediction step. The performance of the methods is tested in a simulated application.

Keywords: Bayesian continuous-discrete filtering, Bayesian continuous-discrete smoothing, Gaussian approximation, Kalman filter, Rauch–Tung–Striebel smoother

1. Introduction

Non-linear continuous-discrete optimal filtering and smoothing refer to applications of Bayesian inference to state estimation in dynamic systems, where the time behavior of the system is modeled as a non-linear stochastic differential equation (SDE), and noise-corrupted observations of the state are obtained from a non-linear measurement model. These kind of continuous-discrete state estimation problems arise in many applications, such as, in guidance systems, integrated inertial navigation and passive sensor based target tracking [1, 2, 3]. Solving these estimation problems is very hard, because the SDEs appearing in the dynamic model or the corresponding Fokker–Planck–Kolmogorov partial differential equations cannot typically be solved analytically and approximations must be used. Here we consider the particularly difficult case of non-additive noise which is intractable to many existing methods in the field (cf. [4, 5]).

In this paper, we show how the numerical integration based discrete-time Gaussian filtering and smoothing frameworks presented in [6, 7, 8] can be applied to continuous-discrete models. We first show how the recent numerical integration methods can be applied to the classical continuous-discrete Gaussian filtering framework [9]. The usage of cubature integration method in continuous-discrete filtering was also recently analyzed by Särkkä and Solin [5], and here we generalize those results. The main contributions of this paper are to derive new continuous-discrete Gaussian smoothers by using two different methods: (i) by forming a Gaussian approximation to the partial differential equations of the smoothing solution [10, 11], and (ii) by computing the continuous limit of the discrete-time Gaussian smoother [8] as in [12]. Both of these smoothers consist of differential equations for the smoother mean and covariance. The third main contribution is to derive a novel computationally efficient smoothing method which only requires forward integration of one additional matrix differential equation during the prediction step of the continuous-discrete filter.

1.1. Problem Formulation

This paper is concerned with *Bayesian optimal filtering and smoothing* of non-linear continuous-discrete state space models [1] of the following form:

$$\begin{aligned} dx &= f(x,t)\,dt + L(x,t)\,d\beta \\ y_k &= h_k(x(t_k), r_k), \end{aligned} \tag{1}$$

*Corresponding Author (E-Mail: simo.sarkka@aalto.fi. Address: P.O.Box 12200, 00076 AALTO, Finland. Tel: +358 50 512 4393. Fax: +358 9 470 23182.)

where $x(t) \in \mathbb{R}^n$ is the state and $y_k \in \mathbb{R}^d$ is the measurement at time instant t_k. The functions $f(x,t)$ and $h_k(x,r)$ define the dynamic and measurement models, respectively. Here $\{\beta(t) : t \geq 0\}$ is a s-dimensional Brownian motion with diffusion matrix $Q(t)$ and $\{r_k : k = 1, 2, \ldots\}$ is a Gaussian $\mathrm{N}(0, R_k)$ white random sequence. The processes $\beta(t)$ and r_k as well as the random initial conditions $x(0) \sim \mathrm{N}(m_0, P_0)$ are assumed to be mutually independent. $L(x,t)$ is a matrix valued function which causes the effective diffusion matrix of the process noise to be

$$\Sigma(x,t) = L(x,t)\, Q(t)\, L^T(x,t), \tag{2}$$

and is thus allowed to be state-dependent. In this paper, we interpret the stochastic differential equations (SDE) as Itô-type stochastic differential equations (see, e.g., [13]).

In *continuous-discrete filtering*, the purpose is to compute the following *filtering distributions* which are defined for all $t \geq 0$, not only for the discrete measurement steps:

$$p(x(t) \,|\, y_1, \ldots, y_k), \quad t \in [t_k, t_{k+1}), \quad k = 1, 2, \ldots \tag{3}$$

The *Bayesian optimal continuous-discrete filter* [14, 1, 9] is actually almost the same as the discrete filter—only the prediction step is replaced with solving of the *Fokker–Planck–Kolmogorov (FPK) partial differential equation*.

In *continuous-discrete smoothing* we are interested in computing smoothing distributions of the form

$$p(x(t) \,|\, y_1, \ldots, y_K), \qquad t \in [t_0, t_K]. \tag{4}$$

The formal Bayesian filtering and smoothing solutions to the state estimation problem—including the continuous-discrete special case—have already been around since the 60's–70's [15, 1, 10, 11, 16] and are in that sense well known. However, the only way to solve the formal Bayesian filtering and smoothing equations is by approximation, as the closed-form solution is available only for the linear Gaussian case [17, 18, 19] and for a few other isolated special cases (see, e.g., [20]). Although non-linear continuous-discrete optimal filtering and smoothing are mature subjects, the approximations have concentrated to Taylor series based methods [21, 22, 23, 16, 9] and other methods have received considerably less attention.

During the last few decades, the speed of computers has increased exponentially, and due to that, numerical integration methods and other computational methods have developed rapidly. Thus more accurate approximations to the formal filtering and smoothing equations are tractable than before. In particular, the sigma-point based unscented transform was introduced as an alternative to Taylor series approximations for discrete-time filters and estimators in [24, 25, 26] and the extension to smoothing problems was presented in [27]. The idea was extended to a full numerical integration based discrete-time Gaussian filtering and smoothing framework in [6, 7, 8]. Note that the Gaussian approximations themselves date back to the 60's–80's [28, 1, 9], but the contributions of the recent articles are in application of more general numerical integration methods to the problem.

Non-linear continuous and continuous-discrete smoothing has been more recently studied in [29, 30], and applications of the unscented (sigma-point) transform and related approximations to continuous-discrete (and continuous-time) filtering and smoothing have been proposed in [30, 31, 32, 12]. The extension of the cubature Kalman filter [33] to continuous-discrete filtering problems—using Itô–Taylor series based approximations—has also been recently studied in [4], and its relationship to the classical approach to continuous-discrete filtering was analyzed in [5].

In this paper we only consider Gaussian approximations, but obviously other approximations exist as well. One possible approach is to use MCMC (Markov chain Monte Carlo) based methods for sampling from the state posterior (see, e.g., [34, 35, 36]). It is also possible to use simulation (sequential Monte Carlo) based particle filtering and smoothing methods (see, e.g., [37, 38]) or approximate the solution of the Fokker–Planck–Kolmogorov equation numerically (see, e.g., [39, 40]). Although these methods are more accurate in some cases, they typically are computationally more demanding than Gaussian approximations.

2. Continuous-Discrete Gaussian Filtering

2.1. Gaussian Filter for the Continuous-Discrete Problem

In the filtering algorithms presented in this paper, we use the classical Gaussian filtering [1, 9, 6, 7] approach where the idea is to employ the approximation

$$p(x(t) \,|\, y_1, \ldots, y_k) \approx \mathrm{N}(x(t) \,|\, m(t), P(t)). \tag{5}$$

That is, we replace the true expectations with respect to $x(t)$ with expectations over the Gaussian approximation. The means $m(t)$ and covariances $P(t)$ are computed via the following algorithm:

1. *Prediction step:* Integrate the following mean and covariance differential equations starting from the

mean $m(t_{k-1})$ and covariance $P(t_{k-1})$ on the previous update time, to the time t_k:

$$\frac{dm}{dt} = \mathrm{E}[f(x,t)]$$
$$\frac{dP}{dt} = \mathrm{E}[(x-m)\, f^T(x,t)] + \mathrm{E}[f(x,t)\,(x-m)^T] + \mathrm{E}[\Sigma(x,t)], \tag{6}$$

where $\mathrm{E}[\cdot]$ denotes the expectation with respect to $x \sim \mathrm{N}(m,P)$. The results of the prediction are denoted as $m(t_k^-)$, $P(t_k^-)$, where the minus at superscript means 'infinitesimally before the time t_k'.

2. *Update step:* This is the same as the discrete-time filter update step (see, e.g., [6, 7]) which in the present non-additive case can be written as

$$\begin{aligned}
\mu_k &= \iint h_k(x,r)\, \mathrm{N}(x \,|\, m(t_k^-), P(t_k^-))\, \mathrm{N}(r \,|\, 0, R_k)\, dx\, dr \\
S_k &= \iint (h_k(x,r) - \mu_k)\,(h_k(x,r) - \mu_k)^T \\
&\quad \times \mathrm{N}(x \,|\, m(t_k^-), P(t_k^-))\, \mathrm{N}(r \,|\, 0, R_k)\, dx\, dr \\
D_k &= \iint (x - m(t_k^-))\,(h_k(x,r) - \mu_k)^T \\
&\quad \times \mathrm{N}(x \,|\, m(t_k^-), P(t_k^-))\, \mathrm{N}(r \,|\, 0, R_k)\, dx\, dr \\
K_k &= D_k\, S_k^{-1} \\
m(t_k) &= m(t_k^-) + K_k\,(y_k - \mu_k) \\
P(t_k) &= P(t_k^-) - K_k\, S_k\, K_k^T.
\end{aligned} \tag{7}$$

If the function $x \mapsto f(x,t)$ is differentiable, we can simplify the covariance prediction equation by using the following property of Gaussian random variables [41]:

$$\mathrm{E}\left[f(x,t)\,(x-m)^T\right] = \mathrm{E}\left[F_x(x,t)\right] P, \tag{8}$$

where $F_x(x,t)$ is the Jacobian matrix of $f(x,t)$ with respect to x, with elements $[F_x]_{ij} = \partial f_i / \partial x_j$, and $\mathrm{E}[\cdot]$ denotes the expectation with respect to $x \sim \mathrm{N}(x \,|\, m, P)$. The prediction Equations (6) can thus be equivalently written as

$$\begin{aligned}
\frac{dm}{dt} &= \mathrm{E}[f(x,t)] \\
\frac{dP}{dt} &= P\, \mathrm{E}[F_x(x,t)]^T + \mathrm{E}[F_x(x,t)]\, P + \mathrm{E}[\Sigma(x,t)].
\end{aligned} \tag{9}$$

2.2. Numerical Integration of Filtering Equations

As we saw in Section 2.1, the Gaussian filter update step is identical in both discrete and continuous-discrete cases. Thus here we only consider implementation of the differential equations on the prediction step. In order to implement the prediction equations given in (6) or (9), we need means to approximate the following kind of integrals over Gaussian distributions:

$$\mathrm{E}[g(x,t)] = \int g(x,t)\, \mathrm{N}(x \,|\, m, P)\, dx. \tag{10}$$

The classical way [1, 9] is to eliminate the non-linearities by forming analytical approximations to the drift and diffusion as follows:

$$\begin{aligned}
f(x,t) &\approx f(m,t) + F_x(m,t)\,(x-m) \\
\Sigma(x,t) &\approx \Sigma(m,t),
\end{aligned} \tag{11}$$

where F_x denotes the Jacobian matrix of f with respect to x. In this case, of course, $f(x,t)$ has to be differentiable with respect to x. The prediction equations (6) now reduce to the classical first-order continuous-discrete extended Kalman filter (EKF) prediction equations [1]:

$$\begin{aligned}
\frac{dm}{dt} &= f(m,t) \\
\frac{dP}{dt} &= P\, F_x^T(m,t) + F_x(m,t)\, P + \Sigma(m,t).
\end{aligned} \tag{12}$$

Exactly the same equations would have been obtained by using the approximation $\mathrm{E}[F_x(x,t)] \approx F_x(m,t)$ in the prediction equations (9). The linearization is obviously exact only for linear functions and for this reason does not work well for highly non-linear problems (see [26, 7, 33]).

Another general way [7] is to approximate the integrals as a weighted sum:

$$\mathrm{E}[g(x,t)] \approx \sum_i W^{(i)}\, g(x^{(i)}, t), \tag{13}$$

where $x^{(i)}$ and $W^{(i)}$ are the sigma-points and weights which have been selected using a method specific deterministic rule. In the multidimensional Gauss–Hermite integration [6], unscented transform [24, 25, 26] and cubature integration [7, 33] the sigma-points are selected as follows:

$$x^{(i)} = m + \sqrt{P}\, \xi_i, \tag{14}$$

where the matrix square root is defined by $P = \sqrt{P}\,\sqrt{P}^T$, and the vectors ξ_i and the weights $W^{(i)}$ are selected as follows:

- *Gauss–Hermite* integration method (the product rule based method) uses set of m^n vectors ξ_i which have been formed as a Cartesian product of the zeros of the Hermite polynomials of order m. The weights $W^{(i)}$ are formed as products of the corresponding one-dimensional Gauss–Hermite integration weights (see, [6, 7], for details).

- *Unscented transform* uses a zero vector and $2n$ scaled coordinate vectors e_i as follows:

$$\xi_0 = 0$$
$$\xi_i = \begin{cases} \sqrt{\lambda + n}\, e_i & , \quad i = 1, \ldots, n \\ -\sqrt{\lambda + n}\, e_{i-n} & , \quad i = n+1, \ldots, 2n, \end{cases} \tag{15}$$

and the weights are defined as follows:

$$W^{(0)} = \begin{cases} \frac{\lambda}{n+\lambda} & , \quad \text{For the 1st term in (10)}, \\ \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta) & , \quad \text{For the 2nd term in (10)}, \end{cases}$$
$$W^{(i)} = \frac{1}{2(n+\lambda)}, \qquad i = 1, \ldots, 2n, \tag{16}$$

where $\lambda = \alpha^2(n + \kappa) - n$ and α, β, and κ are parameters of the method. The parameters α and κ determine the spread of the sigma-points around the mean, and β is an additional parameter that can be used for incorporating prior information on the distribution of x [25].

- *Cubature method* (spherical 3rd degree) uses only $2n$ vectors as follows:

$$\xi_i = \begin{cases} \sqrt{n}\, e_i & , \quad i = 1, \ldots, n \\ -\sqrt{n}\, e_{i-n} & , \quad i = n+1, \ldots, 2n, \end{cases} \tag{17}$$

and the weights are defined as $W^{(i)} = 1/(2n)$ for $i = 1, \ldots, 2n$. Note that the cubature rule is a special case of the unscented transform with the parameters $\alpha = 1, \beta = 0$, and $\kappa = 0$.

The sigma-point methods above lead to the following approximations to the prediction step differential equations:

$$\begin{aligned} \frac{dm}{dt} &= \sum_i W^{(i)}\, f(m + \sqrt{P}\,\xi_i, t) \\ \frac{dP}{dt} &= \sum_i W^{(i)}\, f(m + \sqrt{P}\,\xi_i, t)\,\xi_i^T\, \sqrt{P}^T \\ &\quad + \sum_i W^{(i)}\, \sqrt{P}\,\xi_i\, f^T(m + \sqrt{P}\,\xi_i, t) \\ &\quad + \sum_i W^{(i)}\, \Sigma(m + \sqrt{P}\,\xi_i, t), \end{aligned} \tag{18}$$

where in the case of the unscented transform we need to use different weights for the first two terms in the covariance differential equation (cf. (16) and [31]).

The sigma-point methods above have the advantage that they all are exact for monomials up to order 3 [7, 33], in contrast to linearization, which is only exact up to order 1. It is possible to form higher order approximations using higher order terms in the Taylor series, but the computation of the required higher order derivatives in closed form quickly becomes complicated and error prone. With suitable selection of parameters, the unscented transform can also be made exact for some fourth order monomials as well [26]. The Gauss–Hermite method can be made exact for monomial up to an arbitrary order.

Once the Gaussian integral approximation has been selected, the solutions to the resulting ordinary differential equations (ODE) can be computed, for example, by 4th order Runge-Kutta method or any similar numerical ODE solution method.

3. Continuous-Discrete Gaussian Smoothing

3.1. Gaussian Approximation to Formal Smoothing Equation

According to [10, 11] the expectation of an arbitrary function $g(x)$ over the smoothed distribution solves the partial differential equation

$$\frac{d\mathrm{E}_s[g(x)]}{dt} = \mathrm{E}_s[K_1\, g(x)], \tag{19}$$

where $\mathrm{E}_s[\cdot]$ denotes the expectation with respect to the smoothing distribution and the operator K_1 is given as

$$\begin{aligned} K_1 g(x) = &\sum_i f_i(x,t)\frac{\partial g(x)}{\partial x_i} - \frac{1}{2}\sum_{ij}\Sigma_{ij}(x,t)\frac{\partial^2 g(x)}{\partial x_i \partial x_j} \\ &- \sum_{ij}\frac{\partial g(x)}{\partial x_i}\frac{\partial \Sigma_{ij}(x,t)}{\partial x_j} \\ &- \frac{1}{p(x,t)}\sum_{ij}\Sigma_{ij}(x,t)\frac{\partial g(x)}{\partial x_i}\frac{\partial p(x,t)}{\partial x_j}, \end{aligned} \tag{20}$$

where $p(x, t)$ denotes the filtering distribution. Assume that we have already employed the Gaussian filter to the estimation problem and thus obtained the approximation:

$$p(x,t) \triangleq p(x(t) \,|\, y_1, \ldots, y_k) \approx \mathrm{N}(x(t) \,|\, m(t), P(t)). \tag{21}$$

Then by direct calculation we get

$$\begin{aligned} \frac{\partial p(x,t)}{\partial x} &\approx -P^{-1}\,(x - m)\,\mathrm{N}(x \,|\, m, P) \\ &\approx -P^{-1}\,(x - m)\, p(x,t), \end{aligned} \tag{22}$$

and thus the operator in Equation (20) can be approximated as

$$
\begin{aligned}
K_1 g(x) \approx & \sum_i f_i(x,t) \frac{\partial g(x)}{\partial x_i} - \frac{1}{2} \sum_{ij} \Sigma_{ij}(x,t) \frac{\partial^2 g(x)}{\partial x_i \partial x_j} \\
& - \sum_{ij} \frac{\partial g(x)}{\partial x_i} \frac{\partial \Sigma_{ij}(x,t)}{\partial x_j} \\
& + \sum_{ij} \Sigma_{ij}(x,t) \frac{\partial g(x)}{\partial x_i} [P^{-1}(x-m)]_j.
\end{aligned} \tag{23}
$$

By first selecting $g(x) = x$, then $g(x) = (x - m^s)(x - m^s)^T$, where $m^s = \mathrm{E}_s[x]$, and by finally taking the expectations with respect to the smoothing distribution, we obtain the following approximate differential equations for the smoother mean and covariance:

$$
\begin{aligned}
\frac{dm^s}{dt} &= \mathrm{E}_s[f(x,t)] - \mathrm{E}_s\left[\sum_j \frac{\partial \Sigma_j(x,t)}{\partial x_j}\right] \\
&\quad + \mathrm{E}_s[\Sigma(x,t)\, P^{-1}(x-m)]
\end{aligned} \tag{24}
$$

$$
\begin{aligned}
\frac{dP^s}{dt} &= \mathrm{E}_s[f(x,t)\,(x-m^s)^T] + \mathrm{E}_s[(x-m^s)\, f^T(x,t)] \\
&\quad + \mathrm{E}_s[\Sigma(x,t)\, P^{-1}(x-m)\,(x-m^s)^T] \\
&\quad + \mathrm{E}_s[(x-m^s)\,(x-m)^T\, P^{-1}\,\Sigma(x,t)] \\
&\quad - \mathrm{E}_s\left[\sum_j \left(\frac{\partial \Sigma_j(x,t)}{\partial x_j}\right)(x-m^s)^T\right] \\
&\quad - \mathrm{E}_s\left[\sum_j (x-m^s) \left(\frac{\partial \Sigma_j(x,t)}{\partial x_j}\right)^T\right] - \mathrm{E}_s[\Sigma(x,t)],
\end{aligned} \tag{25}
$$

where Σ_i denotes the ith column of Σ and the expectations are taken with respect to the smoothed distribution of x at time t.

Using integration by parts, the diffusion matrix derivatives in the mean and covariance equations can be eliminated, which results in the following smoothing equations:

$$
\begin{aligned}
\frac{dm^s}{dt} &= \mathrm{E}_s[f(x,t)] - \mathrm{E}_s[\Sigma(x,t)\,(P^s)^{-1}\,(x-m^s)] \\
&\quad + \mathrm{E}_s[\Sigma(x,t)\, P^{-1}\,(x-m)] \\
\frac{dP^s}{dt} &= \mathrm{E}_s[f(x,t)\,(x-m^s)^T] \\
&\quad + \mathrm{E}_s[(x-m^s)\, f^T(x,t)] \\
&\quad + \mathrm{E}_s[\Sigma(x,t)\, P^{-1}\,(x-m)\,(x-m^s)^T] \\
&\quad + \mathrm{E}_s[(x-m^s)\,(x-m)^T\, P^{-1}\,\Sigma(x,t)] \\
&\quad - \mathrm{E}_s[\Sigma(x,t)\,(P^s)^{-1}\,(x-m^s)\,(x-m^s)^T] \\
&\quad - \mathrm{E}_s[(x-m^s)\,(x-m^s)^T\,(P^s)^{-1}\,\Sigma(x,t)] \\
&\quad + \mathrm{E}_s[\Sigma(x,t)]
\end{aligned} \tag{26}
$$

which should be integrated from the terminal condition $m^s(T) = m(T), P^s(T) = P(T)$ to the initial time $t = 0$. Note that if the process noise is additive, that is $\Sigma(x,t) = \Sigma(t)$, the mean and covariance equations reduce to

$$
\begin{aligned}
\frac{dm^s}{dt} &= \mathrm{E}_s[f(x,t)] + \Sigma(t)\, P^{-1}\,[m^s - m] \\
\frac{dP^s}{dt} &= \mathrm{E}_s[f(x,t)\,(x-m^s)^T] \\
&\quad + \mathrm{E}_s[(x-m^s)\, f^T(x,t)] \\
&\quad + \Sigma(t)\, P^{-1}\, P^s + P^s\, P^{-1}\,\Sigma(t) - \Sigma(t).
\end{aligned} \tag{27}
$$

The equations, which we shall call *Type I continuous-discrete Gaussian smoother* equations, are obtained when instead of the true expectations, we use the following approximations in the Equations (26) or in the additive Equations (27):

$$
\mathrm{E}_s[g(x,t)] \approx \int g(x,t)\, \mathrm{N}(x \,|\, m^s, P^s)\, dx \tag{28}
$$

If the function $f(x,t)$ is differentiable, the Equation (8) can be used for rewriting the first two terms in the covariance formulas.

3.2. Continuous-Time Limit of Discrete-Time Smoother

An alternative way to derive continuous-discrete nonlinear smoothing equations is by computing the formal continuous-time limit of the discrete-time smoothing equations, as was done for the unscented Rauch–Tung–Striebel (RTS) smoother in [12]. The idea is to start from the discretized approximation to the dynamic model

$$
\begin{aligned}
x(t+\delta t) = x(t) &+ f(x(t),t)\,\delta t \\
&+ L(x(t),t)\,\delta\beta(t) + o(\delta t),
\end{aligned} \tag{29}
$$

where $\delta\beta(t) \sim \mathrm{N}(0, Q(t)\,\delta t)$, apply a discrete-time smoother, and then take the limit $\delta t \to 0$. If we follow the derivation presented in [12], except that we replace the discrete time RTS smoother [27] with the more general Gaussian smoother [8], we get the following differential equations for the smoother mean and covariance, which are here called the *Type II continuous-discrete Gaussian smoother* equations:

$$\begin{aligned}
\frac{dm^s}{dt} &= \mathrm{E}[f(x,t)] + \{\mathrm{E}[f(x,t)\,(x-m)^T] \\
&\quad + \mathrm{E}[\Sigma(x,t)]\}\,P^{-1}[m^s - m] \\
\frac{dP^s}{dt} &= \{\mathrm{E}[f(x,t)\,(x-m)^T] + \mathrm{E}[\Sigma(x,t)]\}\,P^{-1}\,P^s \\
&\quad + P^s\,P^{-1}\,\{\mathrm{E}[f(x,t)\,(x-m)^T] + \mathrm{E}[\Sigma(x,t)]\}^T \\
&\quad - \mathrm{E}[\Sigma(x,t)].
\end{aligned} \tag{30}$$

Note that here the expectations are taken with respect to the Gaussian approximation to the filtering distribution, that is $x \sim \mathrm{N}(m, P)$. Again, we could rewrite some of the terms in the equations with Equation (8), but there seems to be no obvious advantage in doing that. The additive form of this equation can be simply obtained by replacing the terms $\mathrm{E}[\Sigma(x,t)]$ with $\Sigma(t)$.

3.3. *Forward-Time Differential Equations*

Because the expectations in Equations (30) are with respect to the filtering distributions, not with respect to the smoothing distributions, the equations can be seen to have the form

$$\begin{aligned}
\frac{dm^s}{dt} &= u(t) + A(t)\,[m^s - m] \\
\frac{dP^s}{dt} &= A(t)\,P^s + P^s\,A^T(t) - S(t),
\end{aligned} \tag{31}$$

where the following are functions of the filtering solution only:

$$\begin{aligned}
u(t) &= \mathrm{E}[f(x,t)] \\
A(t) &= \{\mathrm{E}[f(x,t)\,(x-m)^T] + \mathrm{E}[\Sigma(x,t)]\}\,P^{-1} \\
S(t) &= \mathrm{E}[\Sigma(x,t)].
\end{aligned} \tag{32}$$

Let $\Phi(t,s)$ be the transition matrix of the system $dx/dt = A(t)\,x$. Then given the smoothed solution at time t_k, that is $m^s(t_k)$, we can solve the mean equation at time t_{k+1}^- as follows:

$$\begin{aligned}
m^s(t_{k+1}^-) &= \Phi(t_{k+1}^-, t_k)\,m^s(t_k) \\
&\quad + \int_{t_k}^{t_{k+1}^-} \Phi(t_{k+1}^-, s)\,[u(s) - A(s)\,m(s)]\,ds \\
&= \Phi(t_{k+1}^-, t_k)\,m^s(t_k) + m(t_{k+1}^-) - \Phi(t_{k+1}^-, t_k)\,m(t_k).
\end{aligned} \tag{33}$$

Similarly, for the covariance we get

$$\begin{aligned}
P^s(t_{k+1}^-) &= \Phi(t_{k+1}^-, t_k)\,P^s(t_k)\,\Phi^T(t_{k+1}^-, t_k) \\
&\quad - \int_{t_k}^{t_{k+1}^-} \Phi(t_{k+1}^-, s)\,S(s)\,\Phi^T(t_{k+1}^-, s)\,ds \\
&= \Phi(t_{k+1}^-, t_k)\,P^s(t_k)\,\Phi^T(t_{k+1}^-, t_k) \\
&\quad - \Phi(t_{k+1}^-, t_k)\,P(t_k)\,\Phi^T(t_{k+1}^-, t_k) + P(t_{k+1}^-).
\end{aligned} \tag{34}$$

Because the smoother mean and covariance are continuous with respect to time, we can replace the terms t_{k+1}^- in them with t_{k+1}. As the result we get that the filter and smoother means and covariance are related by the backward recursions

$$\begin{aligned}
m^s(t_k) &= m(t_k) + \Phi(t_k, t_{k+1}^-)\,[m^s(t_{k+1}) - m(t_{k+1}^-)] \\
P^s(t_k) &= P(t_k) \\
&\quad + \Phi(t_k, t_{k+1}^-)\,[P^s(t_{k+1}) - P(t_{k+1}^-)]\,\Phi^T(t_k, t_{k+1}^-).
\end{aligned} \tag{35}$$

The differential equation for the transition matrix $\Phi(t_k, t)$ can be written as

$$\begin{aligned}
\frac{\partial \Phi(t_k, t)}{\partial t} &= -\Phi(t_k, t)\,A(t) \\
&= -\Phi(t_k, t)\,\{\mathrm{E}[f(x)\,(x-m)^T] + \mathrm{E}[\Sigma(x,t)]\}\,P^{-1}.
\end{aligned} \tag{36}$$

Let's now define $C_k(t) = \Phi(t_k, t)\,P(t)$. By computing its time derivative, and by using the Equations (6) and (36), we get the differential equation

$$\begin{aligned}
\frac{dC_k}{dt} &= \frac{\partial \Phi(t_k, t)}{\partial t}\,P(t) + \Phi(t_k, t)\,\frac{dP(t)}{dt} \\
&= C_k(t)\,P^{-1}(t)\,\mathrm{E}[f(x)\,(x - m(t))^T]^T.
\end{aligned} \tag{37}$$

By rewriting the transition matrix in Equations (35) in terms of $C_k(t_{k+1}^-)$, the Type II Gaussian smoother in the previous section can be *equivalently written* in the following form which we call *Type III continuous-discrete Gaussian smoother*:

$$\begin{aligned}
\frac{dm}{dt} &= \mathrm{E}[f(x)] \\
\frac{dP}{dt} &= \mathrm{E}[f(x)\,(x-m)^T] \\
&\quad + \mathrm{E}[f(x)\,(x-m)^T]^T + \mathrm{E}[\Sigma(x,t)] \\
\frac{dC_k}{dt} &= C_k\,P^{-1}\,\mathrm{E}[f(x)\,(x-m)^T]^T \\
G_{k+1} &= C_k(t_{k+1}^-)\,P^{-1}(t_{k+1}^-) \\
m^s(t_k) &= m(t_k) + G_{k+1}\,[m^s(t_{k+1}) - m(t_{k+1}^-)] \\
P^s(t_k) &= P(t_k) + G_{k+1}\,[P^s(t_{k+1}) - P(t_{k+1}^-)]\,G_{k+1}^T.
\end{aligned} \tag{38}$$

The initial conditions for the mean and covariance differential equations are the filtered mean and covariance,

$m(t_k)$ and $P(t_k)$, and the initial condition for the third equation is $C(t_k) = P(t_k)$. Note that the first three differential equations above are forward-time differential equations and thus can be solved already during the filtering stage, and actually, the first two equations are just the predicted mean and covariance of the continuous-discrete filter. In the filtering stage, we only need to store the filtered values, $m(t_k)$ and $P(t_k)$, predicted values, $m(t_{k+1}^-)$ and $P(t_{k+1}^-)$, and the gains G_{k+1} to be able to compute the smoothing solution later. That is, we do not need to store or recompute the filter solution for each time instant t to be able to compute the smoothing solution.

The Equation (8) can now be used for eliminating the matrix inverse in the third differential equation of (38), because it can also be written as

$$\frac{dC_k}{dt} = C_k \, \mathrm{E}[F_x(x,t)]^T. \tag{39}$$

By comparing Equations (38) to the discrete-time Gaussian smoother in [8], it is easy to see that the smoother equations have the same form as the discrete-time equations if we identify $m_{k+1}^- \triangleq m(t_{k-1}^-)$, $P_{k+1}^- \triangleq P(t_{k+1}^-)$ and $C_{k+1} \triangleq C_k(t_{k+1}^-)$. Thus we can implement the smoother by replacing the first three equations in the discrete-time smoother with the first three equations in the above smoother.

Note that the equations (38) are actually valid for arbitrary time instants $\tau \in [t_k, t_{k+1})$ and thus we can recover the smoother mean and covariance for arbitrary time instants, not just for the measurement times.

3.4. Computational Complexity of Smoother Types

In target tracking scenarios we are often interested in computing the filtering and smoothing results only at the measurement times or at some other discrete times. In such cases the different types of smoothers have significantly different computational and storage requirements. The computational and storage requirements for the smoothers (including the filtering step) are summarized in Table 1, assuming an n-dimensional state, K measurements and N evaluation steps in the numerical integration. The computational requirements are measured as the number of Gaussian integral computations, and the storage requirements as number of stored floating point numbers. As can be seen from the requirements, the Type III smoother has significantly lower computational and storage requirements than the other smoothers.

The differences in the computational requirements are due to the number of Gaussian integrals required in the smoothing step. In Type II and Type III smoothers we do not need to evaluate additional Gaussian integrals during smoothing, provided that we store the results during the filtering step. The storage requirements are different, because in Type I and Type II smoothers we need to store all the N intermediate filtering results between the K measurements to be able to implement the backward smoothing differential equations. In Type II smoother we also need to store the results of the computed Gaussian integrals. In Type III smoother we need to store the predicted mean, predicted covariance and the gain G_{k+1}, but only at the measurement steps.

Table 1: The computational requirements (= the number of Gaussian integral computations) and the storage requirements for different types of smoothers.

Smoother	Integrals	Storage
Type I	$10\,N\,K$	$N\,K\,(n + n^2)$
Type II	$3\,N\,K$	$N\,K\,(2n + 3n^2)$
Type III	$3\,N\,K$	$K\,(2n + 3n^2)$

3.5. Numerical Approximation of Smoothers

In this section we demonstrate the usage of the theory by showing how certain known Taylor series based continuous-discrete (or continuous) smoothers can be derived from the present theory. In addition to that, we also show how the sigma-point type numerical integration methods can be used for approximating the integrals in a similar manner as was done for filtering equations in Section 2.2. We only consider additive process noises, that is, assume that $\Sigma(x,t) = \Sigma(t)$ for notational convenience.

A Type I Taylor series based smoother can be derived by substituting the second order Taylor series expansion of $f(x,t)$ formed around the smoothed mean into the Equations (27):

$$\begin{aligned}
\frac{dm^s}{dt} &= f(m^s,t) + \frac{1}{2}\sum_i \mathrm{tr}\left\{F_{xx}^{(i)}(m^s,t)\,P^s\right\} e_i \\
&\quad + \Sigma(t)\,P^{-1}\,[m^s - m] \\
\frac{dP^s}{dt} &= F_x(m^s,t)\,P^s + P^s\,F_x^T(m^s,t) - \Sigma(t) \\
&\quad + \Sigma(t)\,P^{-1}\,P^s + P^s\,P^{-1}\,\Sigma(t).
\end{aligned} \tag{40}$$

where $F_{xx}^{(i)}$ is the Hessian of f_i and e_i is a unit coordinate vector. These equations are exactly the same as the minimum variance smoothing equations derived from the non-linear smoothing theory in [11]. The non-linear continuous maximum a posteriori (MAP) fixed-interval smoothers given in [42, 23], which have been derived by the invariant embedding approach, can be recovered from the above equations by neglecting the second order derivative term.

For the Type II Taylor series based smoother we expand the function f around the filter mean instead of the smoother mean, which gives:

$$
\begin{aligned}
\frac{dm^s}{dt} &= f(m,t) + \frac{1}{2} \sum_i \mathrm{tr}\left\{ F_{xx}^{(i)}(m,t)\, P \right\} e_i \\
&\quad + [F_x(m,t) + \Sigma(t)\, P^{-1}]\,[m^s - m] \\
\frac{dP^s}{dt} &= F_x(m,t)\, P^s + P^s\, F_x^T(m^s,t) - \Sigma(t) \\
&\quad + \Sigma(t)\, P^{-1}\, P^s + P^s\, P^{-1}\, \Sigma(t).
\end{aligned}
\tag{41}
$$

If we neglect the second order term, we recover the linearization based continuous non-linear smoothers given in [29] and [30].

Substituting the Taylor series approximation to the Type III smoother Equation (38), we get the following approximations to the first three equations of the smoother:

$$
\begin{aligned}
\frac{dm}{dt} &= f(m,t) + \frac{1}{2} \sum_i \mathrm{tr}\left\{ F_{xx}^{(i)}(m,t)\, P \right\} e_i \\
\frac{dP}{dt} &= F_x(m,t)\, P + P\, F_x^T(m,t) + \Sigma(t) \\
\frac{dC_k}{dt} &= C_k\, F_x^T(m,t),
\end{aligned}
\tag{42}
$$

and the smoothing solution would be then given by the last three of the Equations (38).

Sigma-point type numerical integration approximations to the Type I smoothing equations can be obtained by substituting the following approximation to the Equations (27):

$$
\mathrm{E}_s[f(x,t)] \approx \sum_i W^{(i)}\, f(m^s + \sqrt{P^s}\, \xi_i, t). \tag{43}
$$

The vectors ξ_i can be selected, for example, according to one of the rules given in Section 2.2. For the Type I smoother we get the following backward differential equations:

$$
\begin{aligned}
\frac{dm^s}{dt} &= \sum_i W^{(i)}\, f(m^s + \sqrt{P^s}\, \xi_i, t) + \Sigma(t)\, P^{-1}\, [m^s - m] \\
\frac{dP^s}{dt} &= \sum_i W^{(i)}\, f(m^s + \sqrt{P^s}\, \xi_i, t)\, \xi_i^T\, \sqrt{P^s}^T \\
&\quad + \sum_i W^{(i)}\, \sqrt{P^s}\, \xi_i\, f^T(m^s + \sqrt{P^s}\, \xi_i, t) \\
&\quad + \Sigma(t)\, P^{-1}\, P^s + P^s\, P^{-1}\, \Sigma(t) - \Sigma(t).
\end{aligned}
\tag{44}
$$

For the Type II smoother we need the following approximation:

$$
\mathrm{E}[f(x,t)] \approx \sum_i W^{(i)}\, f(m + \sqrt{P}\, \xi_i, t), \tag{45}
$$

and by substituting into Equations (30), we obtain the following backward differential equations for the Type II smoother:

$$
\begin{aligned}
\frac{dm^s}{dt} &= \sum_i W^{(i)}\, f(m + \sqrt{P}\, \xi_i, t) \\
&\quad + \left[\sum_i W^{(i)}\, f(m + \sqrt{P}\, \xi_i, t)\, \xi_i^T\, \sqrt{P}^T + \Sigma(t) \right] \\
&\quad \times P^{-1}\, [m^s - m] \\
\frac{dP^s}{dt} &= \left[\sum_i W^{(i)}\, f(m + \sqrt{P}\, \xi_i, t)\, \xi_i^T\, \sqrt{P}^T + \Sigma(t) \right] P^{-1}\, P^s \\
&\quad + P^s\, P^{-1} \left[\sum_i W^{(i)}\, f(m + \sqrt{P}\, \xi_i, t)\, \xi_i^T\, \sqrt{P}^T + \Sigma(t) \right]^T \\
&\quad - \Sigma(t).
\end{aligned}
\tag{46}
$$

Finally, by substituting into the Equations (38), we get the following forward differential equations for the Type III smoother:

$$
\begin{aligned}
\frac{dm}{dt} &= \sum_i W^{(i)}\, f(m + \sqrt{P}\, \xi_i, t) \\
\frac{dP}{dt} &= \sum_i W^{(i)}\, f(m + \sqrt{P}\, \xi_i, t)\, \xi_i^T\, \sqrt{P}^T \\
&\quad + \sum_i W^{(i)}\, \sqrt{P}\, \xi_i\, f^T(m + \sqrt{P}\, \xi_i, t) + \Sigma(t) \\
\frac{dC_k}{dt} &= C_k\, \sum_i W^{(i)}\, \sqrt{P}^{-1}\, \xi_i\, f^T(m + \sqrt{P}\, \xi_i, t).
\end{aligned}
\tag{47}
$$

3.6. *Computational Complexity of Numerical Integration*

In addition to the selection of the smoother type (cf. Section 3.4), the computational complexity of the filters and smoothers also depends on the numerical integration methods chosen. In particular, in sigma-point methods such as Gauss–Hermite, Cubature and Unscented methods the complexity is directly proportional to the number of sigma-points (cf. Section 2.2). For the accuracy and complexity analysis of the different Gaussian integration methods reader is referred to [7].

4. Simulation

Radar Tracked Coordinate Turn With Multiplicative Noise in Dynamic Model

To compare the performance of the different methods, we use a modified version of the simulation scenario used in [4]. In the original simulation, the dynamic model was a three-dimensional coordinated turn model which can be written in the form:

$$dx = f(x)\,dt + d\beta, \tag{48}$$

where the state is $x = (\epsilon, \dot{\epsilon}, \eta, \dot{\eta}, \zeta, \dot{\zeta}, \omega)$; ϵ, η and ζ are the position coordinates; $\dot{\epsilon}$, $\dot{\eta}$ and $\dot{\zeta}$ the corresponding velocities and ω is the turn rate in (ϵ, η) space. The drift function was $f(x) = (\dot{\epsilon}, -\omega\,\dot{\eta}, \dot{\eta}, \omega\,\dot{\epsilon}, \dot{\zeta}, 0, 0)$ and the Brownian motion β had the diffusion matrix $Q = \mathrm{diag}(0, \sigma_1^2, 0, \sigma_1^2, 0, \sigma_1^2, \sigma_2^2)$. The radar was assumed to be located at origin measuring the range r, azimuth angle θ and elevation angle ϕ with sampling interval T:

$$\begin{pmatrix} r_k \\ \theta_k \\ \phi_k \end{pmatrix} = \begin{pmatrix} \sqrt{\epsilon^2(t_k) + \eta^2(t_k) + \zeta^2(t_k)} \\ \tan^{-1}\left(\eta(t_k)/\epsilon(t_k)\right) \\ \tan^{-1}\left(\zeta(t_k)/\sqrt{\epsilon^2(t_k) + \eta^2(t_k)}\right) \end{pmatrix} + w_k, \tag{49}$$

where $w_k \sim \mathrm{N}(0, R)$ with $R = \mathrm{diag}(\sigma_r^2, \sigma_\theta^2, \sigma_\phi^2)$. As discussed in [4], this is a suitable scenario for testing nonlinear estimators, because it has nonlinear dynamic and measurement models, and it is relatively high-dimensional.

In the above model, the process noise models the effect of turbulence, winds and other such unmodeled effects. The random effects are assumed to be similar in each direction and in particular, independent of the direction of the motion. However, it may be desirable to assume that the random effects have different magnitude in tangential and normal directions with respect to the trajectory. For this reason, we have modified the model such that the process noise is different in tangential and normal directions. The dynamic model can be then written as

$$dx = f(x)\,dt + L(x)\,d\beta', \tag{50}$$

where the state x and drift functions are still the same, but the diffusion coefficient matrix is

$$L(x) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \dot{\epsilon}/u & \dot{\eta}/u_{\epsilon\eta} & \dot{\epsilon}\dot{\zeta}/(u\,u_{\epsilon\eta}) & 0 \\ 0 & 0 & 0 & 0 \\ \dot{\eta}/u & -\dot{\epsilon}/u_{\epsilon\eta} & \dot{\eta}\dot{\zeta}/(u\,u_{\epsilon\eta}) & 0 \\ 0 & 0 & 0 & 0 \\ \dot{\zeta}/u & 0 & -u_{\epsilon\eta}/u & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{51}$$

where $u = \sqrt{\dot{\epsilon}^2 + \dot{\eta}^2 + \dot{\zeta}^2}$ and $u_{\epsilon\eta} = \sqrt{\dot{\epsilon}^2 + \dot{\eta}^2}$. The Brownian motion β' has the diffusion matrix $Q' = \mathrm{diag}(\sigma_t^2, \sigma_h^2, \sigma_v^2, \sigma_\omega^2)$, where σ_t, σ_h, σ_v, σ_ω are the noise densities in tangential, horizontal and vertical directions, and turn rate, respectively.

The modified model is even more challenging and nonlinear than the original model, because it now has a multiplicative, highly nonlinear term in the dynamic model. Due to the multiplicative term, the method presented in [4] is inapplicable to the model and it would be non-trivial to extend that Itô–Taylor expansion method to cope with the multiplicative term.

The measurement noises were selected to be the same as in [4], that is, $\sigma_r = 50\,\mathrm{m}$, $\sigma_\theta = 0.1^\circ$, and $\sigma_\phi = 0.1^\circ$. The initial conditions were $x(0) = (1000\,\mathrm{m}, 0\,\mathrm{ms}^{-1}, 2650\,\mathrm{m}, 150\,\mathrm{ms}^{-1}, 200\,\mathrm{m}, 0\,\mathrm{ms}^{-1}, 6^\circ\mathrm{s}^{-1})$ which were assumed to be known with standard deviations of $100\,\mathrm{m}$ in the positions, $100\,\mathrm{ms}^{-1}$ in the velocities and $1^\circ\mathrm{s}^{-1}$ in the turn rate. The sampling interval was $T = 8\,\mathrm{s}$ and 26 measurements were simulated. Note that these parameters correspond to the parameters in the most challenging case considered in [4]. The horizontal and vertical noises were selected to be the same as the process noise magnitudes in the original model, $\sigma_h = \sqrt{0.2}$ and $\sigma_v = \sqrt{0.2}$, and the tangential noise was significantly higher $\sigma_t = \sqrt{100}$. The noise in the turn rate was the same as in the original, $\sigma_\omega = 7 \times 10^{-3}$. The data was simulated using 1000 steps of the Euler–Maruyama method between each measurement.

The following methods were tested:

- EKF/ERTS: Linearization based filter (EKF) and smoothers of Type I (ERTS1), Type II (ERTS2) and Type III (ERTS3).

- CKF/CRTS: 3rd order spherical cubature integration based filter (CKF) and smoothers (CRTS1, CRTS2, CRTS3).

- UKF/URTS: Unscented transform based filter (UKF) and smoothers of each type (URTS1, URTS2, URTS3). The UT parameters were selected to be $\alpha = 1$, $\beta = 2$ and $\kappa = 0$.

- GHKF/GHRTS: 3rd order Gauss–Hermite integration based filter (GHKF) and smoothers (GHRTS1, GHRTS2, GHRTS3).

The time integrations were done using the 4th order Runge-Kutta (RK4) method with 100 integration steps between the measurements. We used the initialization presented in the Section 5.5.2 of [3], that is, we drew the initial estimates from the prior distribution. In the estimation methods, the units of position and velocity were the plain meters and meters per second, respectively, but in turn rate we used units of 1/100 degrees per second to have more uniform scaling of the state variables.

The root mean square errors (RMSE) over 100 independent Monte Carlo runs, where both the trajectories and measurements were randomly drawn, are listed in Table 2. As can be seen in the results, the sigma-point methods quite systematically outperform the linearization based methods in the filter case as well as in all types of smoothers. In the filters the error difference is quite small, but in smoothers the difference is considerable. The performance of the different sigma-point methods is practically the same though the high number of sigma-points in Gauss–Hermite does indeed give some benefit and its errors are lowest of all the methods—except in the case of GHRTS1 smoother which has surprisingly high errors. The Type II and Type III smoothers tend to give better results than the Type I smoother in position errors, but in velocities and turn rates the Type I works better in CRTS and URTS cases. The performance of the smoother Types II and III is quite similar with all the methods, which was expected, because the underlying approximations in Types II and III are essentially the same.

In the 100 simulations, EKF diverged 9 times which caused 9 of the Type II and Type III smoother runs also to diverge. However, ERTS1 diverged a total of 34 times, including the 9 divergences of EKF. The sigma-point filter runs (CKF/UKF/GHKF), as well as their Type II and Type III smoother runs were all successful and there were no problems with divergence. However, with Type I smoothers there were numerical problems: CRTS1 and URTS1 diverged 37 times and GHRTS1 diverged 7 times. The latter also explains the poor RMSE performance of GHRTS1: the method indeed did not diverge in some of the runs where CRTS1 and URTS1 did, but it almost diverged and thus produced a result with quite high error. The reason to the divergences seems to be that the Type I smoother equations are numerically more sensitive than the other ones, because the expectations are computed over the smoothed distributions. In this simulation, the covariances tend to become quite ill-conditioned which is likely to be the reason to the divergences.

The time evolution in errors in filters and smoothers is illustrated in Figures 1 and 2, respectively. The errors are the actual mean squared errors (MSE) of three-dimensional positions over the 100 Monte Carlo runs at each time point. As here the purpose is to compare the choice of the Gaussian integration method, we have only included the filters and Type III smoothers to the analysis. The larger errors of the Taylor series based EKF and ERTS can be quite clearly seen also in both the figures. The errors of CKF and UKF as well as of the corresponding smoothers CRTS and URTS can be seen to be very similar. The error of GHKF is lowest of all the filters most of the time, but on some time intervals the error is bigger than the error of CKF and UKF. In smoothers the error of GHRTS remains the lowest almost all the time except for the very end of the trajectory.

To test the consistency of the filters and smoothers we used the two sided chi-square test [3], for the normalized estimation error squared (NEES) averaged over

Table 2: RMSE errors averaged over 100 Monte Carlo runs.

Method	Position	Velocity	Turn Rate
EKF	49.4	13.4	0.092
ERTS1	44.2	15.0	0.048
ERTS2	43.1	10.4	0.067
ERTS3	43.1	10.5	0.067
CKF	46.2	11.7	0.069
CRTS1	28.0	7.1	0.032
CRTS2	25.9	7.4	0.046
CRTS3	25.7	7.4	0.046
UKF	47.0	11.7	0.069
URTS1	28.2	7.1	0.032
URTS2	26.6	7.3	0.046
URTS3	26.4	7.3	0.046
GHKF	41.7	11.1	0.066
GHRTS1	44.4	8.3	0.045
GHRTS2	22.6	6.9	0.045
GHRTS3	22.3	6.9	0.045

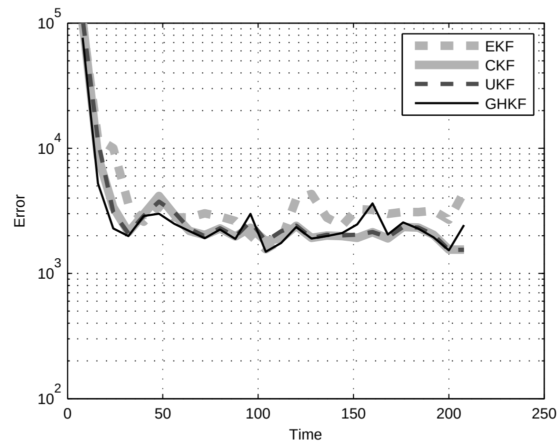

Figure 1: Time evolution of mean squared errors in filters.

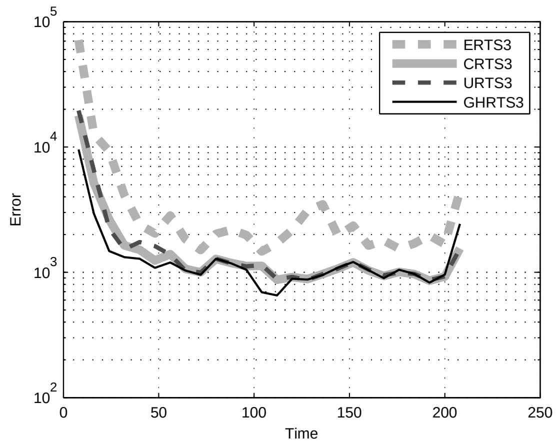

Figure 2: Time evolution of mean squared errors in Type III smoothers.

Table 3: Normalized estimation error squared (NEES) averaged over 100 Monte Carlo runs and measurements. Only GH based methods can be seen to be on the 95% confidence interval $[6.9, 7.1]$ implying consistency.

Method	Average NEES
EKF	1.1×10^6
ERTS3	1.9×10^6
CKF	12.5
CRTS3	12.3
UKF	11.6
URTS3	11.6
GHKF	7.0
GHRTS3	7.1

Monte Carlo samples and time. Under hypothesis that the filter is consistent, with probability 95% the averaged empirical NEES should be on the range $[6.9, 7.1]$. According to the test results, which are shown in Table 3, only the Gauss–Hermite based methods GHKF and GHRTS are indeed consistent. The test statistics of EKF and ERTS are not even close to the 95% confidence interval, and although those of CKF, CRTS, UKF and URTS are much closer, the methods are not consistent according to the statistical test.

5. Discussion and Extensions

5.1. Which Method to Choose?

The theoretical analysis and simulation suggests that for a practical tracking problem with moderately high state dimension, the method of choice would be the Type III smoother with the cubature or unscented transform based Gaussian integration method (i.e., CRTS3 or URTS3). The Type III smoother has the lowest computational cost and also gave very low errors in the simulation. The cubature and unscented methods have the advantage of being computationally quite light, but still their error properties are very good. However, they have the problem that their error estimates might not always be consistent with the actual errors. The unscented transform has more parameters to tune for a particular tracking problem, which can be an advantage or a disadvantage, and in fact, cubature method can be seen as a special case of unscented transform with suitable selection of parameters ($\alpha = 1, \beta = 0, \kappa = 0$). Although, the error of Gauss–Hermite integration is lower than of the other integration methods, and its consistency properties are very good, its computational complexity is too high for many practical tracking problems. In lower dimensional tracking problems the Gauss–Hermite might also be the method of choice or in the cases that the estimator consistency is a very important factor.

5.2. Square-Root and Bryson–Frazier–Bierman Forms

The sigma-point based approximations to the filter and smoother differential equations in Sections 2.2 and 3.5 could also be converted into square-root form using the same procedure as was used in [31, 12] for deriving the continuous square root unscented estimators. Of course, on the update step of the filter we can use the square-root formulations of discrete-time filters [43, 44, 33]. It might also be possible to avoid the inversion of the prediction covariance in the smoother by defining a new variable, $\omega(t) = P^{-1}\,[m^s - m]$, and by converting the smoothing problem into a form compatible with the variational formulation of Bryson and Frazier (see [21, 19, 22, 43]).

5.3. Continuous-Time Models

Continuous-time filtering and smoothing are considered with state space models, where the measurement process is continuous in time also (see, e.g., [1]):

$$\begin{aligned} dx &= f(x,t)\,dt + L(x,t)\,d\beta \\ dz &= h_c(x,t)\,dt + d\nu, \end{aligned} \tag{52}$$

where $\nu(t)$ is a Brownian motion with diffusion matrix $R_c(t)$. If we use the same limiting procedure as was used in [31], we get the following *continuous-time Gaussian filtering equations*:

$$\begin{aligned} K_c(t) &= \mathrm{E}[(x-m)\,h_c^T(x,t)]\,R_c^{-1}(t) \\ dm &= \mathrm{E}[f(x,t)]\,dt + K_c(t)\,(dz - E[h_c(x,t)]\,dt) \\ \frac{dP}{dt} &= \mathrm{E}[(x-m)\,f^T(x,t)] + \mathrm{E}[f(x,t)\,(x-m)^T] \\ &\quad + \mathrm{E}[\Sigma(x,t)] - K_c(t)\,R_c(t)\,K_c^T(t), \end{aligned} \tag{53}$$

which could further be simplified with Equation (8) and approximated, for example, with the methods presented in Section 2.2.

As the non-linear smoothing theory in [11] is applicable both to continuous-discrete and continuous problems, the Type I smoothing equations are also valid in pure continuous-time problems. However, whether the Type II equations are valid also for continuous-time processes is not as clear, because we had to resort to use of ordinary calculus in their derivation. But it can be argued that the equations should be valid in Stratonovich sense [12]. However, if the used filter approximation is such that the covariance approximation is differentiable (as, e.g., in the approximation above), except possibly at finite number of points, then the Type II equations should be valid for continuous-time smoothing as well. This conclusion is also backed up by the fact that the previously presented continuous-time smoothing equations can indeed be considered as approximations to the Type II smoothers (see, Section 3.5).

6. Conclusion

In this paper we have first shown how new numerical integration and sigma-point based methodology can be applied to the classical continuous-discrete Gaussian filtering framework. We have also derived two novel Gaussian smoothers which consist of backward differential equations for the smoother mean and covariance, and shown, how the latter smoother can be converted into a computationally more efficient form. We have also shown how the new numerical methodology can be used for approximating the Gaussian integrals occurring in the smoother equations. The performance of the different approximations was tested in a simulated application.

References

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

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

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

[4] I. Arasaratnam, S. Haykin, T. R. Hurd, Cubature Kalman filtering for continuous-discrete systems: Theory and simulations, IEEE Transactions on Signal Processing 58 (10) (2010) 4977–4993.

[5] S. Särkkä, A. Solin, On continuous-discrete cubature Kalman filtering, in: Proceedings of SYSID 2012, 2012, to appear.

[6] K. Ito, K. Xiong, Gaussian filters for nonlinear filtering problems, IEEE Transactions on Automatic Control 45 (5) (2000) 910–927.

[7] Y. Wu, D. Hu, M. Wu, X. Hu, A numerical-integration perspective on Gaussian filters, IEEE Transactions on Signal Processing 54 (8) (2006) 2910–2921.

[8] S. Särkkä, J. Hartikainen, On Gaussian optimal smoothing of non-linear state space models, IEEE Transactions on Automatic Control 55 (8) (2010) 1938–1941.

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

[10] C. T. Striebel, Partial differential equations for the conditional distribution of a Markov process given noisy observations, Journal of Mathematical Analysis and Applications 11 (1965) 151–159.

[11] C. T. Leondes, J. B. Peller, E. B. Stear, Nonlinear smoothing theory, IEEE Transactions on Systems Science and Cybernetics 6 (1) (1970) 63–71.

[12] S. Särkkä, Continuous-time and continuous-discrete-time unscented Rauch-Tung-Striebel smoothers, Signal Processing 90 (1) (2010) 225–235.

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

[14] A. H. Jazwinski, Filtering for nonlinear dynamical systems, IEEE Transactions on Automatic Control 11 (4) (1966) 765–766.

[15] Y. C. Ho, R. C. K. Lee, A Bayesian approach to problems in stochastic estimation and control, IEEE Transactions on Automatic Control 9 (4) (1964) 333–339.

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

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

[18] H. E. Rauch, Solutions to the linear smoothing problem, IEEE Transactions on Automatic Control 8 (4) (1963) 371–372.

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

[20] F. E. Daum, Exact finite-dimensional nonlinear filters, IEEE Transactions on Automatic Control 31 (7) (1986) 616–622.

[21] A. E. Bryson, M. Frazier, Smoothing for linear and nonlinear dynamic systems, in: Proceedings of the Optimum System Synthesis Conference, 1963, pp. 353–364.

[22] A. E. Bryson, Y.-C. Ho, Applied optimal control : optimization, estimation, and control, Waltham (MA), 1969.

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

[24] 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.

[25] E. A. Wan, R. van der Merwe, The unscented Kalman filter, in: S. Haykin (Ed.), Kalman Filtering and Neural Networks, Wiley, 2001, Ch. 7, pp. 221–280.

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

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

[28] H. J. Kushner, Approximations to optimal nonlinear filters, IEEE Transactions on Automatic Control 12 (5) (1967) 546–556.

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

[30] S. Särkkä, Recursive Bayesian Inference on Stochastic Differential Equations, Doctoral dissertation, Helsinki University of Technology (2006).

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

[32] H. Singer, Nonlinear continuous time modeling approaches in panel research, Statistica Neerlandica 62 (1) (2008) 29–57.

[33] I. Arasaratnam, S. Haykin, Cubature Kalman filters, IEEE Transactions on Automatic Control 54 (6) (2009) 1254–1269.

[34] G. O. Roberts, O. Stramer, On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm, Biometrica 88(3) (2001) 603–612.

[35] O. Elerian, S. Chib, N. Shephard, Likelihood inference for discretely observed nonlinear diffusions, Econometrica 69(4) (2001) 959–993.

[36] B. Eraker, MCMC analysis of diffusion models with applications to finance, Journal of Business and Economic Statistics 19(2) (2001) 177–191.

[37] S. Särkkä, T. Sottinen, Application of Girsanov theorem to particle filtering of discretely observed continuous-time non-linear systems, Bayesian Analysis 3(3) (3) (2008) 555–584.

[38] L. Murray, A. Storkey, Particle smoothing in continuous time: A fast approach via density estimation, IEEE Transactions on Signal Processing 59 (3) (2011) 1017–1026.

[39] J. Gunther, R. Beard, J. Wilson, T. Oliphant, W. Stirling, Fast nonlinear filtering via Galerkin's method, in: American Control Conference. American Automatic Control Council., 1997.

[40] S. Challa, Y. Bar-Shalom, V. Krishnamurthy, Nonlinear filtering via generalized Edgeworth series and Gauss-Hermite quadrature, IEEE Transactions on Signal Processing 48 (6) (2000) 1816–1820.

[41] A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, 1984.

[42] A. P. Sage, Maximum a posteriori filtering and smoothing algorithms, Int. J. Control 11 (4) (1970) 641–658.

[43] G. J. Bierman, Factorization Methods for Discrete Sequential Estimation, Academic Press, 1977.

[44] R. van der Merwe, E. A. Wan, The square-root unscented Kalman filter for state and parameter-estimation, in: International Conference on Acoustics, Speech, and Signal Processing, 2001, pp. 3461–3464.