Zhao, Zheng; Karvonen, Toni; Hostettler, Roland; Särkkä, Simo

Taylor Moment Expansion for Continuous-Discrete Gaussian Filtering

Published in:
IEEE Transactions on Automatic Control

DOI:
10.1109/TAC.2020.3047367

Published: 01/09/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Zhao, Z., Karvonen, T., Hostettler, R., & Särkkä, S. (2021). Taylor Moment Expansion for Continuous-Discrete Gaussian Filtering. *IEEE Transactions on Automatic Control*, *66*(9), 4460-4467. https://doi.org/10.1109/TAC.2020.3047367

Taylor Moment Expansion for Continuous-Discrete Gaussian Filtering

Zheng Zhao, Toni Karvonen, Roland Hostettler, *Member, IEEE*, and Simo Särkkä, *Senior Member, IEEE*

***Abstract*— The note is concerned with Gaussian filtering in non-linear continuous-discrete state-space models. We propose a novel Taylor moment expansion (TME) Gaussian filter which approximates the moments of the stochastic differential equation with a temporal Taylor expansion. Differently from classical linearisation or Itô–Taylor approaches, the Taylor expansion is formed for the moment functions directly and in time variable, not by using a Taylor expansion on the non-linear functions in the model. We analyse the theoretical properties, including the positive definiteness of the covariance estimate and stability of the TME filter. By numerical experiments, we demonstrate that the proposed TME Gaussian filter significantly outperforms the state-of-the-art methods in terms of estimation accuracy and numerical stability.**

***Index Terms*— continuous-discrete state-space model, Gaussian filtering, Kalman filtering, stochastic differential equation, Taylor moment expansion**

I. Introduction

In this note, we study Gaussian filtering of a continuous-discrete state-space model

$$\mathrm{d}\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, t)\,\mathrm{d}t + \mathbf{L}(\mathbf{x}_t, t)\,\mathrm{d}\mathbf{W}_t, \tag{1a}$$

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k, \tag{1b}$$

where $\mathbf{x}_t \in \mathbb{R}^D$ is a D-dimensional Itô process, $\mathbf{y}_k \in \mathbb{R}^Z$ is the measurement at time t_k, and $\mathbf{W}_t$ denotes an S-dimensional Wiener process with diffusion matrix $\mathbf{Q}$. We also assume the non-linear drift and dispersion functions $\mathbf{f}(\mathbf{x}_t, t)$ and $\mathbf{L}(\mathbf{x}_t, t)$ are sufficiently regular so that (1a) has a weakly unique solution [1], [2]. As we are mostly concerned with the continuous-time part (1a), for simplicity, we model the measurement $\mathbf{y}_k$ in (1b) with a non-linear function $\mathbf{h}(\mathbf{x}_k)$ and an additive noise $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_k)$, where $\mathcal{N}$ denotes a Gaussian distribution. Furthermore, we denote $\mathbf{\Gamma}(\mathbf{x}_t, t) = \mathbf{L}(\mathbf{x}_t, t)\,\mathbf{Q}\,\mathbf{L}^{\mathsf{T}}(\mathbf{x}_t, t)$, and Γ_{ij} denotes the i-th row and j-th column entry of $\mathbf{\Gamma}(\mathbf{x}_t, t)$. When $D = 1$ and $S = 1$, we use scalar notations x_t, $f(x_t, t)$ and $\Gamma(x_t, t)$.

The aim is to form Gaussian approximations to the filtering density for any t_k, $k = 1, \dots, T$ as follows:

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) \approx \mathcal{N}(\mathbf{x}_k \mid \mathbf{m}_k, \mathbf{P}_k). \tag{2}$$

Manuscript received Jan, 2020. Corresponding author: Zheng Zhao (email: zheng.zhao@aalto.fi).

Zheng Zhao, Toni Karvonen, and Simo Särkkä were with Department of Electrical Engineering and Automation, Aalto University, Finland.

Toni Karvonen was also with Alan Turing Institute, United Kingdom.

Roland Hostettler was with Department of Engineering Sciences, Uppsala University.

Above, we have used the notation $\mathbf{x}_k \triangleq \mathbf{x}_{t_k}$ at time t_k, and $\mathbf{y}_{1:k} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k\}$. Additionally, $\mathbf{m}_k$, $\mathbf{P}_k$ are the mean and covariance of (2), respectively.

In order to obtain the exact posteriors on the left hand side of (2), it would be necessary to compute the transition densities $p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$ for the continuous model (1a) (see, e.g., [3], [4]). It turns out that the transition density is only analytically tractable in limited cases, such as for linear stochastic differential equations (SDEs). In the Gaussian filtering framework, we are interested in constructing a Gaussian approximation to the transition density:

$$p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) \approx \mathcal{N}(\mathbf{x}_k \mid \mathbb{E}\left[\mathbf{x}_k \mid \mathbf{x}_{k-1}\right], \mathrm{Cov}\left[\mathbf{x}_k \mid \mathbf{x}_{k-1}\right]), \tag{3}$$

which is also the approach that we employ here. Above, $\mathbb{E}\left[\mathbf{x}_k \mid \mathbf{x}_{k-1}\right]$ and $\mathrm{Cov}\left[\mathbf{x}_k \mid \mathbf{x}_{k-1}\right]$ denote the conditional expectation and covariance of $\mathbf{x}_k$ given $\mathbf{x}_{k-1}$, respectively.

One classical way to approximate the transition density (3) is the Itô–Taylor expansion [5]–[7] which can be used to form a discretised solution to the SDE by expanding Itô integrals iteratively using Itô's lemma. The Euler–Maruyama scheme is the simplest instance of this kind of methods, and the mean and covariance in (3) are computed as

$$\begin{aligned} \mathbb{E}\left[\mathbf{x}_k \mid \mathbf{x}_{k-1}\right] &\approx \mathbf{x}_{k-1} + \mathbf{f}(\mathbf{x}_{k-1}, t_{k-1})\,\Delta t, \\ \mathrm{Cov}\left[\mathbf{x}_k \mid \mathbf{x}_{k-1}\right] &\approx \mathbf{\Gamma}(\mathbf{x}_{k-1}, t_{k-1})\,\Delta t. \end{aligned} \tag{4}$$

However, the Euler–Maruyama scheme only works well when the time interval Δt is small enough. Other commonly used choices are, for example, Milstein's method and the strong order 1.5 Itô–Taylor (Itô-1.5) method [4], [5]. However, because of the difficulty of the involved iterated Itô integrals, it is not easy to construct higher order Itô–Taylor expansions [5] and hence this approach is inherently low order in Δt. This motivates us to develop higher order weak approximations by using the proposed Taylor moment expansion.

Another widely used approach is to approximate ODEs for the first two moments of the Itô process [8]–[10]. The mean and covariance of the Itô process (1a) for any $t \in (t_{k-1}, t_k]$ are characterised by

$$\begin{aligned} \frac{\mathrm{d}\mathbf{m}_t}{\mathrm{d}t} &= \mathbb{E}\left[\mathbf{f}(\mathbf{x}_t, t)\right], \\ \frac{\mathrm{d}\mathbf{P}_t}{\mathrm{d}t} &= \mathbb{E}\left[\mathbf{f}(\mathbf{x}_t, t)\,(\mathbf{x}_t - \mathbf{m}_t)^{\mathsf{T}}\right] \\ &\quad + \mathbb{E}\left[(\mathbf{x}_t - \mathbf{m}_t)\,\mathbf{f}^{\mathsf{T}}(\mathbf{x}_t, t)\right] + \mathbb{E}\left[\mathbf{\Gamma}(\mathbf{x}_t, t)\right], \end{aligned} \tag{5}$$

where $\mathbf{m}_t = \mathbb{E}\left[\mathbf{x}_t\right]$ and $\mathbf{P}_t = \mathbb{E}\left[(\mathbf{x}_t - \mathbf{m}_t)\,(\mathbf{x}_t - \mathbf{m}_t)^{\mathsf{T}}\right]$ [4]. The initial values of $\mathbf{m}_t$ and $\mathbf{P}_t$ are given at time t_{k-1}.

Notice that when using this scheme in Gaussian filtering, it is not necessary to directly approximate the transition density (3). By solving the ODEs, they directly give the prediction step of filtering when the initial conditions are given by the previous filtering posterior. Unfortunately, the ODEs are only tractable for linear SDEs along with certain other isolated special cases, because of the expectations in (5). To disentangle the intractability problem of these ODEs, one practical solution is to linearise $\mathbf{f}(\mathbf{x}_t, t)$ and $\mathbf{L}(\mathbf{x}_t, t)$ around $\mathbf{m}_t$, which leads to the continuous-discrete extended Kalman filter (CD-EKF) [8]. Another solution is to assume that the densities are Gaussian, in which case the expectations in the ODEs can be calculated with Gaussian quadrature or sigma-point methods [11]–[13]. Finally, the ODEs are solved with numerical solvers, such as Runge–Kutta (RK) methods. For simplicity of later discussion, we use *Linear-ODE* and *Gauss-ODE* to refer to the methods solving (5) using linearisation and Gaussian assumptions, respectively.

The contributions of this note are as follows. (1) We develop a novel Taylor moment expansion based Gaussian filter for continuous-discrete state-space models. (2) We analyse the positive definiteness of TME covariance estimate and the stability of TME Gaussian filter. (3) We show by numerical experiments that the proposed TME Gaussian filter outperforms the state-of-the-art methods in terms of both estimation accuracy and numerical stability.

II. Taylor Moment Expansion for Gaussian Filtering

As a Gaussian distribution is entirely characterised by its mean and covariance, a reasonable approach to Gaussian filtering is to use moment matching to form a Gaussian approximation to the transition density. The previously presented Itô–Taylor and ODE methods are useful tools for this purpose. In this note, we present another *Taylor moment expansion* (TME) based approach, which allows us to derive higher order approximations to the transition density [4], [14]–[16].

A. Taylor Moment Expansion

Let $\phi(\mathbf{x}_t)$ be an arbitrary twice-differentiable scalar function of the process $\mathbf{x}_t$. By Itô's lemma and taking the expectation yields

$$\begin{aligned} \mathrm{d}\mathbb{E}\left[\phi(\mathbf{x}_t)\right] &= \mathbb{E}\left[\nabla\phi(\mathbf{x}_t)\,\mathbf{f}(\mathbf{x}_t,t)\right]\mathrm{d}t \\ &\quad + \frac{1}{2}\mathbb{E}\left[\mathrm{tr}\left(\nabla\nabla^{\mathsf{T}}\phi(\mathbf{x}_t)\,\mathbf{\Gamma}(\mathbf{x}_t,t)\right)\right]\mathrm{d}t, \end{aligned} \tag{6}$$

where ∇ and $\nabla\nabla^{\mathsf{T}}$ give the Jacobian and Hessian of $\phi(\mathbf{x}_t)$, respectively. With a proper choice of ϕ, this will lead to the moment ODEs as shown in (5) [4]. The aim now is to form a Taylor expansion of the function $\mathbb{E}\left[\phi(\mathbf{x}_t)\right]$. We notice that the right-hand side of (6) can be reformulated with the (generalized) infinitesimal generator

$$\begin{aligned} \mathcal{A}g &= \frac{\partial g}{\partial t} + \nabla g\,\mathbf{f}(\mathbf{x}_t,t) + \frac{1}{2}\,\mathrm{tr}\big(\nabla\nabla^{\mathsf{T}} g\,\mathbf{\Gamma}(\mathbf{x}_t,t)\big), \\ &= \frac{\partial g}{\partial t} + \sum_i \frac{\partial g}{\partial x_i} f_i(\mathbf{x}_t,t) + \frac{1}{2}\sum_{i,j}\frac{\partial^2 g}{\partial x_i \partial x_j}\Gamma_{ij}, \end{aligned} \tag{7}$$

for any regular smooth function g, where x_i is the i-th component of $\mathbf{x}_t$ (see, e.g., [1, Ch. 7] or [4, Ch. 9]). Thus (6) becomes

$$\frac{\mathrm{d}\mathbb{E}\left[\phi(\mathbf{x}_t)\right]}{\mathrm{d}t} = \mathbb{E}\left[\mathcal{A}\phi(\mathbf{x}_t)\right], \tag{8}$$

which requires that $\phi \in C^2$ is twice differentiable. We also denote by $\mathcal{A}^r$ the r-th iteration of the generator. By taking derivatives of (8) to M times, we have

$$\frac{\mathrm{d}^M\mathbb{E}\left[\phi(\mathbf{x}_t)\right]}{\mathrm{d}t^M} = \mathbb{E}\left[\mathcal{A}^M\phi(\mathbf{x}_t)\right], \tag{9}$$

which requires that $\phi \in C^{2M}$ [4]. Notice that (9) above also requires sufficient smoothness of $\mathbf{f}(\mathbf{x}, t)$ and $\mathbf{\Gamma}(\mathbf{x}_t, t)$. We can now form an M-th order Taylor expansion of the function $\mathbb{E}\left[\phi(\mathbf{x}_k)\right]$ at time t_k and centred at time t_{k-1} as follows:

$$\begin{aligned} \mathbb{E}\left[\phi(\mathbf{x}_k)\right] &\approx \sum_{r=0}^{M}\frac{1}{r!}\frac{\mathrm{d}^r\mathbb{E}\left[\phi(\mathbf{x}_{k-1})\right]}{\mathrm{d}t^r}\,\Delta t^r \\ &= \sum_{r=0}^{M}\frac{1}{r!}\mathbb{E}\left[\mathcal{A}^r\phi(\mathbf{x}_{k-1})\right]\,\Delta t^r, \end{aligned} \tag{10}$$

where $\Delta t = t_k - t_{k-1}$. Conditioning (10) on $\mathbf{x}_{k-1}$ gives

$$\mathbb{E}\left[\phi(\mathbf{x}_k) \mid \mathbf{x}_{k-1}\right] \approx \sum_{r=0}^{M}\frac{1}{r!}\mathcal{A}^r\phi(\mathbf{x}_{k-1})\,\Delta t^r. \tag{11}$$

In Gaussian filtering, we are only interested in functions ϕ having certain polynomial forms. For the mean and covariance, we introduce two sets of functions: $\{\phi_i = x_i : i = 1, \ldots, D\}$ and $\{\phi_{ij} = x_i\,x_j : i,j = 1,\ldots,D\}$, where x_i is the i-th component of $\mathbf{x}_k$. Then, we have the mean $\mathbb{E}\left[\mathbf{x}_k \mid \mathbf{x}_{k-1}\right] = \left[\mathbb{E}\left[\phi_1 \mid \mathbf{x}_{k-1}\right], \ldots, \mathbb{E}\left[\phi_D \mid \mathbf{x}_{k-1}\right]\right]^{\mathsf{T}}$ and the covariance $\mathrm{Cov}\left[x_i\,x_j \mid \mathbf{x}_{k-1}\right] = \mathbb{E}\left[\phi_{ij} \mid \mathbf{x}_{k-1}\right] - \mathbb{E}\left[\phi_i \mid \mathbf{x}_{k-1}\right]\mathbb{E}\left[\phi_j \mid \mathbf{x}_{k-1}\right]$. Using the approximation (11), we can now form the Taylor moment expansion (TME) estimator for the transition density as shown in Definition 1.

Definition 1 (Taylor Moment Expansion (TME) of Transition Density)**.** *The M-th order Taylor expansion based estimates of the mean $\mathbf{a}_M$, the second moment $\mathbf{B}_M$, and the covariance $\mathbf{\Sigma}_M$ of the transition density* (3) *are given by*

$$\begin{aligned} \mathbf{a}_M &\triangleq \mathbf{a}_M(\mathbf{x}_{k-1}, \Delta t) \\ &= \sum_{r=0}^{M}\frac{1}{r!}\,\mathcal{A}^r\mathbf{x}_{k-1}\,\Delta t^r \approx \mathbb{E}\left[\mathbf{x}_k \mid \mathbf{x}_{k-1}\right], \\ \mathbf{B}_M &\triangleq \mathbf{B}_M(\mathbf{x}_{k-1}, \Delta t) \\ &= \sum_{r=0}^{M}\frac{1}{r!}\,\mathcal{A}^r\left(\mathbf{x}_{k-1}\,\mathbf{x}_{k-1}^{\mathsf{T}}\right)\Delta t^r \approx \mathbb{E}\left[\mathbf{x}_k\,\mathbf{x}_k^{\mathsf{T}} \mid \mathbf{x}_{k-1}\right], \\ \mathbf{\Sigma}_M &\triangleq \mathbf{\Sigma}_M(\mathbf{x}_{k-1}, \Delta t) \\ &= \mathbf{B}_M - \mathbf{a}_M\,\mathbf{a}_M^{\mathsf{T}} \approx \mathrm{Cov}\left[\mathbf{x}_k \mid \mathbf{x}_{k-1}\right], \end{aligned} \tag{12}$$

respectively. Here the application of the generator $\mathcal{A}$ on vector or matrix input means that we apply the operator elementwise.

Remark 2. *Note that if an M-th order TME approximation is used, the covariance estimator $\mathbf{\Sigma}_M$ in Definition 1 is a* *polynomial of degree $2M$ in Δt, which comes from the product $\mathbf{a}_M\,\mathbf{a}_M^\mathsf{T}$. To keep the order of Δt consistent in mean and covariance, $\mathbf{\Sigma}_M$ needs to be truncated to degree M.*

In addition, it is also important to recover the remainder $R(\mathbf{x}_{k-1}, \Delta t)$ of TME, such that (10) becomes

$$\mathbb{E}\left[\phi(\mathbf{x}_k)\right] = \sum_{r=0}^{M} \frac{1}{r!}\mathbb{E}\left[\mathcal{A}^r \phi(\mathbf{x}_{k-1})\right]\,\Delta t^r + R(\mathbf{x}_{k-1}, \Delta t).$$

By Taylor's theorem, the remainder admits the form

$$R(\mathbf{x}_{k-1}, \Delta t) = \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{u_M}\cdots\int_{t_{k-1}}^{u_1}\mathbb{E}\left[\mathcal{A}^{M+1}\phi(\mathbf{x}_s)\right]\mathrm{d}s\,\mathrm{d}u_1\cdots\mathrm{d}u_M, \tag{13}$$

provided that ϕ and the SDE coefficients are sufficiently smooth [14], [16]. The convergence properties of TME can be analysed in the same way as any other Taylor expansion.

The difference between the TME and the aforementioned ODE and Itô–Taylor schemes is mainly how the Gaussian approximation to the continuous model is done. The Itô–Taylor approach first discretises the SDE solution which gives the discretised approximation $\hat{\mathbf{x}}_k$, and then obtains the moment $\mathbb{E}\left[\phi(\hat{\mathbf{x}}_k)\right]$ using the approximation. In ODE approach we need to postulate certain hypothesis, such as Gaussian assumptions or linearisation before solving the ODEs (5). In contrast, TME gives the estimate of $\mathbb{E}\left[\phi(\mathbf{x}_k)\right]$ directly, without forming the discretised approximation $\hat{\mathbf{x}}_k$ or approximation to the ODEs.

B. TME Gaussian Filtering

Using Definition 1, we now formulate the proposed TME Gaussian filter by utilising an M-th order TME estimate of the transition density $p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) \approx \mathcal{N}(\mathbf{x}_k \mid \mathbf{a}_M, \mathbf{\Sigma}_M)$. Notice that although we are using simplified notations $\mathbf{a}_M$, $\mathbf{B}_M$, and $\mathbf{\Sigma}_M$, those terms are functions of $\mathbf{x}_{k-1}$ and Δt.

Let us assume the filtering posterior from previous time step t_{k-1} is $p(\mathbf{x}_{k-1} \mid \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{x}_{k-1} \mid \mathbf{m}_{k-1}, \mathbf{P}_{k-1})$. We first perform prediction with respect to the continuous model (1a), and thus the prediction density $p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{x}_k \mid \mathbf{m}_k^-, \mathbf{P}_k^-)$ is characterised by

$$\begin{aligned}&\mathbb{E}\left[\mathbf{x}_k \mid \mathbf{y}_{1:k-1}\right] \\ &= \int \mathbf{x}_k \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1})\, p(\mathbf{x}_{k-1} \mid \mathbf{y}_{1:k-1})\,\mathrm{d}\mathbf{x}_{k-1}\,\mathrm{d}\mathbf{x}_k \\ &\approx \mathbb{E}\left[\mathbf{a}_M\right] = \mathbf{m}_k^-,\end{aligned} \tag{14}$$

and also

$$\begin{aligned}&\mathrm{Cov}\left[\mathbf{x}_k \mid \mathbf{y}_{1:k-1}\right] \\ &\approx \mathbb{E}\left[\mathbf{\Sigma}_M + \mathbf{a}_M\,\mathbf{a}_M^\mathsf{T}\right] - \mathbf{m}_k^-\,(\mathbf{m}_k^-)^\mathsf{T} = \mathbf{P}_k^-.\end{aligned} \tag{15}$$

Note that we are using $\mathbf{P}_k^- = \mathbb{E}\left[\mathbf{\Sigma}_M + \mathbf{a}_M\,\mathbf{a}_M^\mathsf{T}\right] - \mathbf{m}_k^-\,(\mathbf{m}_k^-)^\mathsf{T}$ instead of directly $\mathbf{P}_k^- = \mathbb{E}\left[\mathbf{B}_M\right] - \mathbf{m}_k^-\,(\mathbf{m}_k^-)^\mathsf{T}$. Recall from Remark 2 that they are not equal, as we truncate $\mathbf{\Sigma}_M$ to keep the power of Δt consistent. By using $\mathbf{P}_k^- = \mathbb{E}\left[\mathbf{B}_M\right] - \mathbf{m}_k^-\,(\mathbf{m}_k^-)^\mathsf{T}$, it is difficult to perform such truncation. Furthermore, the prediction covariance $\mathbf{P}_k^-$ is not always positive definite, which is an issue that is discussed in Section III-A.

The resulting filtering algorithm is the following, where we also use the general discrete Gaussian filter update step [3].

Algorithm 3 (TME Gaussian Filter)**.** *Starting from initial filtering condition $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{x}_0 \mid \mathbf{m}_0, \mathbf{P}_0)$, the equations of TME Gaussian filter for $k = 1, 2, \ldots, T$ are as follows:*

- *Prediction:*

$$\begin{aligned}\mathbf{m}_k^- &= \int \mathbf{a}_M\,\mathcal{N}(\mathbf{x}_{k-1} \mid \mathbf{m}_{k-1}, \mathbf{P}_{k-1})\,\mathrm{d}\mathbf{x}_{k-1}, \\ \mathbf{P}_k^- &= \int \left(\mathbf{\Sigma}_M + \mathbf{a}_M\,\mathbf{a}_M^\mathsf{T}\right)\mathcal{N}(\mathbf{x}_{k-1} \mid \mathbf{m}_{k-1}, \mathbf{P}_{k-1})\,\mathrm{d}\mathbf{x}_{k-1} \\ &\quad - \mathbf{m}_k^-\,(\mathbf{m}_k^-)^\mathsf{T}.\end{aligned} \tag{16}$$

- *Update:*

$$\begin{aligned}\mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k\,(\mathbf{y}_k - \boldsymbol{\mu}_k), \\ \mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^\mathsf{T},\end{aligned} \tag{17}$$

where $\boldsymbol{\mu}_k = \mathbb{E}[\mathbf{h}(\mathbf{x}_k)]$, $\mathbf{S}_k = \mathbb{E}\left[(\mathbf{h}(\mathbf{x}_k) - \boldsymbol{\mu}_k)\,(\mathbf{h}(\mathbf{x}_k) - \boldsymbol{\mu}_k)^\mathsf{T}\right]$, and $\mathbf{K}_k = \mathbb{E}\left[(\mathbf{x}_k - \mathbf{m}_k^-)\,(\mathbf{h}(\mathbf{x}_k) - \boldsymbol{\mu}_k)^\mathsf{T}\right]\mathbf{S}_k^{-1}$. In this update step, the expectations are taken with respect to the predicted $\mathbf{x}_k \sim \mathcal{N}(\mathbf{m}_k^-, \mathbf{P}_k^-)$.

The calculation of Gaussian integrals in Algorithm 3 is intractable for many non-linear integrands (i.e., $\mathbf{a}_M$, $\mathbf{B}_M$, and $\mathbf{\Sigma}_M$). Herein, we consider numerically approximating them by using quadrature and sigma-point methods, such as Gauss–Hermite [11], unscented transform [17], and spherical cubature method [18]. It is also worth mentioning that as the sigma-point approximation is an operation of linearly weighted summation, the positive definiteness of $\mathbf{P}_k^-$ is preserved from $\mathbf{\Sigma}_M$ provided that the quadrature weights are positive. This is true for Gauss–Hermite quadrature, spherical cubature, and unscented transformation with suitable parameter selection. In this note we assume that such a positive-weight integration rule is used.

III. Theoretical Analysis of Taylor Moment Expansion Gaussian Filter

In this section, we first study the positive definiteness of the covariance estimates produced by the Taylor moment expansion (TME) and then prove the stability of the TME Gaussian filter for a class of non-linear state-space models.

A. Positive Definiteness of Taylor Moment Expansion

In the Gaussian filtering context, it is essential for the covariance estimate to stay positive definite (p.d.). Unfortunately, this is not always true when using TME, as we truncate the full Taylor expansion [19]. We now formulate the following theorem to show conditions for the positive definiteness of the TME covariance estimate.

Theorem 4. *The M-th order TME covariance estimate $\mathbf{\Sigma}_M$ is positive definite for Δt on an interval $U \subseteq \mathbb{R}^+$, if*

$$P_M(\Delta t) = \sum_{r=1}^{M} w_r\,\Delta t^r > 0, \tag{18}$$

for all $\Delta t \in U$, where $w_r = \frac{1}{r!}\lambda_{\min}\left(\mathbf{\Phi}_{\mathbf{x}_t,r}\right)$, and $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of a square matrix. The coefficients are

$$\mathbf{\Phi}_{\mathbf{x}_t,r} = \mathcal{A}^r\left(\mathbf{x}_t\,\mathbf{x}_t^{\mathsf{T}}\right) - \sum_{s=0}^{r}\binom{r}{s}\mathcal{A}^s\mathbf{x}_t\left(\mathcal{A}^{r-s}\mathbf{x}_t\right)^{\mathsf{T}}, \tag{19}$$

where $\binom{r}{s}$ denotes the binomial coefficient.

Proof. By Cauchy product and Definition 1, we can formulate the (u,v)-th entry of $\mathbf{\Sigma}_M$ as

$$[\Sigma_M]_{uv} = \sum_{r=0}^{M}\frac{1}{r!}\left[\mathcal{A}^r(x_u x_v) - \sum_{s=0}^{r}\binom{r}{s}\mathcal{A}^s x_u\,\mathcal{A}^{r-s}x_v\right]\Delta t^r$$

and further rearrange it into matrix form

$$\mathbf{\Sigma}_M = \sum_{r=1}^{M}\frac{1}{r!}\mathbf{\Phi}_{\mathbf{x}_t,r}\Delta t^r. \tag{20}$$

Using Weyl's inequality [20], we have $\lambda_{\min}\left(\mathbf{\Sigma}_M\right) \geq \sum_{r=1}^{M}\frac{1}{r!}\lambda_{\min}\left(\mathbf{\Phi}_{\mathbf{x}_t,r}\right)\Delta t^r$. Thus the positive definiteness of $\mathbf{\Sigma}_M$ is implied by the polynomial on the right in (20). □

Equation (20) reveals that the covariance estimate $\mathbf{\Sigma}_M$ is a polynomial function of Δt with coefficients formed by $\{\mathbf{\Phi}_{\mathbf{x}_t,r}\colon r = 1,\ldots,M\}$. The idea behind Theorem 4 is to construct a lower bound for the minimum eigenvalue of $\mathbf{\Sigma}_M$ in terms of the eigenvalues of the coefficients matrices $\mathbf{\Phi}_{\mathbf{x}_t,r}$. The following Proposition 5 gives applications of using Theorem 4 for the positive definiteness of the second and third order of TME estimates.

Proposition 5. *Let $\mathbf{\Sigma}_2$ and $\mathbf{\Sigma}_3$ be the TME estimates of the covariance with expansion order 2 and 3, respectively. Also let us denote by $\mathbf{f} \triangleq \mathbf{f}(\mathbf{x}_t,t)$ and $\mathbf{\Gamma} \triangleq \mathbf{\Gamma}(\mathbf{x}_t,t)$. Then*

1) *$\mathbf{\Sigma}_2$ is p.d. for $\Delta t > 0$ if $(\mathbf{\Gamma}\,\nabla\mathbf{f})^{\mathsf{T}} + \mathbf{\Gamma}\,\nabla\mathbf{f}$ and $\mathbf{\Gamma}$ are positive semi-definite (p.s.d.), and one of $\mathbf{\Gamma}$ and $(\mathbf{\Gamma}\,\nabla\mathbf{f})^{\mathsf{T}} + \mathbf{\Gamma}\,\nabla\mathbf{f}$ is p.d.*
2) *$\mathbf{\Sigma}_3$ is p.d. for $\Delta t > 0$ if $\mathbf{\Phi}_{\mathbf{x}_t,3}$ is p.s.d. and $\lambda_{\min}(\mathbf{\Phi}_{\mathbf{x}_t,2}) > \frac{-2\sqrt{6}}{3}\sqrt{\lambda_{\min}(\mathbf{\Phi}_{\mathbf{x}_t,1})\,\lambda_{\min}(\mathbf{\Phi}_{\mathbf{x}_t,3})}$.*

Proof. By Definition 1,

$$\begin{aligned}\mathbf{\Sigma}_2 &= \mathbf{\Phi}_{\mathbf{x}_t,1}\Delta t + \frac{1}{2}\mathbf{\Phi}_{\mathbf{x}_t,2}\Delta t^2 \\ &= \mathbf{\Gamma}\Delta t + \frac{1}{2}\left((\mathbf{\Gamma}\,\nabla\mathbf{f})^{\mathsf{T}} + \mathbf{\Gamma}\,\nabla\mathbf{f}\right)\Delta t^2.\end{aligned} \tag{21}$$

Thus $\boldsymbol{\eta}^{\mathsf{T}}\,\mathbf{\Sigma}_2\,\boldsymbol{\eta} > 0$ for any real non-zero vector $\boldsymbol{\eta}$ and $\Delta t > 0$, if $\lambda_{\min}(\mathbf{\Gamma}) > 0$ and $\lambda_{\min}((\mathbf{\Gamma}\,\nabla\mathbf{f})^{\mathsf{T}} + \mathbf{\Gamma}\,\nabla\mathbf{f}) \geq 0$ or $\lambda_{\min}(\mathbf{\Gamma}) = 0$ and $\lambda_{\min}((\mathbf{\Gamma}\,\nabla\mathbf{f})^{\mathsf{T}} + \mathbf{\Gamma}\,\nabla\mathbf{f}) > 0$.

For $\mathbf{\Sigma}_3$, by Theorem 4, we have the polynomial

$$P_3(\Delta t) = w_1\Delta t + w_2\Delta t^2 + w_3\Delta t^3. \tag{22}$$

The polynomial $P_3(\Delta t)$ is positive and has no real roots for $\Delta t > 0$, if and only if $w_2 > -2\sqrt{w_1 w_3}$ and $w_3 \geq 0$, which is equivalent to $\lambda_{\min}(\mathbf{\Phi}_{\mathbf{x}_t,2}) > \frac{-2\sqrt{6}}{3}\sqrt{\lambda_{\min}(\mathbf{\Phi}_{\mathbf{x}_t,1})\,\lambda_{\min}(\mathbf{\Phi}_{\mathbf{x}_t,3})}$ and $\lambda_{\min}(\mathbf{\Phi}_{\mathbf{x}_t,3}) \geq 0$. It follows that $\mathbf{\Sigma}_3$ is p.d. □

Remark 6. *In the limit $\Delta t \to 0$, the TME covariance estimate will always be p.d., provided $\mathbf{\Gamma}$ is p.d.*

Example 7. *The TME variance estimate of SDE*

$$\mathrm{d}x_t = \tanh x_t\,\mathrm{d}t + \mathrm{d}W_t, \tag{23}$$

where W_t is a standard Wiener process, is always p.d. This follows from Theorem 4 and observing that $\Phi_{x,1} = 1 > 0$, $\Phi_{x,2} = 1 - \tanh^2 x_t \geq 0$, and $\{\Phi_{x,r} = 0\colon r \geq 3\}$.

The coefficients $\mathbf{\Phi}_{\mathbf{x}_t,r}$, the expansion order M, and the time interval Δt jointly determine the positive definiteness of TME covariance estimate. The properties of $\mathbf{\Phi}_{\mathbf{x}_t,r}$ are more of interest, as we usually have M and Δt fixed. Next, we show that $\mathbf{\Phi}_{\mathbf{x}_t,r}$ is only concerned with the SDE coefficients.

Lemma 8. *Consider the SDE (1a) with time-homogeneous $\mathbf{f}(\mathbf{x}_t)$ and constant $\mathbf{L}$. Let $\Phi^{u,v}_{\mathbf{x},r} \triangleq [\mathbf{\Phi}_{\mathbf{x},r}]_{uv}$ be the u-th column and v-th row entry of $\mathbf{\Phi}_{\mathbf{x},r}$. We denote $\alpha^u_r \triangleq \alpha_r(x_u) = \mathcal{A}^r(x_u)$, and partial derivative $\partial_i\alpha^u_r \triangleq \partial\alpha^u_r/\partial x_i$. Then a general expression of $\Phi^{u,v}_{\mathbf{x},r\geq 1}$ is*

$$\begin{aligned}\Phi^{u,v}_{\mathbf{x},r} &= \sum_{i,j}^{D}\sum_{s=0}^{r-1}\binom{r-1}{s}\left(\partial_i\alpha^u_s\,\partial_j\alpha^v_{r-s-1}\right)\Gamma_{ij} + \mathcal{A}\Phi^{u,v}_{\mathbf{x},r-1} \\ &= \sum_{s=0}^{r-1}\mathcal{A}^s\sum_{l=0}^{r-s-1}\binom{r-s-1}{l}\operatorname{tr}\left((\nabla\alpha^u_s)^{\mathsf{T}}\,\nabla\alpha^v_{r-s-1-l}\,\mathbf{\Gamma}\right)\end{aligned} \tag{24}$$

starting from $\Phi^{u,v}_{\mathbf{x},0} = 0$.

Proof. See Appendix I. □

Lemma 8 gives an explicit form of $\mathbf{\Phi}_{\mathbf{x}_t,r}$, which is shown to be the function of $\mathbf{f}$, $\mathbf{\Gamma}$, and their partial derivatives. It implies that once M and Δt are given, the positive definiteness of $\mathbf{\Sigma}_M$ fully depends on the SDE coefficients. The functions $\mathbf{f}$ and $\mathbf{\Gamma}$ have to satisfy certain properties for $\mathbf{\Sigma}_M$ to be positive definite.

Example 9. *Let us consider a one-dimensional Itô process*

$$\mathrm{d}x_t = f(x_t)\,\mathrm{d}t + L\,\mathrm{d}W_t, \tag{25}$$

then by Lemma 8,

$$\begin{aligned}\Phi_{x,0} &= 0, \\ \Phi_{x,1} &= \Gamma, \\ \Phi_{x,2} &= 2f'\Gamma, \\ \Phi_{x,3} &= 2(2(f')^2 + 2ff'' + f'''\Gamma)\Gamma, \\ \Phi_{x,4} &= ((9(f'')^2 + 6ff'''' + 16f'''f')\Gamma + 8(f')^3 \\ &\quad + 6f^2f''' + 26ff''f' + \frac{24}{16}f'''''\Gamma^2)\Gamma, \\ &\;\;\vdots\end{aligned} \tag{26}$$

where $f', f'', \ldots, f''''', \ldots$ are the derivatives of $f(x)$ of orders $1, 2, \ldots, 5, \ldots$. Also, $\Gamma = L^2\,Q$, and Q is the diffusion constant of the Wiener process W_t.

B. Stability of TME Gaussian Filter

It is important and useful that the filter is in some sense *stable*. Some classical stability results for linear Kalman filters can be found in [8], [21] while more recent results on the stability of different approximate Gaussian (Kalman) filters for non-linear systems have been analysed in [22]–[24]. In this section, we follow [24] and prove that the TME Gaussian filter is stable in the mean-square sense if a number of assumptions on the system and the sigma-point approximation, verifiable before the filter is run, are satisfied. This means that we show that

$$\sup_{k\geq 1} \mathbb{E}\big[\|\mathbf{x}_k - \mathbf{m}_k\|_2^2\big] < \infty,$$

where the expectation is taken over all state and measurement trajectories, and $\|\cdot\|_2$ denotes the Euclidean norm.

For simplicity, in (1) we assume $\mathbf{f}(\mathbf{x}_t)$ is time-homogeneous (not explicitly depending on time), $\mathbf{\Gamma}$ is a positive definite constant, $\mathbf{V}_k = \mathbf{V}$ for all $k \geq 1$, and that the measurement model is linear: $\mathbf{h}(\mathbf{x}_k) = \mathbf{H}\,\mathbf{x}_k$ for some matrix $\mathbf{H}$. Then the generic continuous-discrete SDE (1) then has the discretised form

$$\begin{aligned} \mathbf{x}_k &= \mathbf{a}(\mathbf{x}_{k-1}, \Delta t) + \boldsymbol{\tau}(\mathbf{x}_{k-1}, \Delta t), \\ \mathbf{y}_k &= \mathbf{H}\,\mathbf{x}_k + \mathbf{v}_k, \end{aligned} \tag{27}$$

where $\boldsymbol{\tau}(\mathbf{x}_{k-1}) \triangleq \boldsymbol{\tau}(\mathbf{x}_{k-1}, \Delta t)$ is a zero-mean random variable whose covariance $\operatorname{Cov}[\boldsymbol{\tau}(\mathbf{x}_{k-1})] = \mathbf{\Sigma}(\mathbf{x}_{k-1}, \Delta t)$. Notice that we denote by $\mathbf{a}(\mathbf{x}_{k-1}) \triangleq \mathbf{a}(\mathbf{x}_{k-1}, \Delta t)$ and $\mathbf{\Sigma}(\mathbf{x}_{k-1}) \triangleq \mathbf{\Sigma}(\mathbf{x}_{k-1}, \Delta t)$ the *exact* mean and covariance functions of $\mathbf{x}_t$, respectively. It follows that $\mathbf{a} = \mathbf{a}_M + \mathbf{R}_M$, where $\mathbf{R}_M(\mathbf{x}_{k-1}) \triangleq \mathbf{R}_M(\mathbf{x}_{k-1}, \Delta t)$ is the Taylor remainder (13). The assumptions needed for the stability analysis of the system (27) are collected below in Assumption 10. If $\mathbf{A}$ is a matrix, $\|\mathbf{A}\|$ stands for the spectral norm.

Assumption 10. *The following properties hold:*

1) *There are non-negative constants C_M, λ_τ, and λ_P such that $\sup_{k\geq 1}\|\mathbf{R}_M(\mathbf{x}_{k-1})\| \leq C_M$ almost surely, $\sup_{k\geq 1}\mathbb{E}[\operatorname{tr}(\mathbf{\Sigma}(\mathbf{x}_{k-1}))] \leq \lambda_\tau$, and $\sup_{k\geq 1}\mathbb{E}[\operatorname{tr}(\mathbf{P}_k)] \leq \lambda_P$.*
2) *There is $C \geq 0$ such that*

$$\|\mathbf{a}_M(\mathbf{x}) - \mathcal{S}_{\mathbf{m},\mathbf{P}}(\mathbf{a}_M)\|_2 \leq \|\nabla \mathbf{a}_M(\mathbf{x})\|^2 \|\mathbf{x} - \mathbf{m}\|_2^2 + C\operatorname{tr}(\mathbf{P})$$

for any vectors $\mathbf{x}$ and $\mathbf{m}$ and any positive semi-definite matrix $\mathbf{P}$, where $\mathcal{S}_{\mathbf{m},\mathbf{P}}(\mathbf{g})$ stands for the sigma-point approximation of the Gaussian integral

$$\int \mathbf{g}(\mathbf{x})\,\mathcal{N}(\mathbf{x} \mid \mathbf{m}, \mathbf{P})\,\mathrm{d}\mathbf{x}.$$

3) *There is $\lambda \geq 0$ such that $\sup_{k\geq 1}\|\mathbf{I} - \mathbf{K}_k\mathbf{H}\| \leq \lambda$ almost surely and*

$$\lambda_f^2 \triangleq \sup_{k\geq 1} \lambda^2 \sup_{\mathbf{x}} \|\nabla \mathbf{a}_M(\mathbf{x})\|^2 < 1/4.$$

The Assumption 10 postulates conditions on the sigma-point approximations and systems. An example to satisfy the assumptions is that the drift function $\mathbf{f}$ is smooth enough and all of its partial derivatives up to certain orders are uniformly bounded. More practical examples that satisfy Assumption 10 can be found in [24]. It is typically necessary that the measurement model matrix $\mathbf{H}$ is a scaled identity matrix and the discretised dynamics $\mathbf{a}$ in (27) defines an exponentially stable system. Although these assumptions are quite restrictive, stability results with more general assumptions are currently not yet available in literature [24].

Theorem 11. *Suppose that Assumption 10 is satisfied. Then the TME Gaussian filter for system* (27) *has*

$$\mathbb{E}\left[\|\mathbf{x}_k - \mathbf{m}_k\|_2^2\right] \leq (4\lambda_f^2)^k \operatorname{tr}(\mathbf{P}_0) + \frac{C_e}{1 - 4\lambda_f^2}$$

for all $k \geq 1$, where C_e is defined in (29).

Proof. It is easy to see that [24, Proof of Theorem IV.3]

$$\kappa \triangleq \sup_{k\geq 1}\|\mathbf{K}_k\| \leq \lambda_P \|\mathbf{H}\| \|\mathbf{V}^{-1}\|.$$

Denote $\mathbf{A}_k = \mathbf{I} - \mathbf{K}_k\mathbf{H}$. Using the discretised system (27), the filtering error can be written as

$$\begin{aligned} \mathbf{e}_k &\triangleq \mathbf{x}_k - \mathbf{m}_k \\ &= \mathbf{a}(\mathbf{x}_{k-1}) + \boldsymbol{\tau}(\mathbf{x}_{k-1}) - \mathbf{m}_k^- - \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{H}\,\mathbf{m}_k^-\right) \\ &= \mathbf{A}_k\left[\mathbf{a}(\mathbf{x}_{k-1}) - \mathcal{S}_{\mathbf{m}_{k-1},\mathbf{P}_{k-1}}(\mathbf{a}_M)\right] \\ &\quad + (\mathbf{I} - \mathbf{K}_k\,\mathbf{H})\,\boldsymbol{\tau}(\mathbf{x}_{k-1}) - \mathbf{K}_k\,\mathbf{v}_k \\ &= \mathbf{A}_k\left[\mathbf{a}_M(\mathbf{x}_{k-1}) - \mathcal{S}_{\mathbf{m}_{k-1},\mathbf{P}_{k-1}}(\mathbf{a}_M)\right] \\ &\quad + \mathbf{A}_k\,\mathbf{R}_M(\mathbf{x}_{k-1}) + \mathbf{A}_k\,\boldsymbol{\tau}(\mathbf{x}_{k-1}) - \mathbf{K}_k\,\mathbf{v}_k. \end{aligned}$$

The inequality $(a_1 + \cdots + a_n)^2 \leq n(a_1^2 + \cdots + a_n^2)$ gives

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{e}_k\|_2^2\right] &\leq 4\mathbb{E}\left[\left\|\mathbf{A}_k\,\left[\mathbf{a}_M(\mathbf{x}_{k-1}) - \mathcal{S}_{\mathbf{m}_{k-1},\mathbf{P}_{k-1}}(\mathbf{a}_M)\right]\right\|_2^2\right] \\ &\quad + 4\mathbb{E}\big[\|\mathbf{A}_k\,\mathbf{R}_M(\mathbf{x}_{k-1})\|_2^2\big] \\ &\quad + 4\mathbb{E}\big[\|\mathbf{A}_k\,\boldsymbol{\tau}(\mathbf{x}_{k-1})\|_2^2\big] + 4\mathbb{E}\big[\|\mathbf{K}_k\,\mathbf{v}_k\|_2^2\big]. \end{aligned} \tag{28}$$

Assumption 10 yields the following bounds:

$$\begin{aligned} \mathbb{E}\Big[\big\|\mathbf{A}_k\left[\mathbf{a}_M(\mathbf{x}_{k-1}) - \mathcal{S}_{\mathbf{m}_{k-1},\mathbf{P}_{k-1}}(\mathbf{a}_M)\right]\big\|_2^2\Big] \\ \leq \lambda_f^2\mathbb{E}\big[\|\mathbf{e}_{k-1}\|_2^2\big] + C\lambda^2\lambda_P, \\ \mathbb{E}\big[\|\mathbf{A}_k\,\mathbf{R}_M(\mathbf{x}_{k-1})\|_2^2\big] \leq C_M^2\lambda^2, \\ \mathbb{E}\big[\|\mathbf{A}_k\,\boldsymbol{\tau}(\mathbf{x}_{k-1})\|_2^2\big] \leq \lambda^2\lambda_\tau, \\ \mathbb{E}\big[\|\mathbf{K}_k\,\mathbf{v}_k\|_2^2\big] \leq \operatorname{tr}(\mathbf{V})\,\kappa^2. \end{aligned}$$

Upon insertion of these estimates into (28) we get the recursive mean-square error inequality

$$\mathbb{E}\big[\|\mathbf{e}_k\|_2^2\big] \leq 4\lambda_f^2\,\mathbb{E}\big[\|\mathbf{e}_{k-1}\|_2^2\big] + C_e,$$

where

$$C_e = 4\big(\lambda^2[C\lambda_P + C_M^2 + \lambda_\tau] + \operatorname{tr}(\mathbf{V})\kappa^2\big). \tag{29}$$

Because we have assumed that $4\lambda_f^2 < 1$, the claim then follows from the discrete Grönwall's inequality (e.g., [24, Theorem IV.2]). $\square$

IV. Numerical Experiments

To examine the effectiveness of the TME estimator in Definition 1, we first conduct experiments on the moment estimation of SDEs. After that, we examine and compare the accuracy and numerical stability of the proposed TME Gaussian filter against state-of-the-art methods.

A. Moment Estimation of SDEs

We consider a non-linear SDE:

$$\mathrm{d}x_t = \tanh x_t \, \mathrm{d}t + \mathrm{d}W_t, \tag{30}$$

where W_t is a standard Wiener process, with known initial condition $x_0 = 0.5$. The aim is to compare the estimates of the moments of the transition densities. The true mean and covariance are estimated using Monte Carlo (MC) sampling with 10^6 independent trajectories. We simulated the samples from the models using Euler–Maruyama with sufficiently small time interval (10^{-5} s). The estimates were examined in the time interval $T = 0$ s to $T = 5$ s.

We chose the following methods as described in Section I to compare with the TME method:

- the ODE approach by solving (5) using Gaussian assumption, 4th order Runge–Kutta solver, and 3rd order Gauss–Hermite integration (Gauss-RK4);
- the ODE approach by solving (5) using linearisation, and 4th order Runge–Kutta solver (Linear-RK4);
- the Itô–Taylor strong order 1.5 based approach from [5], [6] (Itô-1.5).

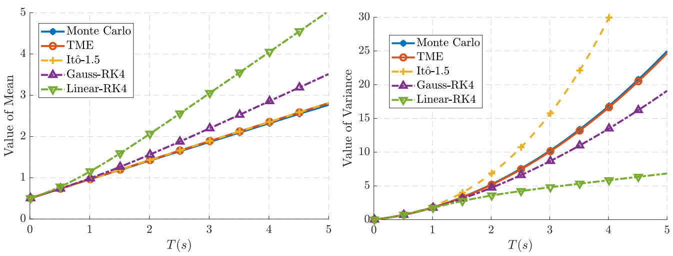

Fig. 1: The mean and variance estimates of model (30).

In Figure 1, we show the mean and variance estimates as functions of time. We observe that the Itô-1.5 and (the second order) TME methods coincide for the estimation of mean function, and are closest to the Monte Carlo result. This is because their formulations for this model are identical, and exact to the true mean of (30). The Gauss-RK4 and Linear-RK4 can only estimate the mean accurately within short time intervals, while Gauss-RK4 is slightly better than Linear-RK4.

When estimating the variance of (30), only the TME method succeeds to follow the Monte Carlo result closely. The variance estimate yield by TME is $\Delta t + (1 - \tanh^2 x_0)\Delta t^2$, which is exact to this model (see, Example 7). The Gauss-RK4, Linear-RK4, and Itô-1.5 all deviate from Monte Carlo for long time intervals.

B. 3D Coordinated Turn Tracking

In this part, we conduct Gaussian filtering on a 3D coordinated turn model. Performing filtering on this model is considered challenging due to its non-linearities and high dimensionality [6], [25]. The model is given by

$$\mathrm{d}\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\mathbf{W}_t, \tag{31a}$$

$$\mathbf{y}_k = \begin{bmatrix} \sqrt{p_x^2 + p_y^2 + p_z^2} \\ \tan^{-1}(p_y/p_x) \\ \tan^{-1}(p_z/\sqrt{p_x^2+p_y^2}) \end{bmatrix} + \boldsymbol{\epsilon}_k, \tag{31b}$$

where the state $\mathbf{x}_t = \begin{bmatrix} p_x & v_x & p_y & v_y & p_z & v_z & \theta \end{bmatrix}^\mathsf{T}$ and

$$\begin{aligned} \mathbf{f}(\mathbf{x}_t) &= \begin{bmatrix} v_x & -\theta v_y & v_y & \theta v_x & v_z & 0 & 0 \end{bmatrix}^\mathsf{T}, \\ \mathbf{L} &= \operatorname{diag}\begin{bmatrix} 0 & \sigma_1 & 0 & \sigma_1 & 0 & \sigma_1 & \sigma_2 \end{bmatrix}. \end{aligned} \tag{32}$$

In this model, we denote by p_x, p_y, and p_z the position of the target in Cartesian coordinates, and v_x, v_y, and v_z are the corresponding velocities. State θ governs the turning rate of the target which controls the non-linearity of this model. In addition, $\mathbf{W}_t$ is a standard Wiener process. The measurement noise $\boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$ and $\mathbf{V} = \operatorname{diag}\begin{bmatrix}\sigma_r^2 & \sigma_\theta^2 & \sigma_\phi^2\end{bmatrix}$.

Name	Description	Method	Integration
EKF-RK	ODE type of Gaussian filter with 4th order Runge–Kutta solver [4], [13]	Solving (5) with linearisation (Linear-ODE)	Not needed
CKF-RK		Solving (5) with Gaussian assumption (Gauss-ODE)	Spherical cubature
UKF-RK			Unscented transform
GHKF-RK			3rd order Gauss–Hermite
CKF-1.5	Itô-1.5 Gaussian filter [6]	Itô–Taylor discretisation with strong order 1.5 [6]	Spherical cubature
UKF-1.5			Unscented transform
GHKF-1.5			3rd order Gauss–Hermite
CKF-T*	TME Gaussian filter (Alg. 3)	$*$-th order TME (Def. 1)	Spherical cubature
UKF-T*			Unscented transform
GHKF-T*			3rd order Gauss–Hermite

TABLE I: The list of the state-of-the-art methods compared in 3D coordinate turn tracking.

The parameters σ_1, σ_2, σ_r^2, σ_θ^2, and σ_ϕ^2 are the same as in [6]. However, we choose the initial turning rate $\theta_0 = 30°/\text{s}$ which is significantly higher than the $\theta_0 = 3$ to $6°/\text{s}$ used in [6]. These two θ_0 settings are illustrated in Figure 2.

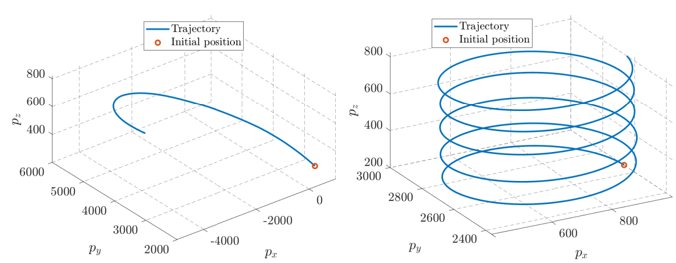

Fig. 2: Examples of trajectory simulation ($t = 0$ s to 60 s). The left figure shows the setting of initial turning rate $\theta_0 = 3°/\text{s}$ [6], while on the right figure, we use $\theta_0 = 30°/\text{s}$.

The other parameters are the same as in [6]. The initial condition is drawn from a normal distribution with mean $\mathbf{m}_0 = \begin{bmatrix} 1000\,\text{m} & 0\,\text{m/s} & 2650\,\text{m} & 150\,\text{m/s} & 200\,\text{m} & 10\,\text{m/s} & 30°/\text{s} \end{bmatrix}^\mathsf{T}$ and covariance $\mathbf{P}_0 = \operatorname{diag}\begin{bmatrix} 100^2 & 100^2 & 100^2 & 100^2 & 100^2 & 100^2 & 10^2 \end{bmatrix}$. We simulate the ground-truth trajectories using Euler–Maruyama with small enough time step $\Delta t \times 10^{-5}$ s, where Δt is the time interval between two measurements. The total time length of the trajectory is fixed to $T = 210$ s. To test the effectiveness of filters, we select the time interval Δt range from 0.5 s to

9 s. We also use additional integration steps $\log_2(M)$ in the prediction steps. For the comparison of the filtering accuracy, we run 100 Monte Carlo trials and calculate the root mean square error (RMSE) of the position states, which is defined as

$$\mathrm{RMSE} = \sqrt{\sum_{i}^{100} \sum_{j}^{T/\Delta t} \sum_{k=1,3,5} \frac{(\mathbf{x}_k^{i,j} - \hat{\mathbf{x}}_k^{i,j})^2}{3 \times 100 \times (T/\Delta t)}}, \tag{33}$$

where $\mathbf{x}_1^{i,j}$, $\mathbf{x}_3^{i,j}$, and $\mathbf{x}_5^{i,j}$ are the ground truth positions at time $j\Delta t$ from i-th Monte Carlo trial, and the $\hat{\mathbf{x}}_1^{i,j}$, $\hat{\mathbf{x}}_3^{i,j}$, and $\hat{\mathbf{x}}_5^{i,j}$ are their corresponding filtering estimates. For the comparison of numerical stability, we record the number of divergences, which is defined by counting the manifestation of non-positive definite covariances and unbounded (NaN) estimates. The state-of-the-art Gaussian filters for comparison are listed in Table I.

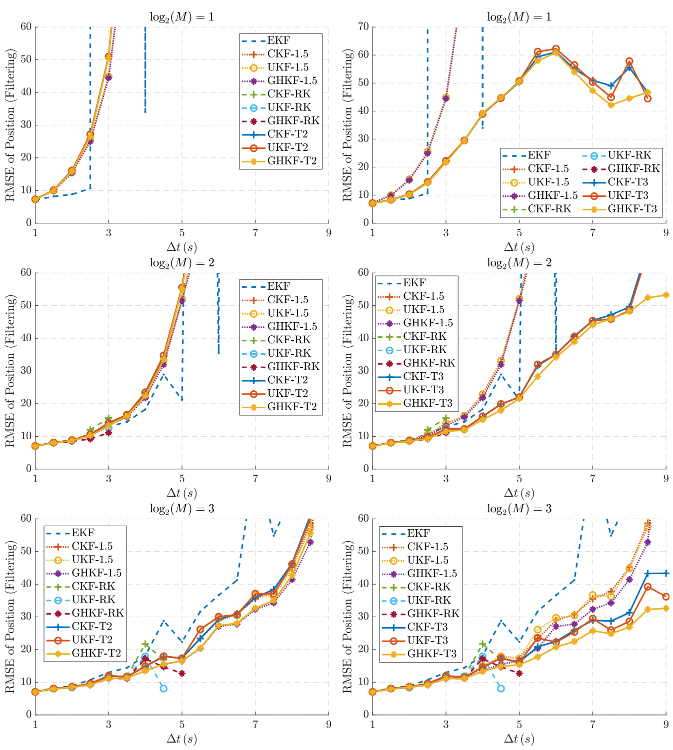

Fig. 3: The RMSE of filtering over 100 Monte Carlo runs. On the rows, we use different integrations steps $\log_2(M)$. On the columns, we use the 2nd and 3rd TME expansion.

In Figure 3, the filtering RMSE of the position is shown. It reveals that (second order) TME-2 filters are comparable to the Itô-1.5 filters, as their performances are almost identical. The (third order) TME-3 filters significantly outperform other filters across different Δt and $\log_2(M)$. More importantly, TME-3 filters especially have less errors for large Δt. On the other hand, the ODE type of filters (i.e., EKF and CKF/UKF/GHKF-RK) perform the worst and diverge for large Δt.

We next examine the numerical stability by recording the number of divergences in filtering as shown in Figure 4. First, we find that the stability of ODE type of filters is the worst. The UKF/CKF/GHKF-RK filters diverge for almost every Δt.

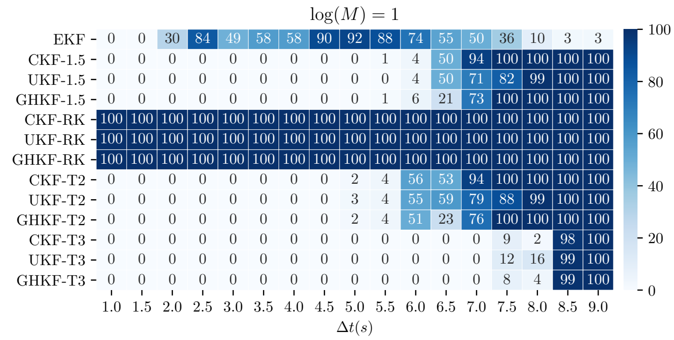

Fig. 4: Number of divergences of the filters. These numbers are recorded once the non-positive definite covariance estimate manifests or the estimate is unbounded (NaN errors).

The stability of Linear-ODE type of filter is also poor and inconsistent. On the contrary, TME-3 filters are the most stable, while Itô-1.5 filters are slightly better than of TME-2.

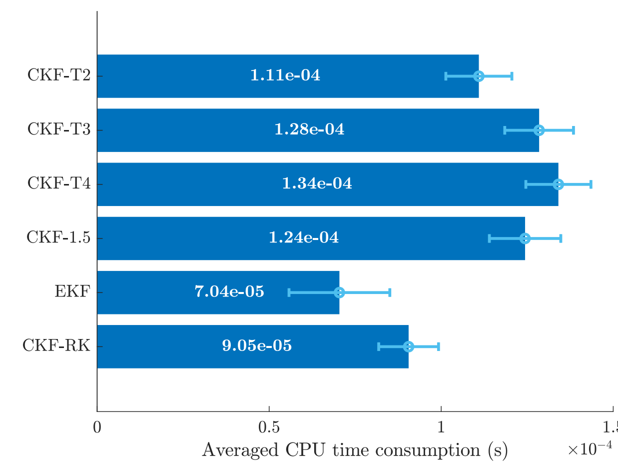

Fig. 5: Averaged CPU time from 10^6 runs on an Xeon® E3-1230v5 workstation. The error bars indicate the standard deviation of the Monte Carlo runs.

To compare the actual computational efficiencies of filters, we perform 10^6 independent runs of the prediction steps and calculate the mean CPU time consumption. We focus on comparing only the prediction step because that is the only difference among the involved Gaussian filters. We uniformly use the same cubature integration method for the Gaussian integral evaluations. The result is illustrated in Figure 5. We observe that the ODE type of filters (EKF and CKF-RK) are the most efficient. For the time-discretisation based methods (TME and Itô-1.5), the CKF-T2 and CKF-T4 are the least and most time consumable, respectively. However, TME and Itô-1.5 performs similarly, for example, the CKF-T4 is only slightly higher than Itô-1.5 with 0.1×10^{-4} s. It does not take significant computation resources by increasing the TME expansion order.

V. Conclusion

We have proposed a novel Gaussian filter for non-linear continuous-discrete models which forms a Gaussian approximation to the transition density of the SDE by using a Taylor moment expansion (TME). We derived the corresponding TME Gaussian filter, and analysed the positive definiteness of the TME covariance estimates and stability of the TME Gaussian filter. The numerical experiments indicate that even a second order TME Gaussian filter is in line with the state-of-the-art. With higher expansion order, the proposed TME filter substantially outperform the state-of-the-art methods, in terms of both estimation accuracy and numerical stability.

Appendix I
Proof of Lemma 8

Let $\beta_r^{u,v} \triangleq \beta_r(x_u\, x_v) = \mathcal{A}^r(x_u\, x_v)$ denote the r-th iteration of operator $\mathcal{A}$ on $x_u x_v$, and $\partial_i \beta_r^{u,v} = \partial\beta_r^{u,v}/\partial x_i$. For $r=0$ and 1, we readily find $\alpha_0^u = x_u$, $\alpha_1^u = f_u$, $\alpha_0^v = x_v$, $\alpha_1^v = f_v$, $\beta_0^{u,v} = x_u x_v = \alpha_0^u\alpha_0^v$, and $\beta_1^{u,v} = \alpha_0^u\alpha_1^v + \alpha_0^v\alpha_1^u + \Gamma_{uv}$.

We can calculate $\Phi_{x,r}^{u,v}$ by (19) and initially reveal $\Phi_{x,0}^{u,v} = 0$, $\Phi_{x,1}^{u,v} = \Gamma_{uv}$. From this pattern above, we will first show $\beta_r^{u,v}$ has a general expression

$$\beta_r = \sum_{s=0}^{r} \binom{r}{s} \alpha_s^u \alpha_{r-s}^v + \Phi_{x,r}^{u,v}, \tag{34}$$

where $\Phi_{x,r}^{u,v}$ is

$$\Phi_{x,r}^{u,v} = \sum_{i,j}\sum_{s=0}^{r-1} \binom{r-1}{s} \left(\partial_i\alpha_s^u\, \partial_j\alpha_{r-s}^v\right)\Gamma_{ij} + \mathcal{A}\Phi_{x,r-1}^{u,v}, \tag{35}$$

It is apparent that (34) and (35) hold for $r=1$. By Algorithm 1, the iteration of $\beta_r^{u,v}$ is

$$\beta_{r+1}^{u,v} = \mathcal{A}(\beta_r^{u,v}) = \sum_i \frac{\partial\beta_r^{u,v}}{\partial x_i} f_i + \frac{1}{2}\sum_{i,j}\frac{\partial^2\beta_r^{u,v}}{\partial x_i\partial x_j}\Gamma_{ij}. \tag{36}$$

With Equation (34), we continue Equation (36) to yield

$$\begin{aligned}\beta_{r+1}^{u,v} &= \sum_{s=0}^{r}\binom{r}{s}\left(\alpha_{s+1}^u\alpha_{r-s}^v + \alpha_s^u\alpha_{r-s+1}^v\right) \\ &\quad + \sum_{i,j}\left(\sum_{s=0}^{r}\binom{r}{s}\partial_i\alpha_s^u\partial_j\alpha_{r-s}^v\right)\Gamma_{ij} + \mathcal{A}\Phi_{x,r}^{u,v} \qquad (37)\\ &= \sum_{s=0}^{r+1}\binom{r+1}{s}\alpha_s\alpha_{r-s+1} + \Phi_{x,r+1}^{u,v} = \beta_{r+1}.\end{aligned}$$

Thus expressions (34) and (35) are proved by induction for $r \geq 1$. From Lemma 4, we now derive an iterated form of $\Phi_{x,r}$ as

$$\begin{aligned}\Phi_{x,r}^{u,v} &= \beta_r^{u,v} - \sum_{s=0}^{r}\binom{r}{s}\alpha_s^u\alpha_{r-s}^v \\ &= \sum_{i,j}\sum_{s=0}^{r-1}\binom{r-1}{s}\left(\partial_i\alpha_s^u\,\partial_j\alpha_{r-s-1}^v\right)\Gamma_{ij} + \mathcal{A}\Phi_{x,r-1}^{u,v} \\ &= \sum_{s=0}^{r-1}\binom{r-1}{s}\operatorname{tr}\left((\nabla\alpha_s^u)^\mathsf{T}\,\nabla\alpha_{r-s-1}^v\,\mathbf{\Gamma}\right) + \mathcal{A}\Phi_{x,r-1}^{u,v} \\ &= \sum_{s=0}^{r-1}\mathcal{A}^s\sum_{l=0}^{r-s-1}\binom{r-s-1}{l}\operatorname{tr}\left((\nabla\alpha_s^u)^\mathsf{T}\,\nabla\alpha_{r-s-1-l}^v\,\mathbf{\Gamma}\right)\end{aligned}$$

starting from $\Phi_{x,0} = 0$.

Acknowledgment

The authors would like to thank Academy of Finland for financial support.

References

[1] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 6th ed. Springer, 2003.

[2] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales, Volume 2: Itô Calculus*, 2nd ed., ser. Cambridge Mathematical Library. Cambridge University Press, 2000.

[3] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[4] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations*, ser. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2019.

[5] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*. Springer–Verlag Berlin Heidelberg, 1992.

[6] I. Arasaratnam, S. Haykin, and T. R. Hurd, "Cubature Kalman filtering for continuous-discrete systems: Theory and simulations," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 4977–4993, Oct 2010.

[7] T. Knudsen and J. Leth, "A new continuous discrete unscented Kalman filter," *IEEE Transactions on Automatic Control*, vol. 64, no. 5, pp. 2198–2205, 2019.

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

[9] G. Kulikov and M. Kulikova, "Accurate continuousdiscrete unscented Kalman filtering for estimation of nonlinear continuous-time stochastic models in radar tracking," *Signal Processing*, vol. 139, pp. 25–35, 2017.

[10] S. Koyama, "Projection smoothing for continuous and continuous-discrete stochastic dynamic systems," *Signal Processing*, vol. 144, pp. 333–340, 2018.

[11] K. Itô and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[12] S. Särkkä, "On unscented Kalman filtering for state estimation of continuous-time nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1631–1641, Sep. 2007.

[13] S. Särkkä and J. Sarmavuori, "Gaussian filtering and smoothing for continuous-discrete dynamic systems," *Signal Processing*, vol. 93, no. 2, pp. 500–510, 2013.

[14] M. Kessler, "Estimation of an ergodic diffusion from discrete observations," *Scandinavian Journal of Statistics*, vol. 24, no. 2, pp. 211–229, 1997.

[15] D. Dacunha-Castelle and D. Florens-Zmirou, "Estimation of the coefficients of a diffusion from discrete observations," *Stochastics*, vol. 19, no. 4, pp. 263–284, 1986.

[16] D. Florens-zmirou, "Approximate discrete-time schemes for statistics of diffusion processes," *Statistics*, vol. 20, no. 4, pp. 547–557, 1989.

[17] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, March 2004.

[18] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, June 2009.

[19] S. M. Iacus, *Simulation and Inference for Stochastic Differential Equations: With R Examples*. Springer-Verlag New York, 2008.

[20] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[21] B. D. O. Anderson and J. B. Moore, "Detectability and stabilizability of time-varying discrete-time linear systems," *SIAM Journal on Control and Optimization*, vol. 19, no. 1, pp. 20–32, 1981.

[22] K. Reif, S. Gnther, E. Yaz, and R. Unbehauen, "Stochastic stability of the discrete-time extended Kalman filter," *IEEE Transactions on Automatic Control*, vol. 44, no. 4, pp. 714–728, Apr 1999.

[23] K. Xiong, H. Zhang, and C. W. Chan, "Performance evaluation of UKF-based nonlinear filtering," *Automatica*, vol. 42, no. 2, pp. 261–270, 2006.

[24] T. Karvonen, S. Bonnabel, E. Moulines, and S. Särkkä, "On stability of a class of filters for non-linear stochastic systems," *SIAM Journal on Control and Optimization, To appear*, 2020.

[25] Y. Bar-Shalom, T. Kirubarajan, and X.-R. Li, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: John Wiley & Sons, Inc., 2002.