Continuous-Discrete Bayesian Filtering for Stochastic Differential Equations

Aniket Das[1]

***Abstract*— This report presents a survey on the Bayesian Filtering problem for continuous-discrete State Space Models, where the state is a continuous-time Markov Process , defined by a Stochastic Differential Equation, while the observation model, representing noisy measurements of the state, is discrete time, defined only at the sampling intervals, which are no longer necessarily uniformly spaced. We elucidate the advantages of such continuous-discrete modelling over the discrete time State Space Models traditionally encountered in Bayesian Filtering, present an intuitive exposition to Stochastic Differential Equations, motivating them as continuous-time Markov Processes and as noise-driven Dynamical Systems, and then discuss the key algorithms for performing Bayesian Filtering in both linear and nonlinear continuous-discrete State Space Models. These algorithms can be intuitively regarded as continuous-time extensions of the well known Bayesian Filtering algorithms (like Kalman Filtering, EKF, CKF) for discrete time models. Lastly, we simulate these algorithms on various well known linear and nonlinear SDE models and compare them with appropriate baselines. The code for the simulations is available at `https://github.com/Aniket1998/SDE-Filtering`**

I. INTRODUCTION

Stochastic Differential Equations (SDEs) are a mathematical tool for describing the dynamics of Markov Processes that occur in continuous time. Apart from being of paramount importance in the theory of stochastic processes, they find immense practical applications in areas such as physics, applied mathematics, machine learning and mathematical finance, among many others. The time evolution of several real world stochastic phenomena can be modelled using Stochastic Differential Equations. In fact, many popular State Space Models encountered in the Bayesian Filtering literature [3] are actually discrete time approximations of continuous time SDE based State Space Models. Hence, a framework for solving the Bayesian Filtering problem for SDE based State Space Models is of both theoretical and practical significance. To this end, this report investigates a special case of the Bayesian Filtering problem in SDEs, known as Continuous-Discrete Bayesian Filtering, where the state transition model is governed by a Stochastic Differential Equation, while the observation model is discrete time and is a function of the current state at discrete, but not necessarily uniform, time intervals. Such a framework is highly practical for a wide range of problem settings that involve one or more sensors (which are often digital and hence, can only operate in discrete time) taking noisy samples or measurements of a continuous time phenomenon (for instance, ambient temperature sensors). Adapting such a Continuous-Discrete framework confers several advantages over the traditional purely discrete approaches. Most notably, by using sophisticated SDE discretization methods, the Continuous-Discrete approach allows one to significantly control the approximation error that the purely discrete framework incurs due to approximating the continuous time phenomenon using a discrete-time Markov Chain. Secondly, Continuous Discrete Bayesian Filtering is *discretization invariant*, i.e., unlike its purely discrete counterpart, it does not require the discrete observations to be made at uniformly spaced time-intervals. The property of discretization invariance is immensely useful for multi-sensor applications, where uniformly spaced samples are generally not available due to the lack of synchronization between the sensors.

A. Contributions

This report makes the following technical contributions

1) It provides an intuitive exposition to Stochastic Differential Equations, motivating them both as the continuous-time limit of Markov Chains, and as white-noise driven Ordinary Differential Equations
2) It introduces the basics of SDE theory such as the Itô Integral and the Fokker Planck Kolmogorov equation and the Euler-Murayama discretization for sampling SDE trajectories
3) It sets up the theoretical framework of the Continuous-Discrete Bayesian Filtering problem for SDE based State Space Models, and discusses the theory behind both exact and approximate Bayesian Filtering algorithms such as the Continuous Discrete Kalman Filter, Continuous Discrete Extended Kalman Filter and the Continuous Discrete Sigma Point Kalman Filter
4) It simulates the abovementioned algorithms on well known SDE models and makes appropriate comparisons with the respective baselines. The source code is publicly available at `https://github.com/Aniket1998/SDE-Filtering`

II. MATHEMATICAL PRELIMINARIES

A. Stochastic Differential Equations

Stochastic Differential Equations (SDEs) are a probabilistic extension of Ordinary Differential Equations where one or more of the driving terms is a time dependent random variable (more rigorously, a stochastic process). As a result, the solutions to SDEs are themselves time dependent random variables as opposed to the familiar deterministic trajectories

[1]Incoming Spring Semester Exchange Student, Department of Computer Science, Aalto University. Student No. 843500 aniket.das@aalto.fi

(or time dependent deterministic variables) obtained by solving ODEs.

The existence of SDEs can either be motivated as a stochastic extension of dynamical systems, by considering SDEs to be describing the time evolution of systems driven by noise, or as a continuous-time limit of a certain class Markov Processes, both of which are discussed as follows

SDEs as Continous-Time Markov Processes: Consider the familiar Markov Process transition equation for a discrete-time stochastic process (or a sequence of random variables) $\{X_k\}_{k=1}^T$

$$X_{k+1} = g(X_k) + L(X_k)w_k \quad w_k \sim \mathcal{N}(0, W_k)$$

Considering the $\{X_k\}$ to be realisations of this Markov Process at time intervals $\delta_t, 2\delta_t$ and so on, we can rewrite it as follows

$$X_{t+\delta_t} = g(X_t) + L(X_t)w_t \quad w_t \sim \mathcal{N}(0, W(t))$$

Before defining the continuous time limit of this Markov Process, we make the following transformations

$$X_{t+\delta_t} - X_t = h(X_t) + L(X_t)w_t \;\; w_t \sim \mathcal{N}(0, W(t)) \; h(X_t) = g(X_t) - X_t$$

Now, we define $h(X_t) = f(x_t)\ \delta_t$ and $W(t) = Q(t)\ \delta_t$, writing the Markov Process transition equation as

$$X_{t+\delta_t} - X_t = f(X_t)\delta_t + L(X_t)w_t \quad w_t \sim \mathcal{N}(0, Q(t)\delta_t)$$

Finally, to obtain the continuous-time limit, we replace δ_t with the time differential dt, $X_{t+\delta_t} - X_t$ with the *stochastic differential* dX_t as follows

$$dX_t = f(X_t)\ dt + L(x_t)\ dW_t \tag{1}$$

Equation 1 is a general (Homogeneous) Stochastic Differential Equation. The term dW_t is a new mathematical object known as the differential of a Wiener Process W_t with dispersion matrix Q, which is a stochastic process defined by the following properties

- **Zero Initial Value** $W_0 = 0$
- **Gaussian Increments** $W_{t+\delta_t} - W_t \sim \mathcal{N}(0, Q\ \delta_t)$
- **Independent Increments** Non-overlapping increments are independent. In other words, for $t_1 < t_2 < t_3 < t_4$, $W_{t_4 - t_3}$ is independent of $W_{t_2 - t_1}$
- **Continuity in Paths** Any realisation of this stochastic process $\{W_t\}_t$ is continuous in t
- **Existence of the Differential** The differential increment dW_t exists and $dW_t \sim \mathcal{N}(0, Q\ dt)$, by taking the infinitesimal limit of the Gaussian Increments property

Due to historical reasons, W_t is also known as *Mathematical Brownian Motion* and is a central object in the theoretical analysis of SDEs. One can also generalise Equation to represent non-homogeneous continuous time Markov Processes as follows

$$dX_t = f(X_t, t)\ dt + L(x_t, t)\ dW_t \tag{2}$$

The term $f(X_t, t)$ is known as *drift* and the term $L(x_t, t)$ is known as *diffusion* or *dispersion*

SDEs as Stochastic Extensions of ODEs: As any ODE of the form $\frac{dx}{dt} = f(x, t)$ can be written in terms of the (deterministic) differential dx as $dx = f(x, t)dt$, it seems natural to rewrite SDEs of the form 2 as

$$\frac{d}{dt}X_t = f(X_t, t) + \frac{dW_t}{dt}$$

As the differential increment of the Wiener Process dW_t is well defined, we define the (Malliavin) derivative of the Wiener Process $n(t) = \frac{dW_t}{dt}$ as *white noise*, which allows us to write SDEs in an ODE-like representation as follows

$$\frac{dX_t}{dt} = f(X_t, t) + L(x_t, t)\ n(t) \tag{3}$$

We refer to Equation 3 as the *Stochastic ODE* representation of an SDE.

We refer the readers to Särkkä and Solin (2019) [1] for a more detailed coverage of Stochastic Differential Equations and Øksendal (2003) [2] for a rigorous treatment.

B. Itô and Stratonovich Stochastic Calculi

We attempt to explicitly solve the SDE in Equation 2 by integrating the stochastic differential as follows.

$$X_{T_1} - X_{T_0} = \int_{T_0}^{T_1} f(X_t, t)dt + \int_{T_0}^{T_1} L(X_t, t)dW_t$$

There are two distinct terms in this integral equation. The first term $\int_{T_0}^{T_1} f(X_t, t)dt$ is the traditional Riemann integral which is well defined. However, the second term $\int_{T_0}^{T_1} L(X_t, t)dW_t$ is an integral with respect to the Wiener Process, which can neither be defined as a Riemann Integral nor as a Lebesgue Integral. There exist two distinct definitions of the integral with respect to a Wiener Process known as the Itô Calculus and Stratonovich Calculus, which define the integral as follows

- **Itô Calculus**
$$\int_{T_0}^{T_1} L(X_t, t)dW_t = \lim_{n\to\infty} \sum_{k=1}^{n-1} L(X_{t_k}, t_k)[W_{t_{k+1}} - W_{t_k}]$$
where $t_0 < t_1 < ... < t_n \in [T_0, T_1]$
- **Stratonovich Calculus**
$$\int_{T_0}^{T_1} L(X_t, t)dW_t = \lim_{n\to\infty} \sum_{k=1}^{n-1} L(X_{\frac{t_k + t_{k+1}}{2}}, \frac{t_k + t_{k+1}}{2})[W_{t_{k+1}} - W_{t_k}]$$
where $t_0 < t_1 < ... < t_n \in [T_0, T_1]$

Under appropriate regularity conditions, both of these integral definitions are convergent (in probability). We use the Itô Calculus framework of SDEs in this report as theoretical analysis in this framework is much easier than in the Stratonovich framework. However, an Itô SDE can be easily converted into a Stratonovich SDE and vice versa. The readers are encouraged to look at [1] and [2] for further details regarding Stratonovich calculus and the inter-conversion process

C. Fokker Planck Kolmogorov Equation

The Fokker Planck Kolmogorov Equation is a Partial Differential Equation for the probability density $p(X_t)$ of a Markov Process defined by a Stochastic Differential Equation, also denoted as $p(x,t)$. For an SDE $dX_t = f(X_t,t)\ dt + L(x_t,t)\ dW_t$ where the dispersion matrix of W_t is Q, and the distribution of the initial state is $p(X_0)$, the Fokker Planck Kolmogorov Equation is as follows.

$$\frac{\partial}{\partial t}p(x,t) = -\sum_{i=1}^{D}\frac{\partial}{\partial x_i}[f_i(x,t)p(x,t)] +$$

$$\frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D}\frac{\partial^2}{\partial x_i \partial x_j}[[L(x,t)QL^T(x,t)]_{ij}p(x,t)]$$

subject to the boundary condition $p(x,0) = p(X_0)$. By choosing the appropriate boundary conditions, Fokker Planck Equation can also be used to solve for transition densities $p(X_t|X_s)$ of Markov Processes as well as the Chapman Kolmogorov Equation. In particular, they are solved as follows

- **Transition Density** $p(X_t|X_s = y)$: Solve Fokker Planck Kolmogorov with the boundary condition $p(x,0) = \delta(x-y)$
- **Chapman Kolmogorov** $\int p(X_t|X_s)p(X_s|y)dX_s$: Solve Fokker Planck Kolmogorov with the boundary condition $p(x,s) = p(X_s = x|y)$

The above solution to the Chapman Kolmogorov equation for SDEs is fundamental to the framework of Continuous-Discrete Bayesian Filtering.

From the Fokker Planck Kolmogorov Equation, one can derive the following ODEs for the mean $m(t) = \mathbb{E}[X_t]$ and the covariance $P(t) = \mathrm{Cov}[\mathrm{X_t}]$ of a Markov Process as follows

$$\frac{dm}{dt} = \mathbb{E}[f(X_t,t)]$$

$$\frac{dP}{dt} = \mathbb{E}[f(X_t,t)(X_t - m)^T] + \mathbb{E}[f(X_t,t)^T(X_t - m)]$$
$$+\mathbb{E}[(L(X_t,t)QL(X_t,t)^T]$$

In general, these equations are not solvable in closed form, as they involve calculating an expectation with respect to the state distribution, which itself is given by the Fokker Planck Kolmogorov Equation. However, one notable and important exception is Linear SDEs of the form

$$dX_t = F(t)X_t dt + L(t)dW_t$$

In this case, the mean and covariance ODEs are simplified as

$$\frac{dm}{dt} = \mathbb{E}F(t)X_t = F(t)\mathbb{E}[X_t] = F(t)m$$

Simplifying similarly,

$$\frac{dP}{dt} = F(t)P + PF(t)^T + L(t)QL(t)^T$$

The simplified ODEs of the mean and covariance, coupled with the fact that, for a Gaussian distributed initial state, all subsequent states of a Linear SDE are also Gaussian distributed(Refer to [1] and [2] for the proof), allows for explicit solutions of Linear SDEs, which in turns, makes the exact Continuous Discrete Bayesian Filtering Problem feasible for Linear SDEs

III. EXACT CONTINUOUS-DISCRETE KALMAN FILTERING

A. Continuous-Discrete Bayesian Filtering

The key property of the State Space Models encountered in the continuous-discrete Bayesian Filtering Framework, is that the dynamics of the state X_t is now a continuous Markov Process defined by an SDE. The observation, or emission model y_k, however, remains discrete time, and depends on the realisation of the state at discrete time steps t_k. The overall model is as follows

$$dX_t = f(X_t,t)\ dt + L(x_t,t)\ dW_t$$

$$y_k \sim p(y_k|X_{t_k})$$

The objective of Continuous Discrete Bayesian Filtering is to infer the marginal posterior distribution $p(X_{t_k}|y_{1:k})$, also known as the filtering distribution. Analogous to traditional Discrete Time Bayesian Filtering, this can be performed recursively, using the Chapman Kolmogorov Equation and Bayes Theorem as follows, provided the filtering distribution of the previous step $p(X_{t_{k-1}}|y_{1:k-1})$ is known

$$p(X_{t_k}|y_{1:k-1}) = \int p(X_{t_k}|X_{t_{k-1}})p(X_{t_{k-1}}|y_{1:k-1})dX_{t_{k-1}}$$

$$p(X_{t_k}|y_{1:k}) = \frac{p(y_k|X_{t_k})p(X_{t_k}|y_{1:k-1})}{\int p(y_k|X_{t_k})p(X_{t_k}|y_{1:k-1})dX_{t_k}}$$

The key difference from a traditional Discrete Bayesian Filter lies in the fact that the transition distribution of the Markov Process $p(X_{t_k}|X_{t_{k-1}})$ is not readily available in closed form for the Continuous Discrete Bayesian Filter. Moreover, the time evolution of the transition density $p(X_{t_k}|X_{t_{k-1}})$ is given by the Fokker Planck Kolmogorov Equation. Hence, given an initial distribution $p(X_{t_0})$, the density $p(X_{t_k}|y_{1:k-1})$ can be computed at each timestep by solving the Fokker Planck Kolmogorov Equation using the boundary condition $p(x,t_{k-1}) = p(X_{t_{k-1}}|y_{1:k-1})$ and then setting $p(X_{t_k}|y_{1:k-1}) = p(x,t_k)$. This leads to the following algorithm for the Continuous Discrete Bayesian Filter.

Algorithm 1 Continuous Discrete Bayesian Filter

Input: $p(X_{t_0}), \{y_k\}_{k=1}^{T}$

for `k=1 ... T` **do**

  Prediction Step

    Compute $p(X_{t_k}|y_{1:k-1}) = p(x,t_k)$ by solving Fokker Planck Kolmogorov equation using the boundary condition

    $p(x,t_{k-1}) = p(X_{t_{k-1}}|y_{1:k-1})$

  Update Step

    $$p(X_{t_k}|y_{1:k}) = \frac{p(y_k|X_{t_k})p(X_{t_k}|y_{1:k-1})}{\int p(y_k|X_{t_k})p(X_{t_k}|y_{1:k-1})dX_{t_k}}$$

end for

Note that the above algorithm does not require the time intervals $t_k - t_{k-1}$ to be constant. This property is known as *discretization invariance* and is extremely useful in multi-sensor systems where measurement / sampling intervals are usually non-uniform as the sensors are often not synchronised.

B. Continuous-Discrete Kalman Filter for Linear SDEs

We consider the case of Continuous Discrete State Space Models governed by a Linear SDE and a Linear Gaussian Observation Model as follows

$$dX_t = F(t)X_t dt + L(t)dW_t$$

$$y_k = H_k X_{t_k} + r_k \quad r_k \sim \mathcal{N}(0, R_k)$$

Let Q be the diffusion matrix of the Wiener Process W_t. It is known that if the initial distribution $p(X_{t_0})$ is a Gaussian, i.e. $p(X_{t_0}) = \mathcal{N}(m_0, P_0)$, then the transition density $p(X_{t_k}|X_{t_{k-1}})$ is also Gaussian. The reader is requested to refer to [1] and [2] for the proof. Consequently, prediction distribution $p(X_{t_k}|y_{1:k-1})$ and the filtering distribution $p(X_{t_k}|y_{1:k})$ are also Gaussian due to the observation model $p(y_k|X_{t_k}) = \mathcal{N}(H(t)X_{t_k}, R_k)$. As a Gaussian Distribution is uniquely determined by its Mean and Covariance, we can use the mean and covariance ODEs of Linear Time Invariant SDEs, namely $\frac{dm}{dt} = F(t)m$ and $\frac{dP}{dt} = F(t)P + PF(t)^T + L(t)QL(t)^T$, with the boundary conditions $m(t_{k-1}) = m_{k-1}$ and $P(t_{k-1}) = P_{k-1}$ to solve for the prediction distribution $p(X_{t_k}|y_{1:k-1}) = \mathcal{N}(m_k^-, P_k^-)$ where $m_k^- = m(t_k)$ and $P_k^- = P(t_k)$. The subsequent update equation is the same as that of the Discrete Kalman Filter as the observation model is discrete. This gives rise to the following algorithm known as the Continuous Discrete Kalman Filter.

Algorithm 2 Continuous Discrete Kalman Filter

Input: $m_0, P_0, \{y_k\}_{k=1}^{T}$

for `k=1 ... T` **do**

  Prediction Step

    Solve for $m(t_k)$ and $P(t_k)$

    $\frac{dm}{dt} = F(t)m \qquad m(t_{k-1}) = m_{k-1}$

    $\frac{dp}{dt} = F(t)P + PF(t)^T + L(t)QL(t)^T \qquad P(t_{k-1}) = P_{k-1}$

    $m_k^- = m(t_k)$

    $P_k^- = P(t_k)$

  Update Step

    $v_k = y_k - H_k m_k^-$

    $S_k = H_k P_k^- H_k^T + R_k$

    $K_k = P_k^- H_k^T S_k^{-1}$

    $m_k = m_k^- + K_k v_k$

    $P_k = P_k^- - K_k S_k K_k^T$

end for

IV. APPROXIMATE CONTINUOUS-DISCRETE KALMAN FILTERING

We now turn our attention to non-linear Continuous Time State Space Models, where the transition model is no longer a Linear Time Invariant SDE and the observation model, although still discrete time and Gaussian, is now a non-linear function of the state, as follows

$$dX_t = f(X_t, t)\ dt + L(x_t, t)\ dW_t$$

$$y_k = h(X_{t_k}) + r_k \quad r_k \sim \mathcal{N}(0, R_k)$$

In this case, neither the transition density nor the filtering distribution is necessarily Gaussian. Moreover, the Fokker Planck Equation for the transitoon density However, for a Gaussian initial state distribution $p(X_{t_0}) = \mathcal{N}(m_0, P_0)$, one can consider a Gaussian approximation for the prediction and filtering distributions as follows

$$p(X_{t_k}|y_{1:k-1}) \approx \mathcal{N}(m_k^-, P_k^-)$$

$$p(X_{t_k}|y_{1:k-1}) \approx \mathcal{N}(m_k, P_k)$$

Estimating the parameters m_k, P_k is known as Approximate Continuous-Discrete Kalman Filtering. The traditional discrete time approaches to Approximate Kalman Filtering can be extended to perform Approximate Continuous-Discrete Kalman Filtering by approximating the mean and covariance ODEs obtained from the Fokker Planck Kolmogorov equations

$$\frac{dm}{dt} = \mathbb{E}[f(X_t, t)]$$

$$\frac{dP}{dt} = \mathbb{E}[(f(X_t,t) - m)(X_t - m)^T] + \mathbb{E}[(f(X_t,t) - m)^T (X_t - m)] + \mathbb{E}[(L(X_t,t)QL(X_t,t)^T]$$

While various approaches for Approximate Continuous-Discrete Kalman Filtering exist, in this report, we investigate the Extended Continuous Discrete Kalman Filter and the Sigma-Point Continuous Discrete Kalman Filter.

A. Continuous-Discrete Extended Kalman Filter

In a manner similar to the discrete time Extended Kalman Filter, the Extended Continuous Discrete Kalman Filter solves the approximate Bayesian Filtering problem by taking a First Order Taylor approximation of the drift term $f(X_t,t)$ around the mean m as follows.

$$f(X_t,t) = f(m) + F_x(m,t)(X_t - m)$$

where $F_x(m)$ denotes the gradient of $f(X_t,t)$ at m. This leads to the following approximated ODEs for the mean and covariance

$$\frac{dm}{dt} = f(m,t)$$

$$\frac{dP}{dt} = F_x(m,t)P + PF_x(m,t)^T + L(m,t)QL(m,t)^T$$

The RHS terms of these ODEs are independent of the state distribution and can be explicitly evaluated by closed form or numerical integration. The linearisation causes the filtering distribution to be (approximately) Gaussian, assuming a Gaussian initial state distribution. Moreover, the update step is the same as that in a traditional Extended Kalman Filter as the observation model is discrete. This results in the following algorithm for the Continuous Discrete Extended Kalman Filter

Algorithm 3 Continuous-Discrete Extended Kalman Filter

Input: $m_0, P_0, \{y_k\}_{k=1}^T$
 for `k=1 ... T` **do**
  Prediction Step
   Solve for $m(t_k)$ and $P(t_k)$
   $\frac{dm}{dt} = f(m,t)$ $m(t_{k-1}) = m_{k-1}$
   $\frac{dp}{dt} = F_x(m,t)P + PF_x(m,t)^T + L(m,t)QL(m,t)^T$
   $P(t_{k-1}) = P_{k-1}$
   $m_k^- = m(t_k)$
   $P_k^- = P(t_k)$
  Update Step
   $v_k = y_k - h_k(m_k^-)$
   $S_k = [H_k]_x(m_k^-, P_k^-)P_k^-[[H_k]_x(m_k^-, P_k^-)]^T + R_k$
   $K_k = P_k^-[[H_k]_x(m_k^-, P_k^-)]^T S_k^{-1}$
   $m_k = m_k^- + K_k v_k$
   $P_k = P_k^- - K_k S_k K_k^T$
 end for

B. Continuous Discrete Sigma Point Kalman Filter

The Continuous Discrete Sigma Point Kalman Filter belong to the general family of Continuous Discrete Gaussian Assumed density filters, which assume that the state distribution $p(X_t)$ at any moment in time is Gaussian, i.e., $p(X_t) = \mathcal{N}(m,P)$ where m and P are time variable. Using the time evolution of m and P obtained from the Fokker Planck Kolmogorov equation, we can write the following.

$$\frac{dm}{dt} = \mathbb{E}[f(x,t)] = \int f(x,t)\mathcal{N}(x|m,P)dx$$

$$\frac{dP}{dt} = \int f(x,t)(x-m)^T\mathcal{N}(x|m,P)dx$$

$$+\int (x-m)f^T(x,t)\mathcal{N}(x|m,P)dx$$

$$+\int L(x,t)QL^T(x,t)\mathcal{N}(x|m,P)dx$$

Although the above standard Gaussian integrals with respect to $\mathcal{N}(x|m,P)$ may not be generally solvable in closed form, one can closely approximate them by means of Sigma Point Methods such as the Unscented Transform, Gauss Hermite Quadrature or Cubature Integration as follows

$$\int f(x,t)\mathcal{N}(x|m,P) \approx \sum_i W^{(i)} f(m+\sqrt{P}\zeta_i,t)$$

The approximation approach taken is analogous to that used in classical discrete time Gaussian Assumed Kalman Filters, such as the Unscented Kalman Filter and the Cubature Kalman Filter. Substituting the Sigma Point approximation results in the following ODEs for m and P

$$\frac{dm}{dt} = \sum_i W^{(i)} f(m+\sqrt{P}\zeta_i,t)$$

$$\frac{dP}{dt} = \sum_i W^{(i)} f(m+\sqrt{P}\zeta_i,t)\zeta_i^T\sqrt{P}^T$$

$$+\sum_i W^{(i)}\sqrt{P}\zeta_i f^T(m+\sqrt{P}\zeta_i,t)$$

$$+\sum_i W^{(i)} L(m+\sqrt{P}\zeta_i,t)QL^T(m+\sqrt{P}\zeta_i,t)$$

Numerical solution of the above ODEs consitutes the Prediction Step of the Sigma Point Kalman Filter. Due to the observation model being discrete, the update equations are the same as that of the traditional discrete time Sigma Point Kalman Filter. This results in the following algorithm for the Continuous Discrete Sigma Point Kalman Filter.

The simulations in this report use the Cubature Approximation for the Gaussian Integrals, which involve the following choice of sigma points nad their respective weights

$$W^{(i)} = \frac{1}{2n} \qquad i = 1\ldots 2n$$

$$\zeta_i = \begin{cases} \sqrt{n}e_i & i = 1\ldots n \\ -\sqrt{n}e_{i-n} & i = n+1\ldots 2n \end{cases}$$

Algorithm 4 Sigma-Point Continuous Discrete Kalman Filter

Input: $m_0, P_0, \{y_k\}_{k=1}^{T}$

for `k=1 ... T` **do**

  Prediction Step

    Solve for $m(t_k)$ and $P(t_k)$

    $\frac{dm}{dt} = \sum_i W^{(i)} f(m + \sqrt{P}\zeta_i, t)$

    $\frac{dp}{dt} = \sum_i W^{(i)} f(m + \sqrt{P}\zeta_i, t)\zeta_i^T \sqrt{P}^T$

    $+ \sum_i W^{(i)} \sqrt{P}\zeta_i f^T(m + \sqrt{P}\zeta_i, t)$

    $+ \sum_i W^{(i)} L(m + \sqrt{P}\zeta_i, t) Q L^T(m + \sqrt{P}\zeta_i, t)$

    $m_k^- = m(t_k)$

    $P_k^- = P(t_k)$

  Update Step

    $\chi_k^{(i)} = m_k^- + \sqrt{P_k^-}\,\zeta^{(i)}$

    $\hat{Y}_k^{(i)} = h(\chi_k^{(i)})$

    $\mu_k = \sum_i W_m^{(i)} \hat{Y}_k^{(i)}$

    $S_k = \sum W_i^{(c)} (\hat{Y}_k^{(i)} - \mu_k)(\hat{Y}_k^{(i)} - \mu_k)^T + R_k$

    $C_k = \sum W_i^{(c)} (\chi_k^{(i)} - m_k^-)(\hat{Y}_k^{(i)} - \mu_k)^T$

    $K_k = C_k S_k^{-1}$

    $m_k = m_k^- + K_k(y_k - \mu_k)$

    $P_k = P_k^- - K_k S_k K_k^T$

end for

V. EXPERIMENTS: LINEAR STATE SPACE MODELS

We simulate Continuous Discrete Kalman Filtering on two Linear State Space Models, the Ornstein Uhlenbeck Process and the Stochastic Forced Harmonic Oscillator. State trajectories are sampled for a time span of 10 seconds, with a time interval of 0.01 seconds. Samples from the SDEs are drawn using the Euler Murayama [4] method with a step size of 0.001 seconds (10 times finer than the time interval of the state space) and the prediction step ODEs are numerically integrated using an Euler Integrator of step size 0.001 seconds.

For each state space model, we calculate the Root Mean Square Error between the actual states and the Kalman Filter means, and report the mean and the standard deviation of the RMSE across 10 independent runs. We compare this to the baseline, which involves computing the RMSE between the noisy observations. As observed in Table 1, the Kalman Filter achieves significantly better results than the baseline. The code for these simulations is available at `https://github.com/Aniket1998/SDE-Filtering`

A. Ornstein Uhlenbeck Process

We use the following State Space Model for the Ornstein Uhlenbeck Process, where the emission model simply constitutes noisy observations of the state.

$$dX_t = -\lambda X_t \; dt + \sigma_l \; dW_t$$

$$y_k = X_{t_k} + r_k \qquad r_k \sim \mathcal{N}(0, \sigma_r^2)$$

The prediction step of the Kalman Filter constitutes the following ODEs

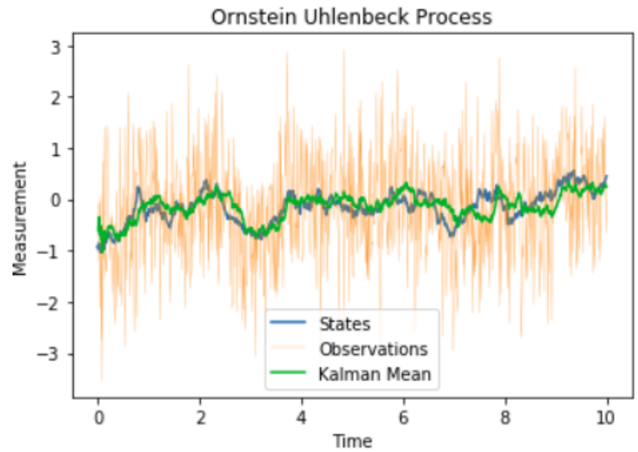

Fig. 1. Kalman Filtering for Ornstein Uhlenbeck Process

$$\frac{dm}{dt} = -\lambda m$$

$$\frac{dP}{dt} = -2\lambda P + (\sigma_q \;\; \sigma_l)^2$$

For the simulations, we set $\lambda = 1$, $\sigma_l = 0.5$ $\sigma_r = 1$ and use a univariate Wiener Process with dispersion $\sigma_q = 1$ Across 10 independent runs, the Kalman Filter achieves an RMSE of 0.04 ± 0.01 compared to the baseline which achieves an RMSE of 1.00 ± 0.04. Figure 1 depicts one such simulation. The code is available at `https://github.com/Aniket1998/SDE-Filtering/blob/master/OrnsteinUhlenbeck.ipynb`

B. Stochastic Forced Harmonic Oscillator

We simulate the Stochastic Forced Harmonic Oscillator which has a second order Stochastic ODE representation as a damped harmonic oscillator being driven by white noise

$$\frac{d^2x}{dt^2} = -\omega^2 x - b\frac{dx}{dt} + n(t)$$

Analogous to the standard approach for deterministic second order ODEs, we introduce the state variable Z as follows, and subsequently rewrite the system in stochastic differential form as an SDE in Z

$$Z = \begin{bmatrix} x \\ \frac{dx}{dt} \end{bmatrix}$$

$$dZ_t = FZ_t dt + LdW_t$$

where

$$F = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -b \end{bmatrix}$$

$$L = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

The observation model constitutes noisy measurements of the position x as follows.

$$y_k = HZ_{t_k} + r_k \qquad r_k \sim \mathcal{N}(0, \sigma_r^2)$$

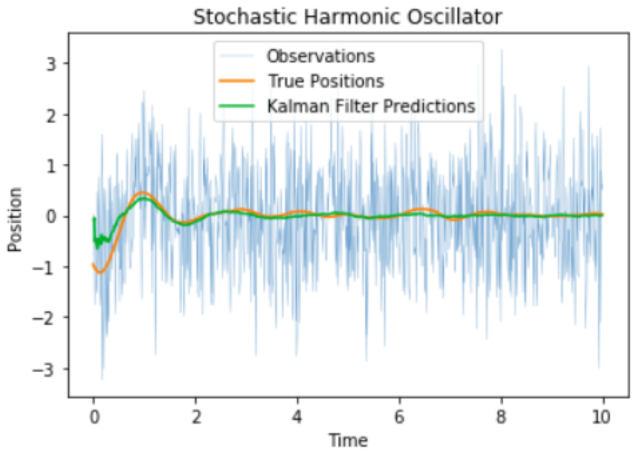

Fig. 2. Kalman Filtering for Stochastic Forced Harmonic Oscillator

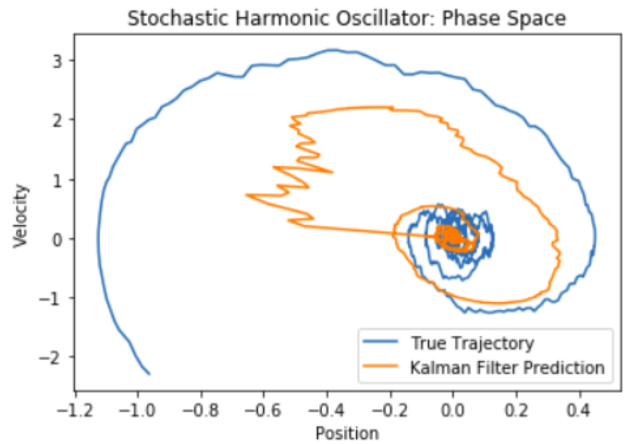

Fig. 3. Phase Space of Stochastic Forced Harmonic Oscillator

where

$$H = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

The prediction step of the Kalman Filter then constitutes the following ODEs

$$\frac{dm}{dt} = Fm$$

$$\frac{dP}{dt} = FP + PF^T + Q$$

where

$$Q = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_q^2 \end{bmatrix}$$

For the simulations, we set $\omega = 4$, $b = 2$ $\sigma_r = 1$ and use a univariate Wiener Process with dispersion $\sigma_q = 0.5$ Across 10 independent runs, the Kalman Filter achieves an RMSE of 0.009 ± 0.003 compared to the baseline which achieves an RMSE of 0.98 ± 0.03. Figure 2 depicts one such simulation and Figure 3 depicts the corresponding phase space representation. The code is available at `https://github.com/Aniket1998/SDE-Filtering/blob/master/StochasticSHM.ipynb`

TABLE I
RMSE FOR LINEAR MODELS

Algorithm	Ornstein-Uhlenbeck	Stochastic SHM
Baseline	1.00 ± 0.04	0.98 ± 0.03
Kalman Filter	**0.04 ± 0.01**	**0.009 ± 0.003**

VI. EXPERIMENTS: NONLINEAR STATE SPACE MODELS

We simulate Continuous Discrete Extended Kalman Filtering and Continuous Discrete Cubature Kalman Filtering on three Nonlinear State Space Models, the Benes Daum Model, the Cox Ingersoll Ross Model and the Stochastic Duffing Van der Pol Oscillator. Similar to the simulations for the linear models, state trajectories are sampled for a time span of 10 seconds, with a time interval of 0.01 seconds. Samples from the SDEs are drawn using the Euler Murayama [4] method with a step size of 0.001 seconds (10 times finer than the time interval of the state space) and the prediction step ODEs are numerically integrated using an Euler Integrator of step size 0.001 seconds.

For each state space model, we calculate the Root Mean Square Error between the actual states and the Extended and Cubature Kalman Filter means, and report the mean and the standard deviation of the RMSE across 10 independent runs. We compare this to the baseline, which involves computing the RMSE between the noisy observations. As observed in Table 2, the Extended and the Cubature Kalman Filters achieve significantly better results than the baseline. Moreover, the Extended and the Cubature Kalman Filters achieve identical results on average. The code for these simulations is available at `https://github.com/Aniket1998/SDE-Filtering`

A. Benes-Daum Model

We use the following State Space Model for the Benes Daum Model, where the emission model simply constitutes noisy observations of the state.

$$dX_t = \mathrm{tanh(X_t)\ dt + dW_t}$$

$$y_k = X_{t_k} + r_k \qquad r_k \sim \mathcal{N}(0, \sigma_r^2)$$

The prediction step ODEs for the Extended Kalman Filter are as follows

$$\frac{dm}{dt} = \mathrm{tanh(m)}$$

$$\frac{dP}{dt} = \mathrm{2(1 - tanh^2(m))P + \sigma_q^2}$$

The following are the prediction step ODEs for the Cubature Kalman Filter

$$\frac{dm}{dt} = \frac{1}{2}[\tanh(m + \sqrt{P}) + \tanh(m - \sqrt{P})]$$

$$\frac{dP}{dt} = \sqrt{P}[\tanh(m + \sqrt{P}) - \tanh(m - \sqrt{P})]$$

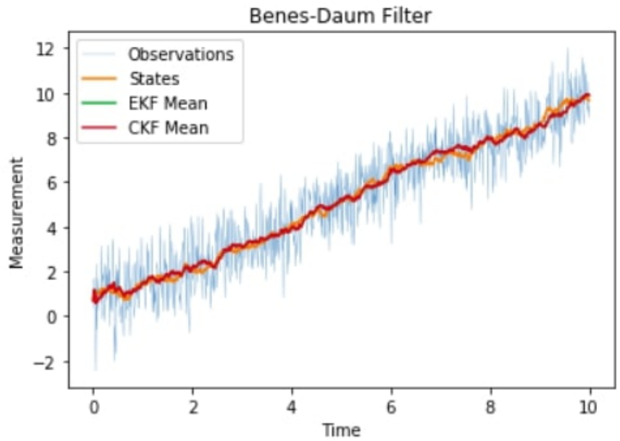

Fig. 4. Approximate Kalman Filtering for Benes-Daum Model

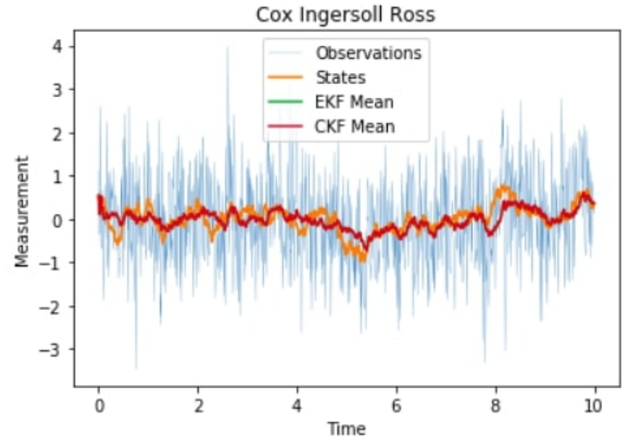

Fig. 5. Approximate Kalman Filtering for Cox-Ingersoll-Ross Model

$$+(\theta_2\sigma_q)^2[(m+\sqrt{P})^2+(m-\sqrt{P})^2]$$

For the simulations, we set $\sigma_r = 1$ and use a univariate Wiener Process with dispersion $\sigma_q = 0.5$ Across 10 independent runs, the Extended Kalman Filter and the Cubature Kalman Filter achieve an RMSE of 0.05 ± 0.01 compared to the baseline which achieves an RMSE of 1.02 ± 0.03. Figure 4 depicts one such simulation. The code is available at `https://github.com/Aniket1998/SDE-Filtering/blob/master/Benes.ipynb`

B. Cox Ingersoll Ross Model

We use the following State Space Model for the Cox Ingersoll Ross Process, where the emission model simply constitutes noisy observations of the state.

$$dX_t = -\theta_1 X_t \ dt + \theta_2\sqrt{1+X_t^2} \ dW_t$$

$$y_k = X_{t_k} + r_k \qquad r_k \sim \mathcal{N}(0, \sigma_r^2)$$

The prediction step ODEs for the Extended Kalman Filter are as follows

$$\frac{dm}{dt} = -\theta_1 m$$

$$\frac{dP}{dt} = -2\theta_1 P + (\theta_2 \ \sigma_q)^2(1+m^2)$$

The following are the prediction step ODEs for the Cubature Kalman Filter

$$\frac{dm}{dt} = -\theta_1 m$$

$$\frac{dP}{dt} = -2\theta_1 P + \frac{1}{2}(\theta_2 \ \sigma_q)^2[2+(m+\sqrt{P})^2+(m-\sqrt{P})^2]$$

For the simulations, we set $\theta_1 = 2$, $\theta_2 = 3$, $\sigma_r = 1$ and use a univariate Wiener Process with dispersion $\sigma_q = 0.2$ Across 10 independent runs, the Extended Kalman Filter and the Cubature Kalman Filter achieve an RMSE of 0.047 ± 0.007 compared to the baseline which achieves an RMSE of 1.00 ± 0.05. Figure 5 depicts one such simulation. The code is available at `https://github.com/Aniket1998/SDE-Filtering/blob/master/Cox.ipynb`

C. Stochastic Duffing Van der Pol Oscillator

We simulate the Stochastic Duffing Van der Pol Oscillator which has the following second order Stochastic ODE representation

$$\frac{d^2x}{dt^2} = x(\alpha - x^2) - \frac{dx}{dt} + x_1 n(t)$$

We define the following augmented state

$$x_1 = x$$

$$x_2 = \frac{dx}{dt}$$

Which allows us to rewrite the dynamics as the following nonlinear SDE

$$d\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = F(x_1,x_2)dt + L(x_1,x_2)dW_t$$

W_t is a bivariate Wiener Process with the dispersion matrix $Q = \sigma_q^2 I$ The observation model constitutes noisy measurements of the position x_1 as follows.

$$y_k = \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} x_{1,t_k} \\ x_{2,t_k} \end{bmatrix} + r_k \qquad r_k \sim \mathcal{N}(0,\sigma_r^2)$$

We define the drift term $f(x_1,x_2)$ and the diffusion term $L(x_1,x_2)$ as follows

$$f(x_1,x_2) = \begin{bmatrix} x_2 \\ x_1(\alpha - x_1^2) - x_2 \end{bmatrix}$$

$$L(x_1,x_2) = \begin{bmatrix} 0 & 0 \\ x_1 & 0 \end{bmatrix}$$

The Jacobian of the drift term is then

$$F_x(x_1,x_2) = \begin{bmatrix} 0 & 1 \\ \alpha - 3x_1^2 & -1 \end{bmatrix}$$

which allows us to define the following prediction step ODEs for the Extended Kalman Filter

$$\frac{dm}{dt} = F(m_1,m_2)$$

$$\frac{dP}{dt} = F_x(m_1,m_2)P + PF_x(m_1,m_2)^T + \sigma_q^2 L(m_1,m_2)L(m_1,m_2)^T$$

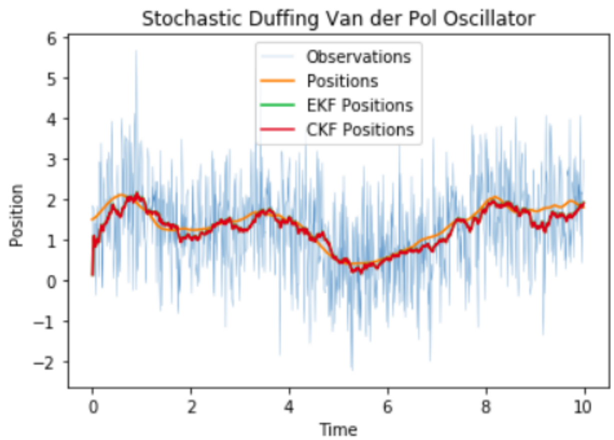

Fig. 6. Approximate Kalman Filtering for Stochastic Duffing Van der Pol Oscillator

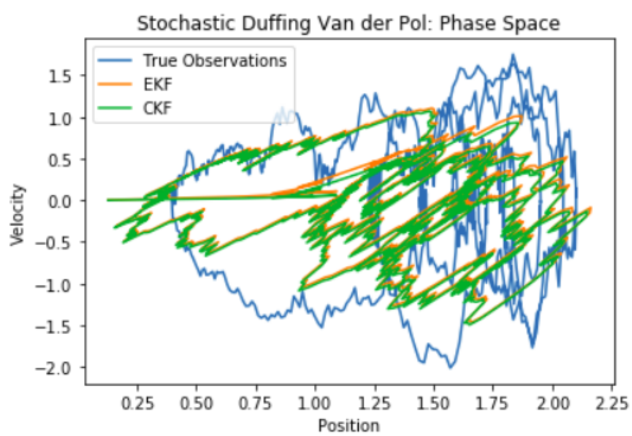

Fig. 7. Phase Space Stochastic Duffing Van der Pol Oscillator

The corresponding prediction step ODEs for the Cubature Kalman Filter are

$$\frac{dm}{dt} = \frac{1}{4}\sum_{i=1}^{4} f([m+\zeta_i]_1, [m+\zeta_i]_2)$$

$$\frac{dP}{dt} = \frac{1}{4}\sum_{i=1}^{4} f([m+\zeta_i]_1, [m+\zeta_i]_2)\zeta_i^T \sqrt{P}^T$$

$$+\frac{1}{4}\sum_{i=1}^{4} \sqrt{P}\zeta_i f^T([m+\zeta_i]_1, [m+\zeta_i]_2)$$

$$+\frac{\sigma_q^2}{4}\sum_{i=1}^{4} L([m+\zeta_i]_1, [m+\zeta_i]_2)L^T([m+\zeta_i]_1, [m+\zeta_i]_2)$$

For the simulations, we set $\alpha = 2$, $\sigma_q = 1.0$ and $\sigma_r = 1$. Across 10 independent runs, the Extended Kalman Filter and Cubature Kalman Filter achieve an RMSE of 0.03 ± 0.01 compared to the baseline which achieves an RMSE of 1.01 ± 0.04. Figure 6 depicts one such simulation and Figure 7 depicts the corresponding phase space representation. The code is available at `https://github.com/Aniket1998/SDE-Filtering/blob/master/Duffing.ipynb`

TABLE II
RMSE FOR NONLINEAR MODELS

Algorithm	Benes-Daum	CIR	Duffing VdP
Baseline	1.02 ± 0.03	1.00 ± 0.05	1.01 ± 0.04
EKF	**0.05 ± 0.01**	**0.047 ± 0.007**	**0.03 ± 0.01**
CKF	**0.05 ± 0.01**	**0.047 ± 0.007**	**0.03 ± 0.01**

VII. CONCLUSION

This report has provided an introductory exposition to the mathematics of Stochastic Differential Equations, presented a theoretical analysis of both exact and approximate continuous discrete Bayesian Filtering, and performed simulations on popular SDE based State Space Models. Possible avenues of extension could be investigating the problem of continuous discrete smoothing, and parameter estimation for SDE based state space models. The integration of approximate continuous discrete Bayesian Filtering with Markov Chain Monte Carlo Methods such as Particle Filters and Rao Blackwellized Particle Filters is also a promising avenue.

REFERENCES

[1] Simo Särkkä and Arno Solin, Applied Stochastic Differential Equations, Cambridge University Press 2019
[2] Bernt Øksendal, Stochastic Differential Equations An Introduction with Applications. New York: Springer-Verlag, 2003.
[3] Simo Särkkä, Bayesian Filtering and Smoothing, Cambridge University Press, 2013.
[4] Peter Kloeden and Eckhard Platen, Numerical Solution of Stochastic Differential Equations, New York: Springer-Verlag, 1992