Non-Linear Continuous-Discrete Smoothing By Basis Function Expansions Of Brownian Motion

Filip Tronarp and Simo Särkkä
Department of Electrical Engineering and Automation
Aalto University, Finland
E-Mail: { filip.tronarp, simo.sarkka }@aalto.fi

***Abstract*—This paper is concerned with inferring the state of a Itô stochastic differential equation (SDE) from noisy discrete-time measurements. The problem is approached by considering basis function expansions of Brownian motion, that as a consequence give approximations to the underlying stochastic differential equation in terms of an ordinary differential equation with random coefficients. This allows for representing the latent process at the measurement points as a discrete time system with a non-linear transformation of the previous state and a noise term. The smoothing problem can then be solved by sigma-point or Taylor series approximations of this non-linear function, implementations of which are detailed. Furthermore, a method for interpolating the smoothing solution between measurement instances is developed. The developed methods are compared to the Type III smoother in simulation examples involving (i) hyperbolic tangent drift and (ii) the Lorenz 63 system where the present method is found to be better at reconstructing the smoothing solution at the measurement points, while the interpolation scheme between measurement instances appear to suffer from edge effects, serving as an invitation to future research.**

***Index Terms*—Non-linear continuous-discrete smoothing, stochastic differential equation, basis expansion, Brownian motion.**

I. Introduction

Inference in continuous-time stochastic dynamic systems is a freqently occuring topic in disciplines such as navigation, tracking, and time series modelling [1], [2], [3], [4], [5]. The system is typically described in terms of a latent Markov process $\{X(s)\}_{s\geq 0}$, governed by a stochastic differential equation (SDE) [6]. Furthermore, the process $\{X(s)\}_{s\geq 0}$ is assumed to be measured at a set of time instants $\{t_k\}_{k=1}^K$ by a collection of random variables $\{Y(t_k)\}_{k=1}^K$, each having a conditional distribution with respect to the process outcome at the corresponding time stamp. For the special case of an affine Gaussian system, the calculation of predictive, filtering, and smoothing distributions amount to computing the first two joint moments of latent process and the measurement process. In the case of filtering this procedure is known as Kalman-Bucy filtering [7]. Subsequently, Rauch, Tung, and Striebel showed that the smoothing moments can be expressed in terms of ordinary differential equations with the filter moments as inputs [8].

Although the theory of filtering and smoothing in linear systems is well established, the non-linear case is still an area of intense research. An early strategy, for filtering, was to linearise the system around the mean trajectory using truncated Taylor series expansions which enables the use of the aforementioned methods for affine systems [9]. This Taylor series approach was later extended to smoothing by combining it with the development of non-linear smoothing theory [10]. Later on, sigma-point approaches to continuous-discrete smoothing emerged based on Euler discretisations and taking the limit of the discrete time smoother [11]. Additionally, in [12] a series of sigma-point smoothers were derived based on the non-linear smoothing theory of [10]. A different line of research is to approximate the family of smoothing distributions with a Gaussian process via variational Bayes [13], [14].

On the other hand, Brownian motion can be approximated using a basis function expansion [15], which results in approximating the underlying stochastic differential equation with an ordinary differential equation with random coefficients [16]. Based on this, another type of sigma-point filter for continuous-discrete systems was developed in [17].

In this paper, non-linear continuous-discrete smoothers are derived based on Wong-Zakai expansions of Brownian motion [15]. This allows the latent stochastic process to be represented as a discrete time system with non-linear transformations of the state and noise terms. Based on this smoothers based on both sigma point and Taylor series approximations are developed, thus generalising the methodology in [17] to smoothing problems. Furthermore, a scheme for interpolating the smoothing distribution between measurement instances is developed. The developed methods are validated in simulation experiments.

The rest of this paper is organised as follows. This section is concluded by establishing notation and formalising the problem formulation. Basis function expansions of Brownian motion and the corresponding ordinary differential equation approximations to stochastic differential equations are reviewed in Section II. In Section III the Gaussian smoother based on basis function expansions of Brownian motion is formally derived, and sigma point and Taylor series implementations are subsequently developed as well as a method for interpolating the smoothing distribution between measurement instances. Simulation results for the derived smoothers are compared to the Type III smoother (see [12]) in Section IV, and finally conclusions are given in Section V.

A. Notation

For the stochastic process $\{Y(t_k)\}_{k=1}^K$ the following sets are defined

$$\mathscr{Y}(t) = \{y(t_k) \mid t_k \le t\}, \tag{1a}$$
$$\mathscr{Y}(t^-) = \{y(t_k) \mid t_k < t\}, \tag{1b}$$

The density of the process $X(t)$ conditioned on the set of outcomes of $Y(t_k)$ up to time τ is then denoted by $f(x, t \mid \mathscr{Y}(\tau))$ and $f(x, t \mid \mathscr{Y}(\tau^-))$ for the outcomes up to just before τ. Conditional expectations, covariances, and cross-covariances are similarly denoted by $\mathbb{E}[\cdot \mid \mathscr{Y}(\tau)]$, $\mathbb{V}[\cdot \mid \mathscr{Y}(\tau)]$, and $\mathbb{C}[\cdot, \cdot \mid \mathscr{Y}(\tau)]$. As a shorthand, the filtering moments at time t are denoted by

$$\mathbb{E}_t^F[\cdot] = \mathbb{E}[\cdot \mid \mathscr{Y}(t)],$$
$$\mathbb{V}_t^F[\cdot] = \mathbb{V}[\cdot \mid \mathscr{Y}(t)],$$
$$\mathbb{C}_t^F[\cdot] = \mathbb{E}[\cdot, \cdot \mid \mathscr{Y}(t)],$$

the mean and covariance of $X(t)$ with respect to the filtering distribution are given special notation according to

$$\bar{x}(t) = \mathbb{E}_t^F[X(t)], \tag{3a}$$
$$\Sigma(t) = \mathbb{V}_t^F[X(t)]. \tag{3b}$$

Similarly, the smoothing moments are denoted by (conditioning on $\mathscr{Y}(t_K)$)

$$\mathbb{E}^S[\cdot] = \mathbb{E}[\cdot \mid \mathscr{Y}(t_K)],$$
$$\mathbb{V}^S[\cdot] = \mathbb{V}[\cdot \mid \mathscr{Y}(t_K)],$$
$$\mathbb{C}^S[\cdot] = \mathbb{E}[\cdot, \cdot \mid \mathscr{Y}(t_K)],$$

and the mean and covariance of $X(t)$ with respect to the smoothing distribution are denoted by

$$\hat{x}(t) = \mathbb{E}^S[X(t)], \tag{5a}$$
$$\Omega(t) = \mathbb{V}^S[X(t)]. \tag{5b}$$

Furthermore, for the Itô process $X(t)$, its Itô differential is denoted by $\mathrm{d}X(t)$, $\mathrm{d}t$ is a time differential, and ∂_t is a time derivative.

B. Problem Formulation

In this paper, systems of the following form are considered

$$\mathrm{d}X(t) = \mu(X(t))\,\mathrm{d}t + \sigma(t)\,\mathrm{d}B(t), \tag{6a}$$
$$Y(t_k) = h(X(t_k)) + V(t_k), \tag{6b}$$

where $f(x, 0) = \mathcal{N}(x; \bar{x}(0^-), \Sigma(0^-))$, $\mu\colon \mathbb{R}^{d_X} \to \mathbb{R}^{d_X}$ is the drift function, $\sigma\colon \mathbb{R}_+ \to \mathbb{R}^{d_X \times d_B}$ is the diffusion matrix, $B(t)$ is a vector of standard Brownian motions, $h\colon \mathbb{R}^{d_X} \to \mathbb{R}^{d_Y}$ is the measurement function, and $V(t_k)$ is a Gaussian white noise sequence with $\mathbb{C}[V(t_k), V(t_l)] = \delta_{kl}\Delta$. Given a measurement sequence, $\{y(t_k)\}_{k=1}^K$, the objective of inference, in the Bayesian sense, is then to find the family of conditional densities

$$f(x, \tau \mid \mathscr{Y}(t_k)),\ k = 1, \ldots, K. \tag{7}$$

When $\tau = t_k$ the probability density in Equation (7) is said to be a filtering distribution, when $\tau > t_k$ it is a predictive distribution, and when $\tau < t_k$ it is a smoothing distribution.

II. Approximation of Stochastic Differential Equations Using Basis Expansion Of Brownian Motion

In this section, the approach based on basis function expansions of Brownian motion to approximating stochastic differential equations is reviewed. In Section II-A basis expansions of Brownian motion are reviewed and in Section II-B the approach to approximating the stochastic differential equation taken in this paper is presented.

A. Basis Function Expansions of Brownian Processes

A standard Brownian motion $B(t)$ can be expanded in terms of a set of orthonormal basis functions, $\{\phi_l(t)\}_{l=1}^\infty$, on the interval $[0, T]$. Using the standard inner product on $L^2([0, T], \mathbb{R})$,

$$\langle \phi, \phi' \rangle = \int_0^T \phi(\tau)\phi'(\tau)\,\mathrm{d}\tau, \tag{8}$$

gives the following basis expansion of $B(t)$ [16]

$$B(t) = \sum_{l=1}^\infty W_l \int_0^t \phi_l(\tau)\,\mathrm{d}\tau,\ t \in [0, T], \tag{9}$$

where the coefficients, $\{W_l\}_{l=1}^\infty$ is a standard Gaussian white noise process. With this representation the increment, $\mathrm{d}B(t)$, can be written as

$$\mathrm{d}B(t) = \sum_{l=1}^\infty W_l \phi_l(t)\,\mathrm{d}t. \tag{10}$$

For instance, the Karhunen-Loeve expansion would correspond to

$$\phi_l(t) = \left(\frac{2}{\delta}\right)^{1/2} \cos\left(\frac{(2l-1)\pi}{2\delta} t\right). \tag{11a}$$

$$\tag{11b}$$

Though other orthonormal bases are also possible such as Haar functions, corresponding to Lévy-Ciesielski constructions of Brownian motion [18].

B. An Approximate Stochastic System

The stochastic differential equation in Equation (6a) can be approximated by an ordinary differential equation by using a basis expansion of Brownian motion [15], [16], [17]. However, making a basis expansion for the entire interval, $[t_1, t_K]$ would require a prohibitive amount of basis functions (large L), hence an interval wise basis expansion is more suitable. That is, $B(t)$ is expanded at each measurement interval, $\{[t_k, t_{k+1}]\}_{k=0}^{K-1}$. Now define

$$\phi^\mathsf{T}(t) = \begin{bmatrix} \phi_1(t) & \ldots & \phi_L(t) \end{bmatrix}, \tag{12}$$

each coordinate of $B(t)$ can then be approximated on the interval $[t_{k+1}, t_k]$ as

$$B_i(t) = \sum_{l=1}^L W_{il}(t_{k+1}) \int_{t_k}^t \phi_l(\tau)\,\mathrm{d}\tau,\ t \in [t_k, t_{k+1}], \tag{13}$$

or in matrix form

$$B(t) = W(t_{k+1}) \int_{t_k}^{t} \phi(\tau)\,\mathrm{d}\tau, \ t \in [t_k, t_{k+1}]. \tag{14}$$

From this it follows that $X(t)$ is given by

$$\begin{aligned} X(t) &\approx X(t_0) + \int_{t_0}^{t} \mu(X(\tau))\,\mathrm{d}\tau \\ &+ \int_{t_0}^{t} \sum_{k=0}^{K-1} \chi_{[t_k, t_{k+1}]}(\tau)\sigma(\tau)W(t_{k+1})\phi(\tau)\,\mathrm{d}\tau, \end{aligned}$$

where χ is an indicator function. The solution $X(t)$ can be obtained by integrating the following ODE on the interval $[t_0, t]$

$$\partial_t X(t) \approx \mu(X(t)) + \sum_{k=0}^{K-1} \chi_{[t_k, t_{k+1}]}(t)\sigma(t)W(t_{k+1})\phi(t).$$

The solution at the measurement points $\{t_k\}_{k=1}^{K}$ can be written in terms of the following recursion

$$X(t_{k+1}) = F[X(t_k), W(t_{k+1})], \tag{15}$$

hence the continuous-discrete smoothing problem can be converted into a discrete-discrete smoothing problem.

III. Continuous-Discrete Smoothing Using basis Function Expansion

In this section, the Gaussian smoothing approach based on basis function expansions of Brownian motion is presented. In Section III-A the equations needed to implement both the filter and the smoother are derived. Sigma-point and Taylor series based approximations to the smoothing equations are then derived in Section III-B and Section III-C, respectively.

A. Derivation

Suppose the filtering distribution at time t_k is given by

$$f(x, t_k \mid \mathscr{Y}(t_k)) = \mathcal{N}(x; \bar{x}(t_k), \Sigma(t_k)). \tag{16}$$

Now, in order to find a Gaussian approximation to the predictive distribution,

$$f(x, t_{k+1} \mid \mathscr{Y}(t_k)) \approx \mathcal{N}(x; \bar{x}(t_{k+1}^-), \Sigma(t_{k+1}^-)), \tag{17}$$

then the moments need to be approximated. The mean is given by

$$\begin{aligned} \bar{x}(t_{k+1}^-) &= \mathbb{E}_{t_k}^F[F[X(t_k), W(t_{k+1})]] \\ &= \bar{x}(t_k) + \int_{t_k}^{t_{k+1}^-} \mathbb{E}_{t_k}^F[\mu(X(\tau))]\,\mathrm{d}\tau, \end{aligned} \tag{18}$$

and the covariance is given by

$$\begin{aligned} &\Sigma(t_{k+1}^-) = \mathbb{V}_{t_k}^F[F[X(t_k), W(t_{k+1})]] \\ &= \Sigma(t_k) + \mathbb{C}_{t_k}^F\left[X(t_k), \int_{t_k}^{t_{k+1}^-} \mu(X(\tau))\,\mathrm{d}\tau\right] \\ &+ \mathbb{C}_{t_k}^F\left[X(t_k), \int_{t_k}^{t_{k+1}^-} \sigma(\tau)W(t_{k+1})\phi(\tau)\,\mathrm{d}\tau\right] \\ &+ \mathbb{C}_{t_k}^F\left[X(t_k), \int_{t_k}^{t_{k+1}^-} \mu(X(\tau))\,\mathrm{d}\tau\right]^{\mathsf{T}} \\ &+ \mathbb{V}_{t_k}^F\left[\int_{t_k}^{t_{k+1}^-} \mu(X(\tau))\,\mathrm{d}\tau\right] \\ &+ \mathbb{C}_{t_k}^F\left[\int_{t_k}^{t_{k+1}^-} \sigma(\tau)W(t_{k+1})\phi(\tau)\,\mathrm{d}\tau, \int_{t_k}^{t_{k+1}^-} \mu(X(\tau))\,\mathrm{d}\tau\right]^{\mathsf{T}} \\ &+ \mathbb{C}_{t_k}^F\left[X(t_k), \int_{t_k}^{t_{k+1}^-} \sigma(\tau)W(t_{k+1})\phi(\tau)\,\mathrm{d}\tau\right]^{\mathsf{T}} \\ &+ \mathbb{C}_{t_k}^F\left[\int_{t_k}^{t_{k+1}^-} \sigma(\tau)W(t_{k+1})\phi(\tau)\,\mathrm{d}\tau, \int_{t_k}^{t_{k+1}^-} \mu(X(\tau))\,\mathrm{d}\tau\right] \\ &+ \mathbb{V}_{t_k}^F\left[\int_{t_k}^{t_{k+1}^-} \sigma(\tau)W(t_{k+1})\phi(\tau)\,\mathrm{d}\tau\right]. \end{aligned}$$

Additionally, for the smoother gain, the cross-covariance between $X(t_k)$ and $X(t_{k+1}^-)$ needs to be computed; it is given by

$$\begin{aligned} &\mathbb{C}_{t_k}^F[X(t_k), X(t_{k+1}^-)] = \Sigma(t_k) \\ &+ \mathbb{C}_{t_k}^F\left[X(t_k), \int_{t_k}^{t_{k+1}^-} \mu(X(\tau))\,\mathrm{d}\tau\right] \\ &+ \mathbb{C}_{t_k}^F\left[X(t_k), \int_{t_k}^{t_{k+1}^-} \sigma(\tau)W(t_{k+1})\phi(\tau)\,\mathrm{d}\tau\right]. \end{aligned} \tag{19}$$

The gain needed for the smoother implementation (cf. [5]) is then given by

$$G(t_k) = \mathbb{C}_{t_k}^F[X(t_k), X(t_{k+1}^-)]\Sigma^{-1}(t_{k+1}^-). \tag{20}$$

Subsequently, approximation strategies for implementing the aforementioned moment computations shall be devised.

B. Sigma-Point Implementation

In order to develop a sigma-point implementation of the smoother, the strategy for developing the corresponding filter is employed [17]. Let vec be the vectorisation operator, then define

$$Z(t_k) = \begin{bmatrix} X(t_k) \\ \operatorname{vec} W(t_k) \end{bmatrix}. \tag{21}$$

Now $Z(t_k)$ is Gaussian distributed according to

$$Z(t_k) \sim \mathcal{N}\left(\begin{bmatrix} \bar{x}(t_k) \\ 0 \end{bmatrix}, \operatorname{blkdiag}[\Sigma(t_k), \mathrm{I}]\right), \tag{22}$$

where blkdiag is the block diagonal matrix containing its arguments on the block diagonals. Now let $\{\mathcal{Z}_j(t_k)\}_{j=1}^{J_Z}$ and

$\{\lambda_j\}_{j=1}^{J_Z}$ be sigma-points and weights associated with $Z(t_k)$, respectively. Furthermore, let $\mathcal{X}_j(t_k)$ and $\mathcal{W}_j(t_k)$ be the sub-vectors of $\mathcal{Z}_j(t_k)$ corresponding to $X(t_k)$ and $\operatorname{vec} W(t_k)$, respectively. The sigma-point, $\mathcal{X}_j(t_{k+1}^-)$, is then given by

$$\begin{aligned}\mathcal{X}_j(t_{k+1}^-) &= F[\mathcal{X}_j(t_k), \mathcal{W}_j(t_k)] \\ &= \mathcal{X}_j(t_k) + \int_{t_k}^{t_{k+1}^-} \mu(\mathcal{X}_j(\tau))\,\mathrm{d}\tau \\ &\quad + \int_{t_k}^{t_{k+1}^-} \phi^\mathsf{T}(\tau) \otimes \sigma(\tau)\mathcal{W}_j(t_k)\,\mathrm{d}\tau,\end{aligned} \tag{23}$$

where $\otimes$ is Kronecker's product and the identity

$$\sigma(\tau)W(t_k)\phi(\tau) = \phi^\mathsf{T}(\tau) \otimes \sigma(\tau)\operatorname{vec} W(t_k), \tag{24}$$

was used. Therefore, $\mathcal{X}_j(t_{k+1}^-)$ can be found by solving the following ordinary differential equation [17]

$$\partial_t \mathcal{X}_j(t) = \mu(\mathcal{X}_j(t)) + \phi^\mathsf{T}(t) \otimes \sigma(t)\mathcal{W}_j(t_k), \tag{25}$$

on the interval $[t_k, t_{k+1}^-]$ with initial condition $\mathcal{X}_j(t_k)$. The quantities necessary for the smoother can then be approximated by

$$\bar{x}(t_{k+1}^-) \approx \sum_{j=1}^{J_Z} \lambda_j \mathcal{X}_j(t_{k+1}^-) \tag{26a}$$

$$\begin{aligned}\Sigma(t_{k+1}^-) \approx \sum_{j=1}^{J_Z} \lambda_j \Big(\mathcal{X}_j(t_{k+1}^-) - \bar{x}(t_{k+1}^-)\Big) \\ \times \Big(\mathcal{X}_j(t_{k+1}^-) - \bar{x}(t_{k+1}^-)\Big)^\mathsf{T}\end{aligned} \tag{26b}$$

$$\begin{aligned}G(t_k) \approx \sum_{j=1}^{J_Z} \lambda_j \Big(\mathcal{X}_j(t_k) - \bar{x}(t_k)\Big) \\ \times \mathcal{X}_j^\mathsf{T}(t_{k+1}^-)\Sigma^{-1}(t_{k+1}^-).\end{aligned} \tag{26c}$$

The smoothing recursion is then given by [5]

$$\hat{x}(t_k) = \bar{x}(t_k) + G(t_k)\Big(\hat{x}(t_{k+1}) - \bar{x}(t_{k+1}^-)\Big) \tag{27a}$$

$$\Omega(t_k) = \Sigma(t_k) + G(t_k)\Big(\Omega(t_{k+1}) - \Sigma(t_{k+1}^-)\Big)G^\mathsf{T}(t_k). \tag{27b}$$

C. Taylor Series Implementation

In this section we derive a Taylor series based approximation to the smoother. First note that

$$\begin{aligned}\mathbb{E}_t^F[F[X(t_k), W(t_{k+1})]] &= \bar{x}(t_k) + \int_{t_k}^{t_{k+1}} \mathbb{E}_\tau^F[\mu(X(\tau))]\,\mathrm{d}\tau \\ &\approx \bar{x}(t_k) + \int_{t_k}^{t_{k+1}} \mu(\bar{x}(\tau))\,\mathrm{d}\tau,\end{aligned} \tag{31}$$

hence the mean solves the same differential equation as ordinary Gaussian filters [9],

$$\partial_t \bar{x}(t) = \mu(\bar{x}(t)). \tag{32}$$

Algorithm 1 Continuous-Discrete Series Expansion based Sigma Point Smoother

Input: Initial parameters $\bar{x}(t_0)$, $\Sigma(t_0)$, sampling intervals $\{\delta t_k\}_{k=0}^{K-1}$, and vector of basis functions $\phi^\mathsf{T}(t) = \big[\phi_1(t) \quad \dots \quad \phi_L(t)\big]$.

Output: Smoothing parameters $\{\hat{x}(t_k)\}_{k=0}^{K}$ and $\{\Omega(t_k)\}_{k=0}^{K}$

{Filter step}

for $k = 0$ **to** $K-1$

{Predict}

Form sigma points, $\{\mathcal{Z}_j(t_k)\}_{j=1}^{J_Z}$, and weights, $\{\lambda_j\}_{j=1}^{J_Z}$ for the Gaussian vector

$$Z^\mathsf{T}(t_k) \leftarrow \Big[X^\mathsf{T}(t_k) \quad (\operatorname{vec} W(t_{k+1}^-))^\mathsf{T}\Big]$$

for $j = 1$ **to** J_Z

Solve the differential equation

$$\partial_t \mathcal{X}_j(t) = \mu(\mathcal{X}_j(t)) + \phi^\mathsf{T}(t) \otimes \sigma(t)\mathcal{W}_j(t_k)$$

on the interval $[t_k, t_{k+1}^-]$ with initial condition $\mathcal{X}_j(t_k)$.

end

Compute the predictive moments and smoother gain

$$\bar{x}(t_{k+1}^-) \leftarrow \sum_{j=1}^{J_Z} \lambda_j \mathcal{X}_j(t_{k+1}^-)$$

$$\begin{aligned}\Sigma(t_{k+1}^-) \leftarrow \sum_{j=1}^{J_Z} \lambda_j \Big(\mathcal{X}_j(t_{k+1}^-) - \bar{x}(t_{k+1}^-)\Big) \\ \times \Big(\mathcal{X}_j(t_{k+1}^-) - \bar{x}(t_{k+1}^-)\Big)^\mathsf{T}\end{aligned}$$

$$G(t_k) \leftarrow \sum_{j=1}^{J_Z} \lambda_j \Big(\mathcal{X}_j(t_k) - \bar{x}(t_k)\Big)\mathcal{X}_j^\mathsf{T}(t_{k+1}^-)\Sigma^{-1}(t_{k+1}^-).$$

{Update}

Form sigma points $\{\mathcal{X}_j(t_{k+1}^-)\}_{j=1}^{J_X}$ and weights $\{\lambda_j\}_{j=1}^{J_X}$ for $X(t_{k+1}^-)$ and compute

$$\bar{y}(t_{k+1}^-) \leftarrow \sum_{j=1}^{J_X} \lambda_j h(\mathcal{X}j(t_{k+1}^-))$$

$$\begin{aligned}S(t_{k+1}^-) \leftarrow \sum_{j=1}^{J_X} \lambda_j (h(\mathcal{X}j(t_{k+1}^-)) - \bar{y}(t_{k+1}^-)) \\ \times (h(\mathcal{X}j(t_{k+1}^-)) - \bar{y}(t_{k+1}^-))^\mathsf{T} + \Delta\end{aligned}$$

$$\begin{aligned}K(t_{k+1}^-) \leftarrow \sum_{j=1}^{J_X} \lambda_j (\mathcal{X}j(t_{k+1}^-) - \bar{x}(t_{k+1}^-)) \\ \times h^\mathsf{T}(\mathcal{X}j(t_{k+1}^-))S^{-1}(t_{k+1}^-)\end{aligned}$$

$$\bar{x}(t_{k+1}) \leftarrow \bar{x}(t_{k+1}^-) + K(t_{k+1}^-)(y(t_{k+1}) - \bar{y}(t_{k+1}^-))$$

$$\Sigma(t_{k+1}) \leftarrow \Sigma(t_{k+1}^-) - K(t_{k+1}^-)S(t_{k+1}^-)K^\mathsf{T}(t_{k+1}^-).$$

end

{Smoother step}

for $k = K-1$ **to** 0

$$\hat{x}(t_k) \leftarrow \bar{x}(t_k) + G(t_k)\Big(\hat{x}(t_{k+1}) - \bar{x}(t_{k+1}^-)\Big)$$

$$\Omega(t_k) \leftarrow \Sigma(t_k) + G(t_k)\Big(\Omega(t_{k+1}) - \Sigma(t_{k+1}^-)\Big)G^\mathsf{T}(t_k).$$

end

Furthermore, in order to compute the covariance and smoother gain, F is expanded up to first order around $\bar{x}(t)$,

$$F[X(t_k), W(t_{k+1})] \approx \frac{\partial F[\bar{x}(t_k), 0]}{\partial \bar{x}(t_k)}(X(t_k) - \bar{x}(t_k)) + \frac{\partial F[\bar{x}(t_k), 0]}{\partial \operatorname{vec} W(t_{k+1})} \operatorname{vec} W(t_{k+1}). \tag{33}$$

Using the chain-rule and assuming the operator ∂_t commutes with $\partial/(\partial X(t_k))$ and $\partial/(\partial \operatorname{vec} W(t_{k+1}))$ gives

$$\begin{aligned}\partial_t \frac{\partial X(t)}{\partial X(t_k)} &= \frac{\partial \mu(X(t))}{\partial X(t_k)} \\ &= \frac{\partial \mu(X(t))}{\partial X(t)} \frac{\partial X(t)}{\partial X(t_k)},\end{aligned} \tag{34}$$

and

$$\begin{aligned}\partial_t \frac{\partial X(t)}{\partial \operatorname{vec} W(t_{k+1})} &= \frac{\partial \mu(X(t))}{\partial \operatorname{vec} W(t_{k+1})} + \phi^\mathsf{T}(t) \otimes \sigma(t) \\ &= \frac{\partial \mu(X(t))}{\partial X(t)} \frac{\partial X(t)}{\partial \operatorname{vec} W(t_{k+1})} \\ &\quad + \phi^\mathsf{T}(t) \otimes \sigma(t).\end{aligned} \tag{35}$$

Therefore,

$$\partial F[\bar{x}(t_k), 0]/(\partial \bar{x}(t_k)) \text{ and } \partial F[\bar{x}(t_k), 0]/(\partial \operatorname{vec} W(t_{k+1}))$$

can be obtained by integrating the following ordinary differential equations on the interval $[t_k, t_{k+1}^-]$

$$\partial F_x(t) = M(\bar{x}(t)) F_x(t), \tag{36a}$$

$$\partial F_w(t) = M(\bar{x}(t)) F_w(t) + \phi^\mathsf{T}(t) \otimes \sigma(t), \tag{36b}$$

with initial conditions $F_x(t_k) = \mathrm{I}$ and $F_w(t) = 0$, and where $M(x) = \partial \mu(x)/(\partial x)$, then

$$\frac{\partial F[\bar{x}(t_k), 0]}{\partial \bar{x}(t_k)} = F_x(t_{k+1}^-), \tag{37a}$$

$$\frac{\partial F[\bar{x}(t_k), 0]}{\partial \operatorname{vec} W(t_{k+1})} = F_w(t_{k+1}^-). \tag{37b}$$

Therefore, to summarise, the predictive moments and smoother gain are given by

$$\bar{x}(t_{k+1}^-) = \bar{x}(t_k) + \int_{t_k}^{t_{k+1}-} \mu(\bar{x}(\tau)) \,\mathrm{d}\tau, \tag{38a}$$

$$\Sigma(t_{k+1}^-) = F_x(t_{k+1}^-)\Sigma(t_k)F_x^\mathsf{T}(t_{k+1}^-) + F_w(t_{k+1}^-)F_w^\mathsf{T}(t_{k+1}^-), \tag{38b}$$

$$G(t_k) = \Sigma(t_k)F_x^\mathsf{T}(t_{k+1}^-)\Sigma^{-1}(t_{k+1}^-). \tag{38c}$$

D. Reconstruction of Trajectory Between Measurements

The procedure in the previous sections only provides the smoothing solution at the discretisation points. However, we might also be interested in reconstructing the solution between the points. Recall that in the interval $[t_k, t_{k+1}]$, $X(t)$ is given by

$$X(t) = X(t_k) + \int_{t_k}^{t} \mu(X(\tau)) \,\mathrm{d}\tau + \int_{t_k}^{t} \phi^\mathsf{T}(\tau) \otimes \sigma(\tau) \,\mathrm{d}\tau \operatorname{vec} W(t_{k+1}).$$

Algorithm 2 Continuous-Discrete Series Expansion based Taylor Series Smoother

Input: Initial parameters $\bar{x}(t_0^-)$, $\Sigma(t_0^-)$, sampling intervals $\{\delta t_k\}_{k=0}^{K-1}$, and vector of basis functions $\phi^\mathsf{T}(t) = \begin{bmatrix}\phi_1(t) & \dots & \phi_L(t)\end{bmatrix}$.

Output: Smoothing parameters $\{\hat{x}(t_k)\}_{k=0}^{K}$ and $\{\Omega(t_k)\}_{k=0}^{K}$

{Filter step}

for $k = 0$ **to** $K-1$

{Predict}

Solve the differential equations

$$\begin{aligned}\partial_t \bar{x}(t) &= \mu(\bar{x}(t)), \\ \partial_t F_x(t) &= M(\bar{x}(t))F_x(t), \\ \partial_t F_w(t) &= M(\bar{x}(t))F_w(t) + \phi^\mathsf{T}(t) \otimes \sigma(t),\end{aligned}$$

on the interval $[t_k, t_{k+1}^-]$ with initial conditions $\bar{x}(t_k)$, $F_x(t_k) = \mathrm{I}$, and $F_w(t_k) = 0$.

Compute the predictive covariance and smoother gain

$$\begin{aligned}\Sigma(t_{k+1}^-) &\leftarrow F_x(t_{k+1}^-)\Sigma(t_k)F_x^\mathsf{T}(t_{k+1}^-) + F_w(t_{k+1}^-)F_w^\mathsf{T}(t_{k+1}^-) \\ G(t_k) &\leftarrow \Sigma(t_k)F_x^\mathsf{T}(t_{k+1}^-)\Sigma^{-1}(t_{k+1}^-).\end{aligned}$$

{Update}

$H_x(t_{k+1}^-) \leftarrow \partial h(\bar{x}(t_{k+1}^-))/(\partial x)$

$$\begin{aligned}\bar{y}(t_{k+1}^-) &\leftarrow h(\bar{x}(t_{k+1}^-)) \\ S(t_{k+1}^-) &\leftarrow H_x(t_{k+1}^-)\Sigma(t_{k+1}^-)H_x^\mathsf{T}(t_{k+1}^-) + \Delta \\ K(t_{k+1}^-) &\leftarrow \Sigma(t_{k+1}^-)H_x^\mathsf{T}(t_{k+1}^-)S^{-1}(t_{k+1}^-) \\ \bar{x}(t_{k+1}) &\leftarrow \bar{x}(t_{k+1}^-) + K(t_{k+1}^-)(y(t_{k+1}) - \bar{y}(t_{k+1}^-)) \\ \Sigma(t_{k+1}) &\leftarrow \Sigma(t_{k+1}^-) - K(t_{k+1}^-)S(t_{k+1}^-)K^\mathsf{T}(t_{k+1}^-).\end{aligned}$$

end

{Smoother step}

for $k = K-1$ **to** 0

$$\begin{aligned}\hat{x}(t_k) &\leftarrow \bar{x}(t_k) + G(t_k)\Big(\hat{x}(t_{k+1}) - \bar{x}(t_{k+1}^-)\Big) \\ \Omega(t_k) &\leftarrow \Sigma(t_k) + G(t_k)\Big(\Omega(t_{k+1}) - \Sigma(t_{k+1}^-)\Big)G^\mathsf{T}(t_k).\end{aligned}$$

end

Now define

$$U(t_k) = \begin{bmatrix} X(t_k) \\ \operatorname{vec} W(t_{k+1}) \end{bmatrix}. \tag{43}$$

If a Gaussian approximation to the smoothing distribution $U(t_k)$ can be obtained, then a set of sigma points $\{\mathcal{U}_j(t_k)\}_{j=1}^{J_U}$ with corresponding weights, $\{\lambda_j\}_{j=1}^{J_U}$ can be drawn. Furthermore, denote

$$\mathcal{U}_j(t_k) = \begin{bmatrix} \mathcal{X}_j(t_k) \\ \mathcal{W}_j(t_{k+1}) \end{bmatrix}, \tag{44}$$

then the sigma points of the smoothing distribution for $\mathcal{X}_j(t),\ t \in [t_k, t_{k+1}]$ can be obtained according

$$\partial_t \mathcal{X}_j(t) = \mu(\mathcal{X}_j(t)) + \phi^\mathsf{T}(t) \otimes \sigma(t)\mathcal{W}_j(t_{k+1}), \tag{45}$$

which should be solved with initial condition given by $\mathcal{X}_j(t_k)$. The smoothing mean and covariance of $X(t)$ can then be obtained by weighted sums of the solutions $\mathcal{X}_j(t)$, as usual.

For the Taylor series approach, the Jacobian of $X(t)$ with respect to $X(t_k)$ and $\operatorname{vec} W(t_{k+1})$ need to be computed. For short-hand, define

$$F_{x(t_k)}(t) = \frac{\partial X(t)}{\partial X(t_k)}, \tag{46a}$$

$$F_{w(t_{k+1})}(t) = \frac{\partial X(t)}{\partial \operatorname{vec} W(t_{k+1})}, \tag{46b}$$

and

$$F_{u(t_k)}(t) = \begin{bmatrix} F_{x(t_k)}(t) & F_{w(t_{k+1})}(t) \end{bmatrix}. \tag{47}$$

Then by the same argument as before, $F_{u(t_k)}(t)$ evaluated at the smoothing mean can be approximated by

$$\partial_t F^{\mathsf{T}}_{u(t_k)}(t) = \mathrm{I}_2 \otimes M(\bar{x}(t)) F^{\mathsf{T}}_{u(t_k)}(t) + \begin{bmatrix} 0 \\ (\phi^{\mathsf{T}}(t) \otimes \sigma(t))^{\mathsf{T}} \end{bmatrix},$$

where I_2 is a 2 by 2 identity matrix and

$$\hat{x}(t) = \mu(\hat{x}(t)) + \phi^{\mathsf{T}}(t) \otimes \sigma(t) \mathbb{E}^S[\operatorname{vec} W(t_{k+1})] \tag{48a}$$

$$\Omega(t) = F_{u(t_k)}(t) \mathbb{V}^S[U(t_k)] F^{\mathsf{T}}_{u(t_k)}(t), \; t \in [t_k, t_{k+1}]. \tag{48b}$$

For both the sigma point and Taylor series approaches, the smoothing distribution of $U(t_k)$ needs to be retrieved. The key to finding this posterior distribution is to exploit the Markovianity of the model

$$\begin{aligned} &f(w(t_{k+1}), x, t_k; x', t_{k+1} \mid \mathscr{Y}(t_K)) \\ &= f(w(t_{k+1}) \mid x, t_k; x', t_{k+1}, \mathscr{Y}(t_K)) \\ &\quad \times f(x, t_k; x', t_{k+1} \mid \mathscr{Y}(t_K)) \\ &= f(w(t_{k+1}) \mid x, t_k; x', t_{k+1}) \\ &\quad \times f(x, t_k; x', t_{k+1} \mid \mathscr{Y}(t_K)), \end{aligned} \tag{49}$$

where the second factor can be retrieved from Algorithm 1 or Algorithm 2 (see [5]). The first factor can be approximated at the filtering step. More specifically, define

$$\Xi^F(t_k) = \begin{bmatrix} \Sigma(t_k) & \mathbb{C}^F_{t_k}[X(t_k), X(t_{k+1})] \\ \mathbb{C}^F_{t_k}[X(t_k), X(t_{k+1})]^{\mathsf{T}} & \Sigma(t^-_{k+1}) \end{bmatrix} \tag{50a}$$

$$\Xi^S(t_k) = \begin{bmatrix} \Omega(t_k) & G(t_k)\Omega(t_{k+1}) \\ \Omega(t_{k+1}) G^{\mathsf{T}}(t_k) & \Omega(t_{k+1}) \end{bmatrix} \tag{50b}$$

$$\begin{aligned} &\Gamma(t_k) = \\ &\begin{bmatrix} \mathbb{C}^F_{t_k}[\operatorname{vec} W(t_{k+1}), X(t_k)] & \mathbb{C}^F_{t_{k+1}}[\operatorname{vec} W(t_{k+1}), X(t_{k+1})] \end{bmatrix} \\ &\times (\Xi^F(t_k))^{-1} \\ &= \begin{bmatrix} 0 & \mathbb{C}^F_{t_{k+1}}[\operatorname{vec} W(t_{k+1}), X(t_{k+1})] \end{bmatrix} (\Xi^F(t_k))^{-1}, \end{aligned} \tag{50c}$$

then

$$\begin{aligned} &\mathbb{E}^F_{t_k}[\operatorname{vec} W(t_{k+1}) \mid X(t_k), X(t_{k+1})] \\ &\approx \Gamma(t_k) \begin{bmatrix} X(t_k) - \bar{x}(t_k) \\ X(t_{k+1}) - \bar{x}(t^-_{k+1}) \end{bmatrix} \end{aligned} \tag{51a}$$

$$\begin{aligned} &\mathbb{V}^F_{t_k}[\operatorname{vec} W(t_{k+1}) \mid X(t_k), X(t_{k+1})] \\ &\approx \mathrm{I} - \Gamma(t_k) \Xi^F(t_k) \Gamma^{\mathsf{T}}(t_k). \end{aligned} \tag{51b}$$

Therefore, using the law of iterated expectation (see [19]), the smoothing moments for $\operatorname{vec} W(t_{k+1})$ and its cross-covariance with $X(t_k)$ and $X(t_{k+1})$ are given by

$$\mathbb{E}^S[\operatorname{vec} W(t_{k+1})] = \Gamma(t_k) \begin{bmatrix} \hat{x}(t_k) - \bar{x}(t_k) \\ \hat{x}(t_{k+1}) - \bar{x}(t^-_{k+1}) \end{bmatrix} \tag{52a}$$

$$\begin{aligned} &\mathbb{V}^S[\operatorname{vec} W(t_{k+1})] = \mathrm{I} \\ &+ \Gamma(t_k)(\Xi^S(t_k) - \Xi^F(t_k))\Gamma^{\mathsf{T}}(t_k) \end{aligned} \tag{52b}$$

$$\mathbb{C}^S\left[\begin{bmatrix} X(t_k) \\ X(t_{k+1}) \end{bmatrix}, \operatorname{vec} W(t_{k+1})\right] = \Gamma(t_k)\Xi^S(t_k). \tag{52c}$$

From all of the aforementioned quantities the joint smoothing distribution of $X(t_k)$, $X(t_{k+1})$, and $W(t_{k+1})$ can be obtained. The joint smoothing distribution of $X(t_k)$ and $W(t_{k+1})$ can then be obtained by selecting appropriate sub-matrices.

IV. Experimental Results

In this section the proposed methods are evaluated in two simulation experiments. In the first experiment (Section IV-A), the between measurement reconstruction is evaluated for the sigma-point implementation of the proposed smoother. In the second experiment (Section IV-B), the reconstruction at the measurement instances is evaluated.

A. Simulation: Hyperbolic tangent drift

In this experiment, the following system is considered.

$$\mathrm{d}X(t) = \tanh X(t)\,\mathrm{d}t + \frac{5}{100}\,\mathrm{d}B(t), \tag{53a}$$

$$Y(t_k) = X(t_k) + V(t_k), \tag{53b}$$

where $V(t_k) \sim \mathcal{N}(0, 1/10)$. The system is simulated 100 times for 10 time units using an Euler-Maruyama scheme with a step size of 1/100. The system is then downsampled by a factor 100 which gives the measurement intervals the competing smoothers are operating on. The proposed smoother is implemented with the sigma-point method using 1 (SE1SP), 5 (SE5SP), 10 (SE10SP), and 15 (SE15SP) cosine basis functions, respectively. These are then compared to sigma point (T3SP) and Taylor series (T3TS) implementations of the Type III smoothers (see [12]). All smoothers are implemented with first order exponential integrators and reconstruct the smoothing solution on the dense grid given by the simulation step size. An example of the reconstructions on a dense time grid (step size is 1/100) is shown in Figure 1.

As can be seen in Figure 3, the competing smoothers are fairly similar on average, with the basis expansion variants having a larger propensity for outliers in performance. Plausible

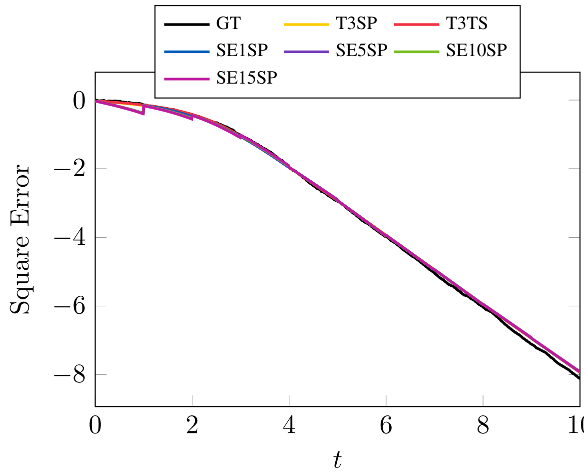

Figure 1. Reconstruction on a dense grid for the aforementioned smoothers.

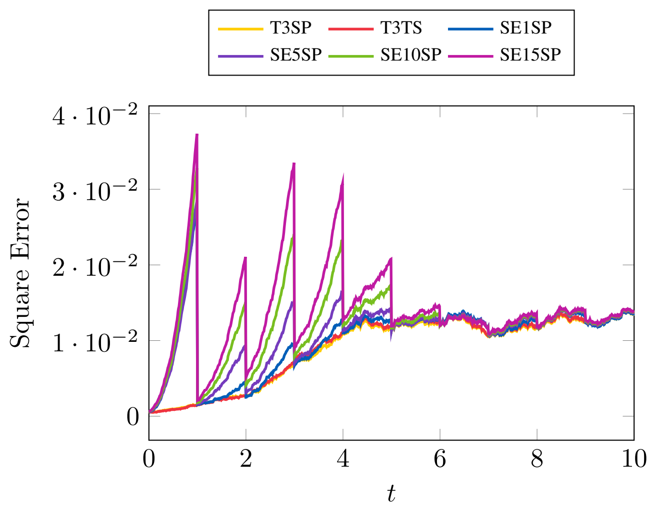

Figure 2. The square error of the trajectories for the aforementioned smoothers, averaged over Monte Carlo trials.

causes of this may be discerned from Figure 2 where the basis expansion approaches appear to be suffering from edge effects at the measurement points. This suggests improvements can be made to the method by which they reconstruct the trajectories between measurement points.

B. Simulation: Lorenz 63

In this experiment a chaotic system is considered. Namely the Lorenz 63 system, given by

$$\mathrm{d}X(t) = \begin{bmatrix} \sigma(X_2(t) - X_1(t)) \\ X_1(t)(\rho - X_3(t)) \\ X_1(t)X_2(t) - \beta X_3(t) \end{bmatrix} \mathrm{d}t + \begin{bmatrix} 0 & 0 \\ \frac{1}{100} & 0 \\ 0 & \frac{1}{100} \end{bmatrix} \mathrm{d}B(t),$$

with measurements given by

$$Y(t_k) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} X(t_k) + V(t_k), \tag{54}$$

where $V(t_k) \sim \mathcal{N}(0, 1/10)$. The system is simulated 100 times for 3.5 time units using an Euler-Maruyama scheme

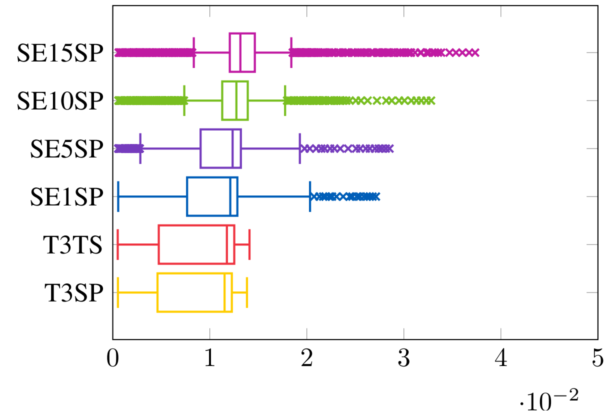

Figure 3. Boxplots of the root mean-square errors for the aforementioned smoothers.

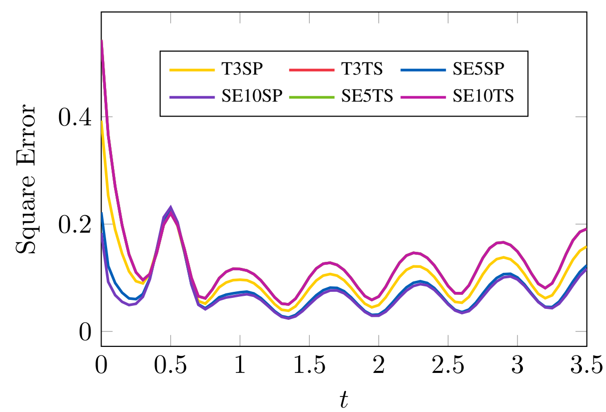

Figure 4. The square error of the trajectories for the aforementioned smoothers, averaged over Monte Carlo trials. As can be seen, both sigma point implementations of the basis expansion approach are almost uniformly superior to their competitors.

with a step size of $1/1000$. The system is then downsampled by a factor 50 which gives the measurement interval the competing smoothers are operating on. Two sigma-point implementations of the proposed smoother are implemented, using 5 (SE5SP) and 10 (SE10SP) basis functions, respectively. Additionally, two Taylor series implementations of the proposed smoother are implmented, again using 5 (SE5TS) and 10 (SE10TS) basis functions, respectively. All the aforementioned smoothers use the cosine basis. This collection of proposed smoothers are compared to sigma point (T3SP) and Taylor series (T3TS) implementations of the Type III smoothers. All the smoothers are use a step-size of $1/1000$ for integration and are implemented with first order exponential integrators.

As can be seen in Figures 4 and 5, the series expansion based sigma-point smoothers performs the best, followed by the sigma point implementation of the Type III smoother, and then the Taylor series based smoothers where its hard to discern a performance difference between them. This suggests that the error incurred by the Taylor series approximations dominates any other sources of errors.

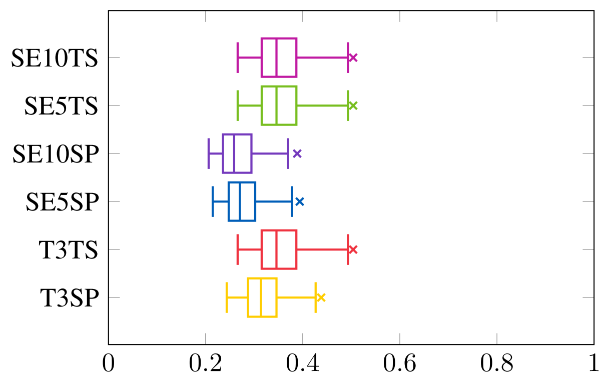

Figure 5. Boxplots of the root mean-square errors for the aforementioned smoothers. Note that the difference between the different Taylor series methods is quite small while the series expansion based sigma-point smoother offers a significant improvement over all the other alternatives for both 5 and 10 basis functions.

V. Conclusion

A novel class of non-linear continuous-discrete smoothers were derived based on series expansions of Brownian motion. We developed sigma-point and Taylor series versions of the smoothers. The sigma-point implementations were shown to be outperform other state-of-the-art smoothers in reconstructions at the measurement points, while the Taylor series implementation performed similarly to other Taylor series approaches. This suggests the error incurred by the Taylor expansion dominates any other source of errors. Furthermore, the method for reconstructing the trajectory between measurements, used by the basis expansion approaches, appear to suffer from edge effects at the measurement points, inviting further research into the topic. Another line of future research is to generalise the discrete time iterative smoothers to the current frame work [20], [21], [19].

Acknowledgment

Financial support by the Academy of Finland and ELEC Doctoral School at Aalto is acknowledged.

References

[1] D. H. Titterton and J. L. Weston, *Strapdown Inertial Navigation Technology*. The Institute of Electrical Engineers, 2004.
[2] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems*. Chapman & Hall/CRC, 2004.
[3] L. D. Stone, R. L. Streit, T. L. Corwin, and K. L. Bell, *Bayesian Multiple Target Tracking*. Artech House, 2014.
[4] E. Lindström, H. Madsen, and J. N. Nielsen, *Statistics for Finance*. Chapman and Hall/CRC, 2015.
[5] S. Särkkä, *Bayesian Filtering and Smoothing*, ser. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2013.
[6] B. Øksendal, *Stochastic Differnetial Equations - An Introduction with Applications*. Springer, 2003.
[7] R. Kalman and R. Bucy, "New results in linear filtering and prediction theory," *Transactions of the ASME, Journal of Basic Engineering*, vol. 83, pp. 95–108, 1961.
[8] H. Rauch, F. Tung, and C. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, 1965.
[9] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
[10] C. T. Leondes, J. B. Peller, and E. B. Stear, "Nonlinear smoothing theory," *IEEE Transactions on system science and cybernetics*, vol. 6, no. 1, pp. 63–71, 1970.
[11] S. Särkkä, "Continuous-time and continuous-discrete-time unscented Rauch-Tung-Striebel smoothers," *Signal Processing*, vol. 90, pp. 225–235, 2010.
[12] S. Särkkä and J. Sarmavuori, "Gaussian filtering and smoothing for continuous-discrete dynamic systems," *Signal Processing*, vol. 93, pp. 500–510, 2013.
[13] C. Archambeau, M. Opper, Y. Shen, D. Cornford, and J. S. Shawe-taylor, "Variational inference for diffusion processes," in *Advances in Neural Information Processing Systems 20*, J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, Eds. Curran Associates, Inc., 2008, pp. 17–24.
[14] J. Ala-Luhtala, S. Särkkä, and R. Piché, "Gaussian filtering and variational approximations for Bayesian smoothing in continuous-discrete stochastic dynamic systems," *Signal Processing*, vol. 111, no. Supplement C, pp. 124–136, 2015.
[15] E. Wong and M. Zakai, "On the convergence of ordinary integrals to stochastic integrals," *The Annals of Mathematical Statistics*, vol. 36, no. 5, pp. 1560–1564, 1965.
[16] W. Luo, "Wiener chaos expansion and numerical solutions of stochastic partial differential equations," Ph.D. dissertation, California Institute of Technology, 2006.
[17] S. M. Lyons, S. Särkkä, and A. J. Storkey, "Series expansion approximations of Brownian motion for non-linear Kalman filtering of diffusion processes," *IEEE Transactions on Signal Processing*, vol. 62, no. 6, pp. 1514–1524, 2014.
[18] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales, Volume I: Foundations*. Cambridge university Press, 2000.
[19] F. Tronarp, A. F. Garcia-Fernandez, and S. Särkkä, "Iterative filtering and smoothing in non-linear and non-Gaussian systems using conditional moments," *IEEE Signal Processing Letters*, vol. 25, no. 3, pp. 408–412, 2018.
[20] B. M. Bell, "The iterated Kalman smoother as a Gauss–Newton method," *SIAM Journal on Optimization*, vol. 4, no. 3, pp. 626–636, 1994.
[21] Á. F. García-Fernández, L. Svensson, and S. Särkkä, "Iterated posterior linearisation smoother," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 2056–2063, 2017.