Stochastic Differential Equations with State Dependent Diffusion - 2 Order Statistics and State Estimation

Knudsen, Torben

Published in:
IFAC-PapersOnLine

DOI (link to publication from Publisher):
10.1016/j.ifacol.2023.10.623

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 5943–5950

Stochastic Differential Equations with State Dependent Diffusion - 2 Order Statistics and State Estimation

Torben Knudsen *

* *Aalborg University, Department of Electronic Systems, Section of Automation and Control, Denmark (e-mail: tk@es.aau.dk).*

Abstract: Estimation of states in stochastic differential equations with state dependent diffusion is known to be difficult. Previous research recommend the higher order extended Kalman filter or the Lamperti transform method for this case. This paper shows that a new developed method, based on the unscented Kalman filter, is superior for two simulated stochastic differential equation systems.

Keywords: Estimation and filtering; Stochastic system identification; Continuous time system estimation; Nonlinear system identification; Stochastic differential equations; Unscented Kalman filter; Lamperti transform; Higher order Kalman filter.

1. INTRODUCTION

Stochastic state space models are extensively used for state and parameter estimation for forecasting or control. For Gaussian linear systems the Kalman filter (KF) is used. This method has a strong theoretical base and is known to be the mean square state error optimal estimator (Maybeck, 1982). This holds for both discrete and continuous state space models. Here we shall focus on the latter which is called a stochastic differential equation (SDE). If the drift term in a SDE is nonlinear in the state, approximation to the optimal state estimator must be used. The most well known is the extended KF (EKF). However, even this filter has a poor performance if the model has state dependent diffusion (SDD) according to Baadsgaard et al. (1997); Nielsen and Madsen (2001); Møller and Madsen (2010). In the case of SDD Jazwinski (1970); Maybeck (1982) suggest to use higher order EKF filters (HOEKF). Unfortunately, Møller and Madsen (2010) report that HOEKF tend to have numerical and stability issues for SDD SDEs. For this case Baadsgaard et al. (1997); Nielsen and Madsen (2001); Møller and Madsen (2010) instead suggest to use the Lamperti transform to provide a state transformation that turns the SDD SDE into a non SDD SDE. However, this is only possible for a limited class of SDDs. The unscented KF (UKF) has a good performance for SDEs with highly nonlinear drift (Knudsen and Leth, 2019).

The main contribution in this paper is the new UKF method for SDD SDEs based on the unscented transform. And how this performance compares to the Lamperti and HOEKF solution. The brief conclusion is that the new UKF method is the superior choice.

The state estimation is divided in a time update and a measurement update(Knudsen and Leth, 2019; Maybeck, 1982). The time update amounts to updating the conditional state mean and covariance from one time to another. The measurement update can differ between the methods depending on the specific measurement equation. However, for the chosen two example systems the measurement equation is common for all methods, and it will not be further discussed in this paper. All the problems discussed above stems from the time update part which is therefore in focus for the rest of the paper.

The remaining part of the paper starts with the mathematical formulation of the model, and some basic theory which is used in the paper. Section 3 and 4 presents the approaches from the literature for the time update, and section 5 presents the new UKF based approach. Section 6 shows examples of the methods ability to calculate long term mean and variance for one dimensional SDEs with known solutions. Section 7 compares the methods performance when used for state estimation based on sampled measurements. A conclusion is drawn in section 8.

2. STOCHASTIC DIFFERENTIAL EQUATIONS AND BASIC THEORY

The SDE is written in the common way (1)

$$dx = f(x,t)dt + g(x,t)dw \,, \tag{1}$$

where $x \in R^n$, $w \in R^l$ is a Wiener process with incremental covariance I and f, g are functions of appropriate dimensions. The SDE is in the sense of Ito.

As x is a Markov process, the complete description is given by the transition probability $\varphi(x,t|x_0,t_0)$, which is related to the SDE (1) by the forward Kolmogorov equation (2), which is also called the Fokker-Planck equation. In (2) $[\cdot]_{ij}$ means component i,j.

$$\begin{aligned}\frac{\partial \varphi(x,t|x_0,t_0)}{\partial t} &= -\sum_{i=1}^{n} \frac{\partial \varphi(x,t|x_0,t_0)[f(x,t)]_i}{\partial x_i} \\ &+ \tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 \varphi(x,t|x_0,t_0)[g(x,t)g(x,t)^{\mathrm{T}}]_{ij}}{\partial x_i \, \partial x_j}\end{aligned} \tag{2}$$

10.1016/j.ifacol.2023.10.623

A simpler second order description of x given by the mean $m(t)$ and covariance $P(t)$, can be derived from the forward Kolmogorov equation, which results in the moment ordinary differential equations (MODE) below (Maybeck, 1982).

$$\dot{m}(t) = \mathrm{E}(f(x(t),t)) \tag{3a}$$

$$\begin{aligned}\dot{P}(t) = &\mathrm{E}((f(x(t),t)x(t)^{\mathsf{T}})) - \mathrm{E}(f(x(t),t))m(t)^{\mathsf{T}} \\ &+ \left(\mathrm{E}((f(x(t),t)x(t)^{\mathsf{T}})) - \mathrm{E}(f(x(t),t))m(t)^{\mathsf{T}}\right)^{\mathsf{T}} \\ &+ \mathrm{E}(g(x(t),t)g(x(t),t)^{\mathsf{T}})\end{aligned} \tag{3b}$$

Both equations (2) and (3) only have a closed form solution in rare cases as e.g. the linear model, but the latter (3) is used to derive the HOEKF.

A straight forward discrete time approximation to the SDE (1), is given by the Euler-Maruyama (EM) method (4), where the version (4b) uses the Δ operator.

$$\begin{aligned}\varkappa(t_{i+1}) - \varkappa(t_i) =& f(\varkappa(t_i),t_i)(t_{i+1}-t_i) \\ &+ g(\varkappa(t_i),t_i)(w(t_{i+1}) - w(t_i))\end{aligned} \Leftrightarrow \tag{4a}$$

$$\Delta\varkappa(t_i) = f(\varkappa(t_i),t_i)\Delta t_i + g(\varkappa(t_i),t_i)\Delta w(t_i) \tag{4b}$$

As $\Delta w(t_i)$ are independent increments $\varkappa(t_i)$ is a discrete time Markov process. According to Kloeden and Platen (1992) the time discrete approximation (4) converges to the continuous time solution (1) as given by (5), where $\delta = t_{i+1} - t_i$ is the time step and h is any function where the expected value exists.

$$\begin{aligned}&\exists K, \delta_0 > 0 : \delta < \delta_0 \Rightarrow \\ &\quad \mathrm{E}(|x(t_i) - \varkappa(t_i)|) \leq K\delta^{\gamma}\ ,\ \gamma = \tfrac{1}{2}\end{aligned} \tag{5a}$$

$$\begin{aligned}&\exists K, \delta_0 > 0 : \delta < \delta_0 \Rightarrow \\ &\quad |\,\mathrm{E}(h(x(t_i))) - \mathrm{E}(h(\varkappa(t_i)))| \leq K\delta^{\gamma}\ ,\ \gamma = 1\end{aligned} \tag{5b}$$

The EM method converges slowly path wise (5a), and faster moment wise (5b).

The last important theory component is the Ito differentiation rule (IDR). Introduce a function $z(x,t)$, which is continuously differentiable in t, and twice continuously differentiable in x. If x is given by the SDE (1), and z is assumed scalar for simplicity, it will be given by the following SDE (Åström, 1970).

$$\begin{aligned}dz = &\left(\frac{\partial z}{\partial t} + \left(\frac{\partial z}{\partial x}\right)^{\mathsf{T}} f(x,t) + \right. \\ &\left.\tfrac{1}{2}\,\mathrm{trace}\left(\frac{\partial^2 z}{\partial x\,\partial x^{\mathsf{T}}} g(x,t)g(x,t)^{\mathsf{T}}\right)\right)dt + \left(\frac{\partial z}{\partial x}\right)^{\mathsf{T}} g(x,t)dw\end{aligned} \tag{6}$$

3. THE LAMPERTI TRANSFORM APPROACH

The key idea in the Lamperti approach, is to exploit the IDR (6) to obtain a state independent diffusion term. In the scalar case this is obtained by using the transform

$$z(x,t) = \int \frac{1}{g(v,t)}dv\bigg|_{v=x} \tag{7}$$

The Lamperti approach consist of the following steps:

(1) Using (7), find a transformation that turns the SDD SDE into a non SDD SDE.
(2) Transform the initial state mean and covariance using this transformation.
(3) Use IDR to find the non SDD SDE for the transformed process.
(4) Based on the transformed SDE, use linearization i.e. EKF, to calculate future conditional mean and covariance given present values.
(5) Use the inverse transformation to calculate the conditional mean and covariance for the original process.

The Lamperti transformation approach are only possible for a limited class of SDEs (Møller and Madsen, 2010). In the scalar case the integral (7) is well defined and can have an analytical solution. As seen in the examples in section 6, it is necessary to derive the inverse of the transform $z(x,t)$ i.e. $x(z,t)$, which is not always possible (Møller and Madsen, 2010, Example 4). Also time dependent diffusion gives rise to very complicated transform SDEs (Møller and Madsen, 2010). For vector valued SDEs the Lamperti approach does not work for the general SDE (1). However, if the diffusion term can be factorized as (8) there is a solution (Møller and Madsen, 2010). Notice that d is a diagonal matrix where element i d_{ii} only depend on state element i x_i.

$$\begin{aligned}&g(x,t) = d(x,t)r(t)\ ,\ d,r \in R^{n\times n}\ , \\ &d(x,t) = \mathrm{diag}\,(d_{11}(x_1,t)\ ,\ \cdots\ ,\ d_{nn}(x_n,t))\end{aligned} \tag{8}$$

4. THE EKF AND HOEKF APPROACH

There are a number of different HOEKF methods. They can be interpreted as choosing simplifying assumptions, and using the MODE (3).

For the linear time invariant SDE where $f(x,t) = Fx$ and $g(x,t) = G$ in (1) the MODE (3) reduces to

$$\dot{m}(t) = Fm(t) \tag{9a}$$

$$\dot{P}(t) = FP(t) + P(t)F^{\mathsf{T}} + GG^{\mathsf{T}} \tag{9b}$$

The normal convention for what is called the EKF is to assume a non linear $f(x,t)$ but at most time varying diffusion $g(t)$, i.e. a non SDD SDE (Maybeck, 1982, Section 9.5), (Jazwinski, 1970, Theorem 8.1), (Grewal and Andrews, 2001, Chapter 5), (Kulikova and Kulikov, 2022). The EKF is obtained by (3) using the following first order Taylor approximation for f and a zero order Taylor approximation for g

$$\begin{aligned}&f(x(t),t) \sim f(m(t),t) + F(x(t) - m(t))\ , \\ &F = \frac{\partial f(m(t),t)}{\partial x}\ ,\ g(x,t) \sim g(m(t),t) = G(m(t),t)\end{aligned} \tag{10}$$

which gives

$$\dot{m}(t) = f(m(t),t) \tag{11a}$$

$$\begin{aligned}\dot{P}(t) =& F(m(t),t)P(t) \\ &+ P(t)F(m(t),t)^{\mathsf{T}} + G(m(t),t)G(m(t),t)^{\mathsf{T}}\end{aligned} \tag{11b}$$

The reason to avoid SDD, is probably that g is already a matrix, so the linearization becomes complicated.

If instead a scalar SDD SDE is assumed, and a first order Taylor linearization is used for both drift f and diffusion g, the following truncated first order EKF (TFOEKF) is obtained

$$\dot{m}(t) = f(m(t)) \tag{12a}$$

$$\begin{aligned}\dot{P}(t) =& 2F(m(t),t)P(t) + G(m(t),t)^2 \\ &+ \left(\frac{\partial g(m)}{\partial x}\right)^2 P(t)\end{aligned} \tag{12b}$$

There are different assumptions and corresponding versions of HOEKFs. Here the truncated second order filter (TSOEKF) by Maybeck (1982, Section 12.3) is used. This is also the choice made by Baadsgaard et al. (1997).

For the TSOEKF, the Taylor approximation of both f and g is expanded to second order, and moments above second order are neglected in (3). For the scalar case this gives (Maybeck, 1982, Section 12.3)

$$\dot{m}(t) = f(m(t),t) + \tfrac{1}{2}\frac{\partial^2 f(m(t))}{\partial x^2}P(t) \tag{13a}$$

$$\dot{P}(t) = \left(2\frac{\partial f(m(t))}{\partial x} + \left(\frac{\partial g(m(t))}{\partial x}\right)^2 + g(m(t))\frac{\partial^2 g(m(t))}{\partial x^2}\right)P(t) + g(m(t),t)^2 \tag{13b}$$

5. THE UKF APPROACH

The motivation for the UKF approach start with the EM discretization (4), and the convergence results (5). Introduce the function ϕ by

$$\begin{aligned}\phi(\varkappa(t_i),\Delta w(t_i)) =& \varkappa(t_i) + f(\varkappa(t_i),t_i)(t_{i+1}-t_i) \\ &+ g(\varkappa(t_i),t_i)\Delta w(t_i)\end{aligned} \tag{14}$$

then the EM discretization can be written

$$\varkappa(t_{i+1}) = \phi(\varkappa(t_i),\Delta w(t_i)) \tag{15}$$

where the dependence on t_{i+1} and t_i has been omitted. Applying the mean and covariance in (15), gives the below recursion.

$$\mathrm{E}(\varkappa(t_{i+1})) = \mathrm{E}(\phi(\varkappa(t_i),\Delta w(t_i)))\ , \tag{16a}$$

$$\mathrm{Cov}(\varkappa(t_{i+1})) = \mathrm{Cov}(\phi(\varkappa(t_i),\Delta w(t_i))) \tag{16b}$$

Again, there is no analytical solution to this. However, if there was, the mean and covariance for $\varkappa$ would converge to the mean and covariance for x as $\Delta t \to 0$ according to (5b).

The point of introducing the function ϕ (14), is the recursion (16), but also to emphasize that the right hand side involves a function of two stochastic variables whose second order description is known for $\Delta w(t_i)$, and calculated at the previous step for $\varkappa(t_i)$.

The unscented transform is generally a good approximator for mean and covariance of non linear functions. The UKF approach suggested here is simply to approximate the right hand side of (16) using the unscented transform. In other words, the unscented transform is used for the EM discretization of the SDD SDE. To the authors knowledge this is not found in the present literature.

6. MEAN AND COVARIANCE TIME EVOLUTION BY EXAMPLES

This section investigates the methods ability to calculate the long term time evolution of the state mean and covariance given initial values. To access and compare the different approaches, the correct solution must be known. One way to obtain the solution would be to simulate the SDE many times using EM and calculate the average and covariance estimates. It is however easier, and more precise, to use examples with known solutions.

One way to obtain this, is to use a transformation of a linear process. Here the most simple linear process, the Ornstein-Uhlenbeck (OU) process (17a) is used.

$$dx = axdt + sdw \Rightarrow \tag{17a}$$

$$\dot{m}_x(t) = am_x(t) \Rightarrow m_x(t) = m_x(0)e^{at} \tag{17b}$$

$$\dot{p}_x(t) = 2ap_x(t) + s^2 \Rightarrow p_x(t) = e^{2at}\left(p_x(0) + \frac{s^2}{2a}\right) - \frac{s^2}{2a} \tag{17c}$$

6.1 Squared OU

A non central Chi squared distributed process is obtained by $z = x^2$ where x is given by (17a).

$$z = x^2 \Rightarrow \tag{18a}$$

$$\mathrm{E}(z(t)) = m_x(t)^2 + p_x(t) \tag{18b}$$

$$\mathrm{V}(z(t)) = 2p_x(t)(p_x(t) + 2m_x(t)^2) \tag{18c}$$

The SDD SDE for z is derived from the IDR (6) to

$$dz = (2az + s^2)dt + 2\sqrt{z}sdw \tag{19}$$

The Lamperti transform will be

$$x(z) = \int \frac{1}{2\sqrt{z}s}dz = \frac{\sqrt{z}}{s} \tag{20}$$

Notice that the s part is only a matter of scaling.

The SDE for x in (20) can be calculated by the IDR to

$$\begin{aligned}dx =& \left(\tfrac{1}{2}z^{-\frac{1}{2}}s^{-1}(2az+s^2) - \frac{1}{8}z^{-\frac{3}{2}}s^{-1}4zs^2\right)dt \\ &+ \tfrac{1}{2}z^{-\frac{1}{2}}s^{-1}2\sqrt{z}sdw \\ =& axdt + dw\end{aligned} \tag{21}$$

If for example s is to be estimated it should be avoided in the Lamperti transform which then could be

$$x(z) = \int \frac{1}{2\sqrt{z}}dz = \sqrt{z} \tag{22}$$

which would give

$$dx = axdt + sdw \tag{23}$$

This is exactly the OU process (17).

The relation (18) going from mean and covariance for x to z can be inverted to give (24) where only the squared mean for x is obtained.

$$p_x(t_0) = \mathrm{E}(z(t_0)) - \sqrt{\mathrm{E}(z(t_0))^2 - \tfrac{1}{2}\mathrm{V}(z(t_0))} \tag{24a}$$

$$m_x(t_0)^2 = \sqrt{\mathrm{E}(z(t_0))^2 - \tfrac{1}{2}\mathrm{V}(z(t_0))} \tag{24b}$$

The Lamperti approach will then start by inserting initial values $\mathrm{E}(z(t_0))$, $\mathrm{V}(z(t_0))$ into (24) to get values for x, and then the exact KF algorithm can calculate the time evolution of the mean for x, which then can be transformed into mean and covariance for z by (18). Notice that the result of this is unchanged, whether the positive or negative solution for $m_x(t_0)$ from (24b) is used.

For the truncated filtering approaches, the first and second order derivatives for f and g are needed. They are therefore listed below.

$$f(z) = 2az + s^2 \Rightarrow \frac{\partial f(z)}{\partial z} = 2a\ ,\ \frac{\partial^2 f(z)}{\partial z^2} = 0\ , \tag{25}$$

$$g(z) = 2\sqrt{z}s \Rightarrow \frac{\partial g(z)}{\partial z} = sz^{-\frac{1}{2}}\ ,\ \frac{\partial^2 g(z)}{\partial z^2} = -\tfrac{1}{2}sz^{-\frac{3}{2}} \tag{26}$$

This leads to the following MODE for EKF, FOEKF and TSOEKF respectively

$$\dot{m}(t) = 2am(t) + s^2 \,, \tag{27a}$$
$$\dot{P}(t) = 4aP(t) + 4s^2 m(t) \tag{27b}$$

$$\dot{m}(t) = 2am(t) + s^2 \,, \tag{28a}$$
$$\dot{P}(t) = (4a + s^2 m(t)^{-1})P(t) + 4s^2 m(t) \tag{28b}$$

$$\dot{m}(t) = 2am(t) + s^2 \,, \tag{29a}$$
$$\dot{P}(t) = 4aP(t) + 4s^2 m(t) \tag{29b}$$

In this squared OU example, all the mean value MODEs are the same. Also, the exact mean value is obtained if started in the correct initial value i.e. $m(0) = m_x(0)^2 + p_x(0)$ where the sub script x refers to the OU process in (17). The variance is the same for EKF and TSOEKF, and for these it is easily verified that the asymptotic variance obtains the correct value ${}^{s^4}/_{2a^2}$. The FOEKF turns out to have a asymptotic variance which is twice as large as the correct value i.e. ${}^{s^4}/_{a^2}$.

It is surprising that the EKF and TSOEKF MODE is exactly the same. The mean value part are equal because the second derivative $\frac{\partial^2 f(z)}{\partial z^2}$ in (13a) is zero, and the variance part are equal because the g parts in (13b) cancel out each other. As seen in section 6.2 this is not the case in general.

Mean and variance for the process resulting from the UKF method (section 5) and the MODE methods EKF, TFOEKF and TSOEKF (27)-(29) are shown in the top of figure 1. Notice that only the UKF mean and variance and the variance for TFOEKF is not exactly correct, and the legends for the correct results are hidden behind the last plotted which is TSOEKF. The bottom figure shows the relative error, which for the UKF method are less than 20% within the time constant at 10 seconds. Results, not shown here, reveals that the relative error is the same, even though the variance on the OU process increases to at least 10^2.

6.2 Exponential OU

A log normal distributed process is obtained by $z = e^x$

$$z = e^x \Rightarrow \tag{30a}$$
$$\mathrm{E}(z) = e^{m_x(t) + \frac{p_x(t)}{2}} \tag{30b}$$
$$\mathrm{V}(z) = (e^{p_x(t)} - 1)\,\mathrm{E}(z)^2 \tag{30c}$$

The SDD SDE for z is derived from the IDR (6) to

$$dz = z\left(a\log(z) + \tfrac{1}{2}s^2\right)dt + zs\,dw \tag{31}$$

The Lamperti transform will be

$$x(z) = \int \frac{1}{zs} dz = s^{-1}\log(z) \tag{32}$$

As for the first example the s can be removed, which results in x having the original OU SDE (17).

The necessary derivatives of f and g are

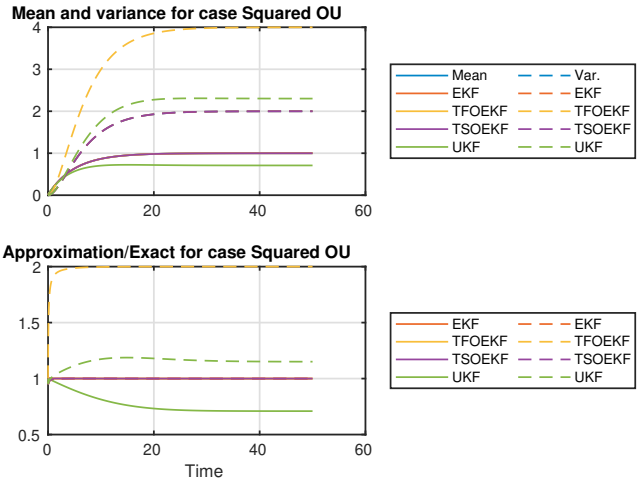

Fig. 1. Mean and variance calculated by the methods compared to the correct values for the squared OU process. The OU system (17) has the parameters $a = -1/\tau$, $\tau = 10$, $\sigma_x^2 = -s^2/2a = 1$, $m_x(0) = 0$, $\sigma_x^2(0) = 0.01$.

$$f(z) = z\left(a\log(z) + \tfrac{1}{2}s^2\right) \Rightarrow$$
$$\frac{\partial f(z)}{\partial z} = \left(a\log(z) + \tfrac{1}{2}s^2\right) + a \tag{33}$$
$$= a(\log(z) + 1) + \tfrac{1}{2}s^2 \,, \quad \frac{\partial^2 f(z)}{\partial z^2} = \frac{a}{z}$$
$$g(z) = zs \Rightarrow \frac{\partial g(z)}{\partial z} = s \,, \quad \frac{\partial^2 g(z)}{\partial z^2} = 0 \tag{34}$$

This leads to the following EKF, FOEKF and TSOEKF respectively

$$\dot{m}(t) = m(t)\left(a\log(m(t)) + \tfrac{1}{2}s^2\right) \,, \tag{35a}$$
$$\dot{P}(t) = (2a(\log(m(t)) + 1) + s^2)P(t) + m(t)^2 s^2 \tag{35b}$$

$$\dot{m}(t) = m(t)\left(a\log(m(t)) + \tfrac{1}{2}s^2\right) \,, \tag{36a}$$
$$\dot{P}(t) = 2(a(\log(m(t)) + 1) + s^2)P(t) + m(t)^2 s^2 \tag{36b}$$

$$\dot{m}(t) = m(t)\left(a\log(m(t)) + \tfrac{1}{2}s^2\right) + \tfrac{1}{2}\frac{a}{m(t)}P(t) \,, \tag{37a}$$
$$\dot{P}(t) = 2(a(\log(m(t)) + 1) + s^2)P(t) + m(t)^2 s^2 \tag{37b}$$

The asymptotic mean and variance for the exponential OU process are

$$m_z = e^{\frac{p_x}{2}} = e^{-\frac{s^2}{4a}} \,, \tag{38a}$$
$$p_z = (e^{p_x} - 1)m_z^2 = (e^{-\frac{s^2}{2a}} - 1)e^{-\frac{s^2}{2a}} \tag{38b}$$

Some results on the asymptotic values for the MODE can be derived, but are left out here to save space.

Graphical results for the four different methods, similar to figure 1, are seen in figure 2. The TFOEKF and The TSOEKF diverges, the former slower than the latter, where the mean even goes negative shortly after the time constant. The EKF and the UKF converges, and the latter has the least error.

The above example based analysis of the methods performance for doing mean and covariance calculations on the "long run" gives the following main results:

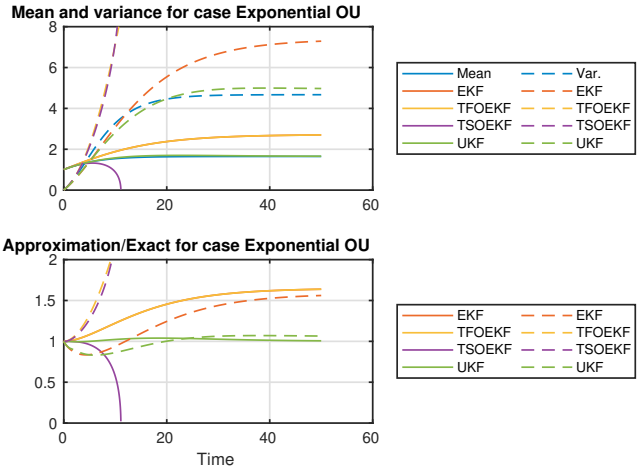

Fig. 2. Mean and variance calculated by the methods compared to the correct values for the exponential OU process. The corresponding OU system (17) has the parameters seen in figure 1. Notice that the y-axis is truncated at 8 and 2 respectively.

- By construction of the examples, the Lamperti method works perfectly.
- For the squared OU process, the EKF and TSOEKF are superior, as they are correct, but the relative errors for the UKF is modest.
- For the exponential OU process, the UKF are superior, as it works for large variances and have modest relative errors. The EKF works better than the TFOEKF and TSOEKF.
- From these two examples the safe choice is the UKF.

7. STATE ESTIMATION BY EXAMPLES

The state mean and covariance time update methods in section 3, 4 and 5, can be used for the time update step, and combined with a measurement update step, they form a collection of state estimation methods. These methods will be compared based on a number of performance measures.

To define the performance measures, the following variables are needed.

$$y_n \triangleq y(t_n)\ ,\ x_n \triangleq x(t_n)\ ,\ Y_n \triangleq (y_n\ y_{n-1}\ \cdots\ y_1) \tag{39a}$$

$$\hat{y}_{n|n-1} \triangleq \widehat{\mathrm{E}}(y_n|Y_{n-1})\ ,\ \tilde{y}_{n|n-1} \triangleq y_n - \hat{y}_{n|n-1}\ ,\ P_{y,n} \triangleq \widehat{\mathrm{V}}(\tilde{y}_{n|n-1}) \tag{39b}$$

$$\hat{x}_{n|n-1} \triangleq \widehat{\mathrm{E}}(x_n|Y_{n-1})\ ,\ \tilde{x}_{n|n-1} \triangleq x_n - \hat{x}_{n|n-1}\ ,\ P_{n|n-1} \triangleq \widehat{\mathrm{V}}(\tilde{x}_{n|n-1}) \tag{39c}$$

$$\hat{x}_{n|n} \triangleq \widehat{\mathrm{E}}(x_n|Y_n)\ ,\ \tilde{x}_{n|n} \triangleq x_n - \hat{x}_{n|n}\ ,\ P_{n|n} \triangleq \widehat{\mathrm{V}}(\tilde{x}_{n|n}) \tag{39d}$$

In (39) $\widehat{\mathrm{E}}, \widehat{\mathrm{V}}$ are the mean and variance estimates provided by the estimation algorithms.

The following performance measures are used for the evaluation of methods:

RY : RMS for $\tilde{y}_{n|n-1}$, which should be small.

RYN : RMS for $\tilde{y}_{n|n-1}/\sqrt{P_{y,n}}$, which should be close to one, indicating that $P_{y,n}$ matches the uncertainty for the method.

PV : Portmanteau test p value, which should be in the "likely" range e.g. 0.05 to 0.95.

RXM : RMS for $\tilde{x}_{n|n-1}$, which should be small.

RXMN : RMS for $\tilde{x}_{n|n-1}/\sqrt{P_{n|n-1}}$, which should be close to one, indicating that $P_{n|n-1}$ matches the uncertainty for the method.

RXP : RMS for $\tilde{x}_{n|n}$, which should be small.

RXPN : RMS for $\tilde{x}_{n|n}/\sqrt{P_{n|n}}$, which should be close to one, indicating that $P_{n|n-1}$ matches the uncertainty for the method.

NAFF : Flag indicating the algorithm had a run error, and could not finish normally due to invalid operations as $\sqrt{z}$ or $\log z$, where the state z has turned negative. This can happen in the time update part of the algorithm.

The state estimation is split in time and measurement update (Maybeck, 1982, p225-227), (Madsen, 2008, p293). The time update is explained above in section 3-5. All the approaches, except UKF, uses integration of the MODE equations. For this the Runge-Kutta 4th order (RK4) method is used. As shown in section 6, some of the methods can produce non positive state covariance and mean, which will give a run time error, and a premature program termination which is caught by setting a flag (Not-all-finite-flag (NAAF)), and then proceeding to the next simulation run.

For non Lamperti methods, the (transformed) state z is estimated, while for the Lamperti method the state estimated is x i.e. the OU process state. To be comparable with the other methods, this state mean and variance estimate is transformed to mean and variance estimate of z using the proper transformation.

For the general non linear estimation problem, a non linear measurement model is allowed

$$y(n) = h(x(n)) + v(n)\ ,\ v(n) \sim \mathrm{IID}(\underline{0}, R) \tag{40}$$

where IID means independent identical distributed i.e. white noise.

The measurement model assumed here is the simple additive noise model

$$y(n) = z(n) + v(n) \tag{41}$$

where z is the squared (19) or exponential (31) OU process.

The TSOEKF has extra terms in the measurement update which does not come into play with the chosen examples as it disappears for $\frac{\partial^2 h}{\partial z^2} = 0$ (Maybeck, 1982, eqn. (12-28)-(12-30)). With this identity measurement function, the non Lamperti methods has virtually equivalent measurement update steps (see e.g. the interpretation in Knudsen and Leth (2019, sec. V)). The Lamperti method will give the measurement models

$$y(n) = x(n)^2 + v(n)\ , \tag{42}$$

$$y(n) = e^{x(n)} + v(n) \tag{43}$$

for the squared and exponential OU process respectively. Recall that x is the OU process. The convention for the Lamperti method, is to use the EKF for the transformed problem, so this is what is done here.

The following methods are tested:

Lamp-EKF : The EKF is used on the Lamperti transformed model.

EKF : The EKF (11) is used on the original model.
TFOEKF : The TFOEKF (12) is used on the original model.
TSOEKF : The TSOEKF (13) is used on the original model.
UKF : The UKF (sec. 5 and Knudsen and Leth (2019)) is used on the original model.

7.1 Squared OU

The methods are compared by generating simulated data with the same parameters as in figure 1. From each simulation of 100 seconds, the signals are logged with a sampling time of 1. Based on this, the performance measures mentioned above is calculated. This is done for 100 repetitions, which leads to the average statistics seen in table 1.

Table 1. Estimation performance measures for the methods used on the squared OU process. The numbers are averages over 100 runs of each 100 seconds.

Method	RY	RYN	PV	RXM	RXMN	RXP	RXPN	NAFF
LampEKF	1.12	4.32	0.67	1.14	0.95	0.95	3.65	0
EKF	0.77	1.04	0.62	0.76	1.08	0.1	1.04	0
TFOEKF	1.48	NA	0	1.48	NA	1.45	NA	1
TSOEKF	1.47	NA	0	1.46	NA	1.43	NA	1
UKF	0.77	1.03	0.62	0.76	1.07	0.1	1.04	0

Before discussing statistic results from table 1, selected time series plots from the first run is presented, and differences between the methods are highlighted. Figure 3 shows the OU process and its square plus the measurements.

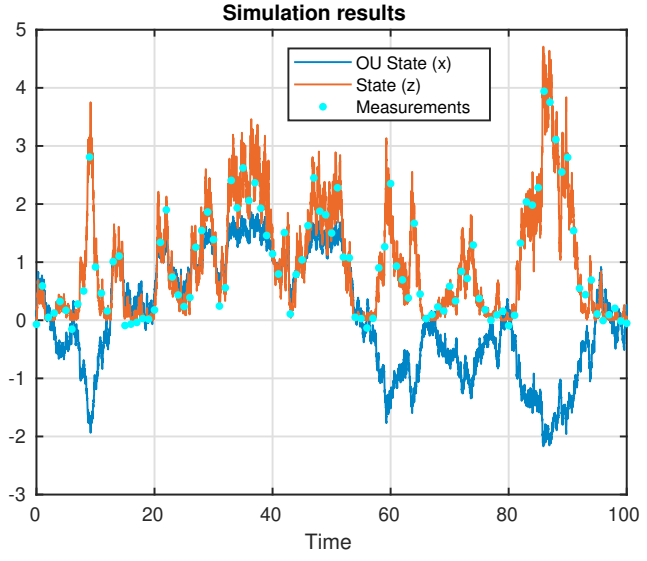

Fig. 3. Simulated OU process (x), squared process (z) and measurements. Notice that the states are shown with the simulation sample time step 0.01 as a connected line, whereas the measurements are shown with the measurement sampling time step 1 as dots.

Figure 4 shows the results for the Lamp-EKF algorithm. Recall that here the Lamperti transform is $\sqrt{z}$ with the result, that the SDE for z is turned into the OU process, and the measurement equations is $y = x^2 + v$. This means the sign of x can not be determined. This is why the top sub figure shows that the state prediction confidence interval for the OU state (x) fits well with the real state except around time 10, where the real and predicted states has opposite signs. The second sub figure shows that the Kalman gain varies a lot due to the non linear measurement equation. The third sub figure shows that the state prediction std. does not vary so much, except for some bumps where the state prediction variance increases, because the state prediction is close to zero, so the measurement provides less information about the state. The bottom sub figure shows that the corresponding squared state (z) prediction confidence interval fits well with the real z, and the uncertainty varies a lot due to the square transformation.

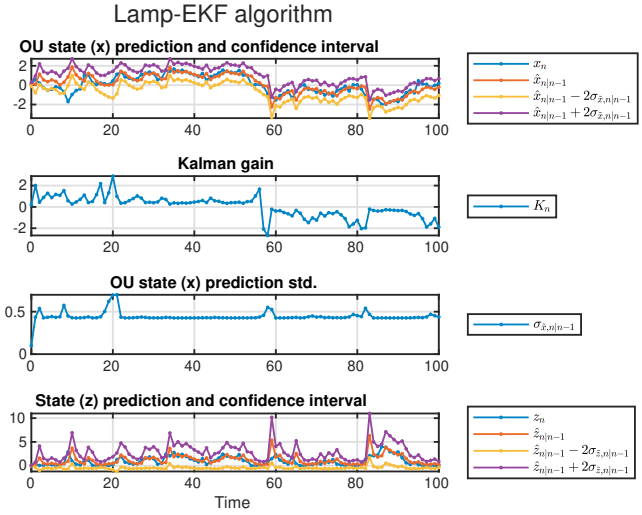

Fig. 4. Results from using the Lamp-EKF algorithm on the data in figure 3. Notise that the data are really in discrete time which is indicated with a dot on the line connecting the samples for visability.

Figure 5 shows the results from using the TSOEKF algorithm on the data in figure 3. The bottom sub figure shows the state prediction variance goes negative, which makes the algorithm stop prematurely, leaving the rest of the values at the zero initial values.

Figure 6 shows the results from using the UKF algorithm on the data in figure 3. The top sub figure shows the state prediction confidence interval is similar to the one for the Lamp-EKF algorithm in figure 4. However, there are some time point e.g. a little before 60 and after 80 where the performance of the UKF is superior. The middle sub figure shows a Kalman gain mostly close to one, due to the measurement equation, except when the state prediction variance is very small due to a small state prediction, then the Kalman gain drops a bit. However, the state prediction variance changes a lot due to the nonlinear and SDD SDE. It is very clear that the state prediction std. follows the predicted state level. For this algorithm also the convincing auto correlation plot, for the output prediction errors, are shown in the bottom sub figure.

To save space, results from the two last algorithm are not shown.

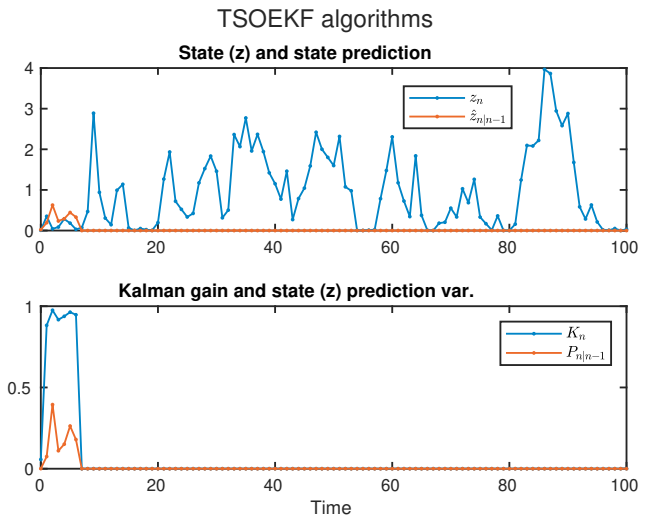

Fig. 5. Results from using the TSOEKF algorithm on the data in figure 3.

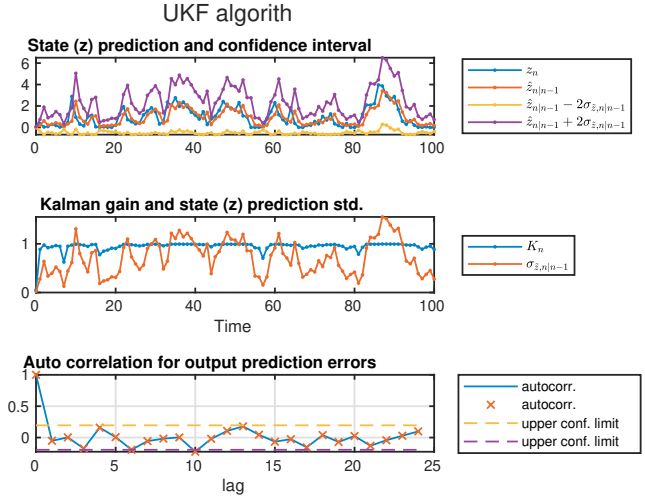

Fig. 6. Results from using the UKF algorithm on the data in figure 3.

Returning to the statistical results, many can be extracted from table 1. EKF and UKF works best and very similar. Also they perform similar to a KF for a linear system. The normalized errors for output and state has a RMS close to one, showing that the variance estimates are very good. The p-value for the Portmanteau test Madsen (2008) is also very convincing. The other methods are clearly inferior. TFOEKF and TSOEKF ends up with negative state variance estimates when the state estimate becomes small. This happens in all the runs, and results in the Not-all-finite-flag fraction being 1. Notice that all state estimates and corresponding variances are set to zero before each run. This is why the normalized RMS errors can not be calculated in case the method does not finish. Also in this case the remaining output predictions is zero, which leads to large average RMS for the output error. Perhaps the most important for TFOEKF and TSOEKF is not the RMS, but that they stopped prematurely due to negative variance estimates. The Lamp-EKF is underestimating the variance on $\tilde{y}_{n|n-1}$ by approximately a factor 4. This is probably because the gradient of the measurement function x_n^2 i.e. $2x_n$, gives to small variance for $\tilde{y}_{n|n-1}$ when x_n is close to 0. The problem that the Lamperti method can transform a otherwise linear measurement, and/or state, model into a non linear one, is also mentioned in the conclusion of Nielsen and Madsen (2001) and Baadsgaard et al. (1997).

7.2 Exponential OU

Again to save space, only plots of the simulated signals and the results for UKF is shown. The simulated signals are seen in figure 7. Notice the OU process (x) is the same as in figure 3, as the default random number start is used in MATLAB. In this case, small OU states (x) gives close to zero measurements, where then the measurement noise has relatively more impact. Figure 8 shows the result of

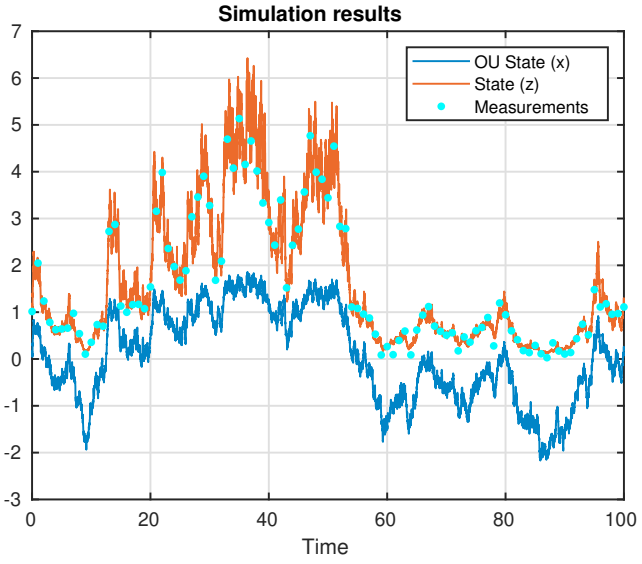

Fig. 7. Simulated OU process (x), exponential process (z) and measurements. This plot is corresponding to figure 3 in the square process case.

the UKF algorithm used for the data in figure 7. Again the performance seems good with white output prediction errors. As for the squared case, the Kalman gain is close to one, except when the state prediction error variance is very small due to state predictions being close to zero.

Results corresponding to table 1, is shown in table 2 for the exponential OU state process. EKF, TFOEKF and UKF works best and very similar. Also in this case they perform similar to a KF for a linear system. The other methods are clearly inferior. TSOEKF ends up with negative state variance estimates when the state estimate becomes small. This happens not all the time, as in the squared OU case, but in 12% of the runs. Also in this case the Lamp-EKF is underestimating the variance on $\tilde{y}_{n|n-1}$, but this problem is less severe compared to the squared OU case.

8. CONCLUSION

This paper discusses stochastic differential equations with state dependent diffusion. Specifically, methods for approximating the state second order statistics, and finally estimating the state from measurements are analyzed.

The paper includes a method that exploits the unscented transform to approximate the time evolution of mean

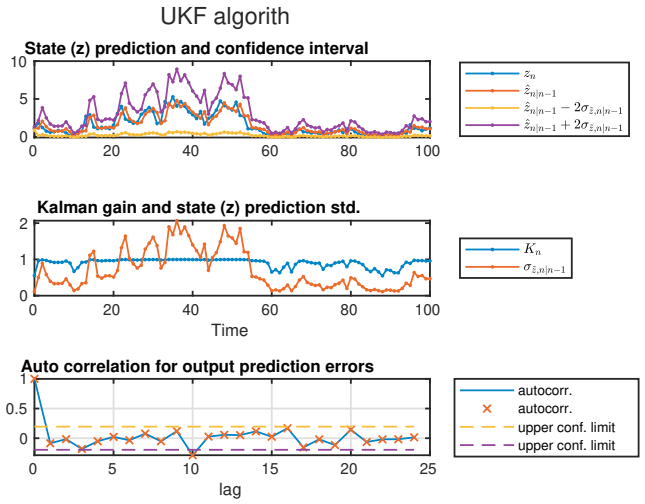

Fig. 8. Results from using the UKF algorithm on the data in figure 7.

Table 2. Estimation performance measures for the methods used on the exponential OU process. The numbers are averages over 100 runs of each 100 seconds.

Method	RY	RYN	PV	RXM	RXMN	RXP	RXPN	NAFF
Lamp-EKF	1.15	1.48	0.61	1.23	0.96	0.96	4.25	0
EKF	1	1.03	0.53	0.99	1.03	0.1	1	0
TFOEKF	1	0.98	0.53	0.99	0.98	0.1	0.99	0
TSOEKF	1.05	NA	0.49	1.04	NA	0.18	NA	0.12
UKF	1	1.03	0.53	0.99	1.03	0.1	1	0

and covariance for state dependent diffusion stochastic differential equations. To the authors best knowledge, this is a new contribution.

The paper starts by presenting the methods. This is followed by analyzing the methods ability to calculate long term development of mean and covariance from initial state values. Because the two examples are transformations of the linear OU process, the Lamperti method gives exactly the correct results. For the squared OU process, it was surprising that the EKF and TSOEKF methods gave correct results. Here the UKF method gave relative errors below 30%. For the exponential OE process, the UKF method was clearly superior, with a relative error below 20%. The EKF method gave much larger errors, the TFOEKF gave a variance that grew unlimited, and the TSOEKF gave a variance going negative after about the time constant.

The last part of the paper analyses the performance estimating states using measurements. The results can be compared to the scarce literature. Baadsgaard et al. (1997); Nielsen and Madsen (2001); Møller and Madsen (2010) states that EKF can have poor performance for SDD SDE but does not give examples. The results obtained here can not support this. On the contrary, The superior methods for the two examples are UKF and EKF, which are very similar, and surprisingly have features found for linear systems, as white prediction errors and correct error variance estimates. Møller and Madsen (2010) states that HOEKF can have numerical issues. This comply with the TSOEKF filter crashes due to negative variance in both examples. Baadsgaard et al. (1997); Nielsen and Madsen (2001) compares only the Lamperti and TSOEKF approach for one SDE model, and conclude the performance is similar. Here the Lamperti method has a significantly lower performance compared to UKF and EKF. The overall conclusion from these two examples, is that the UKF method is preferable as it in both examples are among the best methods, and it does not need any derivatives or transformations and works for any state dimension and diffusion function.

ACKNOWLEDGEMENTS

Thanks to my colleague John Leth for valuable suggestions for improvements.

REFERENCES

Åström, K.J. (1970). *Introduction to Stochastic Control Theory*, volume 70 of *Mathematics in science and engineering*. Academic Press.

Baadsgaard, M., Nielsen, J.N., Spliid, H., Madsen, H., and Preisel, M. (1997). Estimation in stochastic differential equations with a state dependent diffusion term. *IFAC Proceedings Volumes*, 30(11), 1369 – 1374. doi: https://doi.org/10.1016/S1474-6670(17)43033-3. IFAC Symposium on System Identification (SYSID'97), Kitakyushu, Fukuoka, Japan, 8-11 July 1997.

Grewal, M.S. and Andrews, A.P. (2001). *Kalman Filtering*. Wiley, Hoboken, New Jersey, second edition.

Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory*, volume 64 of *Mathematics in science and engineering*. Academic Press, Inc.

Kloeden, P.E. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Stochastic Modelling and Applied Probability*. Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-662-12616-5.

Knudsen, T. and Leth, J. (2019). A new continuous discrete unscented kalman filter. *IEEE Transactions on Automatic Control*, 64(5), 2198–2205. doi: 10.1109/TAC.2018.2867325.

Kulikova, M.V. and Kulikov, G.Y. (2022). Continuous-discrete unscented kalman filtering framework by matlab ode solvers and square-root methods. *Automatica*, 142, 110396. doi: https://doi.org/10.1016/j.automatica.2022.110396.

Madsen, H. (2008). *Time Series Analysis*. Chapman & Hall. ISBN: 978-1-4200-5967-0.

Maybeck, P.S. (1982). *Stochastic Models, Estimation, and Control*, volume 1-3 of *Mathematics in science and engineering*. Academic Press, Inc. Is republished by Navtech http://www.navtechgps.com/supply/books.asp.

Møller, J.K. and Madsen, H. (2010). From state dependent diffusion to constant diffusion in stochastic differential equations by the lamperti transform. Technical Report IMM-Technical Report-2010-16, Technical University of Denmark, DTU Informatics.

Nielsen, J.N. and Madsen, H. (2001). Applying the ekf to stochastic differential equations with level effects. *Automatica*, 37(1), 107–112.