A Fourier State Space Model for Bayesian ODE Filters

Hans Kersting [1][2][*]   **Maren Mahsereci** [3]

Abstract

Gaussian ODE filtering is a probabilistic numerical method to solve ordinary differential equations (ODEs). It computes a Bayesian posterior over the solution from evaluations of the vector field defining the ODE. Its most popular version, which employs an integrated Brownian motion prior, uses Taylor expansions of the mean to extrapolate forward and has the same convergence rates as classical numerical methods. As the solution of many important ODEs are periodic functions (oscillators), we raise the question whether Fourier expansions can also be brought to bear within the framework of Gaussian ODE filtering. To this end, we construct a Fourier state space model for ODEs and a 'hybrid' model that combines a Taylor (Brownian motion) and Fourier state space model. We show by experiments how the hybrid model might become useful in cheaply predicting until the end of the time domain.

1. Introduction

Ordinary differential equations (ODEs) appear in many machine learning algorithms. In recent years, there has been a particular surge of interest in ODEs for normalizing flows (Rezende & Mohamed, 2015). This development is driven by neural ODEs (Chen et al., 2018), which allow for maximum-likelihood estimation and variational inference. Neural ODEs replace learning by gradient descent with learning by ODE sensitivity analysis (Rackauckas et al., 2018).

A recent recast of ODEs as a stochastic filtering problems has made it possible to solve initial value problems (IVPs) by all available Bayesian filtering methods (Tronarp et al., 2019; 2020). The resulting class of methods, called *ODE fil-*

---
[*]work primarily performed during an internship at Amazon Research, Cambridge, UK. [1]University of Tübingen, Germany [2]Max Planck Institute for Intelligent Systems, Tübingen, Germany [3]Amazon Web Services, Berlin, Germany. Correspondence to: Hans Kersting <hans.kersting@uni-tuebingen.de>.

ters, has not only fulfilled the goal of probabilistic numerics (PN) (Hennig et al., 2015) to quantify numerical uncertainty in a Bayesian way, but has also identified the dynamic model as the fundamental internal modeling assumption of ODE solvers. This dynamic model determines how the solver extrapolates forward in time and is equivalent to a prior over the ODE solution (Kersting et al., 2020, Appendix A).

Early PN research has, to show that its new methods are indeed practical, focused mostly on creating probabilistic analogues of classical methods. This line of inquiry has discovered that the integrated Brownian Motion (IBM) prior gives rise to ODE filters whose mean coincides with standard classical methods; see Schober et al. (2019). This is due to the fact that—like e.g. Runge–Kutta method—ODE filters with the IBM prior use Taylor expansions to locally predict forward; see Section 5 in Kersting et al. (2019). In the meantime, other local expansions based on the Matérn covariance function have been studied; see Tronarp et al. (2020). Fourier expansions, however, have not been investigated in the context of ODE filters although it is known how to incorporate them in a dynamic model; see (Solin & Särkkä, 2014). With this paper, we aim to begin filling in this gap. Since so many important ODEs are oscillators with periodic solutions, we consider Fourier expansions a promising research direction in the context of ODE filtering.

2. ODE Filtering for Initial Value Problems

We consider the following IVP

$$\dot{x}(t) = f\left(x(t)\right), \ \forall t \in [0, T], \qquad x(0) = x_0 \ \in \mathbb{R}^d, \ (1)$$

with vector field $f : \mathbb{R}^d \to \mathbb{R}^d$. For notational convenience, we restrict w.l.o.g. the below presentation to $d = 1$. As the solution $x : [0, T] \to \mathbb{R}^d$ in general lies in an infinite dimensional function space such as $C^2([0, T]; \mathbb{R}^d)$, a finite dimensional representation is needed to extrapolate from $x(t)$ to $x(t + h)$. Runge–Kutta methods, for example, use Taylor expansions (i.e. projections to polynomial spaces) as finite dimensional approximations of x. While classical methods only do so implicitly, Gaussian ODE filtering represents $x(t)$ explicitly in a D-dimensional *state vector*, i.e. in a stochastic process $X(t)$ from which a model of $x(t)$ can be linearly extracted:

$$x(t) \ \sim \ H_0 X(t), \quad \text{for some } H_0 \in \mathbb{R}^{d \times D}. \quad (2)$$

Algorithm 1 Gaussian ODE Filtering

> **Input:** IVP(x_i, m, T), step size $h > 0$
> Initialize, $t = 0$, $H_0 X(0) = x_0$ and $H X(0) = f(x_0)$
> **repeat**
>   **predict** state X, $t \rightarrow t + h$, along Equation (5)
> $t = t + h$
>   **update** $X(t)$ by Equations (8) and (9)
> **until** $t + h > T$

Moreover, for ODEs, the derivative has also to be linearly extractable

$$\dot{x}(t) \sim H X(T) \quad \text{for some } H \in \mathbb{R}^{d \times D}. \tag{3}$$

As ODE filtering is a Bayesian method, the state $X(t)$ is modeled by a stochastic process, which is usually represented by a linear time-invariant stochastic differential equation (SDE)

$$dX(t) = F X(t)\, dt + L\, dB(t) \tag{4}$$

with Gaussian initial condition on $X(0)$, where the drift and diffusion matrices $F, L \in \mathbb{R}^{D \times D}$ detail the deterministic and stochastic part of the dynamics respectively. This SDE prior can be thought of as a localized definition of a Gauss–Markov process. In its discretized form, it defines a *dynamic model*

$$p\left(X(t + h) \mid X(t)\right) = \mathcal{N}(A(h)X(t), Q(h)), \tag{5}$$

with matrices $A(h), Q(h) \in \mathbb{R}^{D \times D}$ which are implied in closed form by F and L. To update this model, a *measurement model* is added

$$p(Z(t) \mid X(t)) = \mathcal{N}\left(f(H_0 X(t)) - H X(t), R\right), \tag{6}$$

$R \geq 0$, which is conditioned on the data

$$Z(t) := 0. \tag{7}$$

As f is non-linear, the above measurement model is intractable (Tronarp et al., 2019, Section 2). By substituting $f(H_0 \mathbb{E}[X(t)])$ for $f(H_0 X(t))$, we obtain the following tractable measurement model

$$p(Z(t) \mid X(t)) = \mathcal{N}\left(H X(t), R\right) \tag{8}$$
$$Z(t) := f(H_0 \mathbb{E}[X(t)]). \tag{9}$$

We will employ this measurement model in this paper. The dynamic and measurement model together are called a *probabilistic state space model* (SSM), which Gaussian ODE filtering uses to infer x as detailed in Algorithm 1.

The classical Taylor SSM

In previous research, the most commonly recommended model uses an integrated Brownian motion as a dynamic

model (prior). It is defined by inserting the following matrices into Equation (5):

$$A(h)_{ij} = \mathbb{I}_{i \leq j} \frac{h^{j-i}}{(j-i)!}, \tag{10}$$

$$Q(h)_{ij} = \sigma^2 \frac{h^{2q+1-i-j}}{(2q+1-i-j)(q-i)!(q-j)!}, \tag{11}$$

where σ^2 is the variance scale of the underlying Brownian motion; see Kersting et al. (2019, Appendix A). Here, the state vector

$$\left[x^{(0)}(t), \dots, x^{(q)}(t)\right] \sim X(t) \tag{12}$$

models the first $q \in \mathbb{N}$ derivatives of $x(t)$. Therefore, the mean prediction $X(t + h) = A(h)X(t)$ is a *Taylor expansion* of the numerical estimates of these derivatives. Consequently, since Runge–Kutta methods also use Taylor approximations of $x(t)$ (Hairer et al., 1987, Section II.2), the posterior mean is similar to Runge–Kutta methods with local convergence rates of $q + 1$ and global convergence rates of q. It is hence a very good method to solve generic ODEs; see Schober et al. (2019) and Kersting et al. (2019).

3. Fourier Models for ODEs

In this paper, we are concerned with oscillating ODEs. Hence, let the ODE be such that its solution $x(t)$ is a periodic function. Let us denote the period of x by $p > 0$, and its angular velocity by $w_0 = 2\pi/p$. For such periodic functions, a J-th order *Fourier series*, $J \in \mathbb{N}$, of x is the standard approximation:

$$x^J(t) = x_0 + \sum_{j=1}^{J} x_j(t) \overset{J \to \infty}{\to} x(t), \tag{13}$$

almost everywhere, where $x_0 = a_0/2$ and

$$x_j(t) = a_j \cos(w_0 j t) + b_j \sin(w_0 j t). \tag{14}$$

The derivative of Equation (13) is

$$\dot{x}^J(t) = y_0 + \sum_{j=1}^{J} -w_0 j y_j(t) \overset{J \to \infty}{\to} x(t), \tag{15}$$

almost everywhere, where $y_0 = 0$ and

$$y_j(t) = a_j \sin(w_0 j t) - b_j \cos(w_0 j t). \tag{16}$$

The exact *Fourier coefficients* are given by the integrals

$$a_j = \frac{2}{p} \int_0^p x(t) \cos(j w_0 t)\, dt, \tag{17}$$

$$b_j = \frac{2}{p} \int_0^p x(t) \sin(j x_0 t)\, dt, \tag{18}$$

which are, in general, intractable. Thus, learning a periodic approximation of x amounts to inferring the coefficients (a_j, b_j). To reproduce the model from Solin & Särkkä (2014, Section 3.2), we will however run inference on the corresponding harmonic oscillators $(x_j(t), y_j(t))$ instead. To this end, we first observe that, for each $j = 0, \ldots, J$, $[x_j(t), y_j(t)]$ satisfies the following differential equations

$$\frac{d}{dt} \begin{bmatrix} x_j(t) \\ y_j(t) \end{bmatrix} = \begin{bmatrix} 0 & -jw_0 \\ jw_0 & 0 \end{bmatrix} \begin{bmatrix} x_j(t) \\ y_j(t) \end{bmatrix}. \quad (19)$$

Note that, if we set $[x_0(0), y_0(0)] = [a_0/2, 0]$ and $[x_j(0), y_j(0)] = [a_j, -b_j]$, then the only solution of Equation (19) is indeed $[x_j(t), y_j(t)]$ as defined in Equations (14) and (16). Since we (in general) do not know the Fourier coefficients (a_j, b_j), we model the initial values with a Gaussian probability distribution: $[x_j(0), y_j(0)] \sim \mathcal{N}(\mathbf{0}, q_j^2 \mathbf{I})$, for some $q_j^2 > 0$. We then model these oscillators by a stochastic process $X(t)$, i.e.

$$[x_0(t), y_0(t), x_1(t), y_1(t), \ldots, x_J(t), y_J(t)] \sim X(t) \quad (20)$$

which (according to Equation (19)) follows the SDE, Equation (4), if and only if

$$F = \text{diag}(F_1, \ldots, F_J), \quad \text{with blocks} \quad (21)$$

$$F_j = \begin{bmatrix} 0 & -jw_0 \\ jw_0 & 0 \end{bmatrix}, \quad \text{and} \quad (22)$$

$$L = \mathbf{0} \in \mathbb{R}^{2(J+1) \times 2(J+1)}, \quad (23)$$

and if the initial condition is $[X(0)_{2j}, X(0)_{2j+1}] \sim \mathcal{N}(\mathbf{0}, q_j^2 \mathbf{I})$. It is natural that there is no diffusion ($L = 0$), as the Fourier coefficients (unlike Taylor coefficients) of a periodic signal $x(t)$ do not change in t. Since we want the prior defined by this SDE to be a zero-mean Gaussian Process with the canonical periodic covariance function

$$k_p(t, t') = \sigma^2 \exp \left(-\frac{2 \sin^2 \left(w_0 \frac{t-t'}{2} \right)}{l^2} \right), \quad (24)$$

we have to set

$$q_j^2 = \frac{2 I_j(l^{-2})}{\exp(l^{-2})}, \quad \text{for } j = 1, \ldots, J, \quad (25)$$

where $I_j(z)$ is the modified Bessel function of the first kind of order j; see Solin & Särkkä (2014, Eq. 27).
Dynamic model The implied matrices for the dynamic model, Equation (5), are now given by

$$A = \text{diag}(A_0, \ldots, A_J), \quad \text{with blocks} \quad (26)$$

$$A_j = \begin{bmatrix} \cos(w_0 jt) & -\sin(w_0 jt) \\ \sin(w_0 jt) & \cos(w_0 jt) \end{bmatrix}, \quad \text{and} \quad (27)$$

$$Q = \mathbf{0} \in \mathbb{R}^{2(J+1) \times 2(J+1)}. \quad (28)$$

Measurement Model This dynamic model is, like all dynamic models, combinable with all measurement models. For ODEs, we need, by Equation (8), a model H_0 that extracts $x(t)$ and a model H that extracts the derivative from the state $X(t)$ of Equation (20). By Equations (13) and (15), this is satisfied by

$$H_0 = [1, 0, 1, 0, \ldots, 1, 0] \in \mathbb{R}^{1 \times 2(J+1)}, \quad \text{and} \quad (29)$$
$$H = \begin{bmatrix} 0, 0, 0, -1w_0, 0, -2w_0, \ldots, 0, -Jw_0 \end{bmatrix} \in \mathbb{R}^{1 \times 2(J+1)}.$$

3.1. Discussion of the Fourier model

As their coefficients do not depend on a support point, Fourier models are (unlike Taylor models) global expansions. Hence, they are best at extrapolating globally, while Taylor methods excel at extrapolating locally. As ODE methods are usually designed for a small step size $h > 0$, the Taylor approximation is the standard approximation in ODE solvers, such as Runge–Kutta methods. Hence, we expect the Fourier SSM to be more useful for global extrapolation with larger step sizes. Accordingly, we suggest a hybrid ODE solver which combines the Taylor and the Fourier state space model in the next section. This model could be used to extrapolate from a certain time, after learning the Fourier coefficients with data from the Taylor model.

4. The Hybrid Taylor-Fourier Model

As Taylor approximations excel at local approximations and Fourier approximations at global approximations of periodic signals, we combine both to the hybrid Taylor-Fourier model. The idea is that one can learn the Fourier coefficients (a_j, b_j) from Equations (17) and (18) with data from the Taylor SSM and then extrapolate with the Fourier SSM using the learned coefficients. Let us denote the Taylor SSM by $(A^{\text{Tay}}, Q^{\text{Tay}}, H^{\text{Tay}})$ and the Fourier SSM $(A^{\text{Four}}, Q^{\text{Four}}, H^{\text{Four}})$. The hybrid Filter works now as follows: It splits the time domain $[0, T]$ of the ODE into two parts $[0, T_p]$ and $[T_p, T]$ for some *prediction time point* $T_p \in (0, T)$. On the first interval $[0, T_p]$, we solve the ODE with the classical Gaussian ODE filter with the Taylor SSM, and we, simultaneously, train the Fourier SSM with the data from the Taylor model. On the second interval $[T_p, T]$, we just predict along the Fourier dynamical model defined by $(A^{\text{Four}}, Q^{\text{Four}})$.

As the computation on the time interval $[T_p, T]$ does not require additional evaluations of f and is therefore almost free, we hope that that this model turns out useful to reduce the computational time of solving periodic ODEs. In the next section we present some experiments which, while not practical yet, highlight that this hybrid model in principle works.
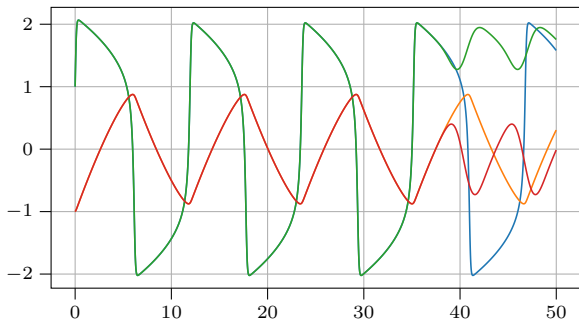
Figure 1. Hybrid Filter on Van der Pol ODE, Equation (30). Blue and yellow line are true curves x_1 and x_2. Red and green line are hybrid filter mean with prediction from $T_p = 37.5$.
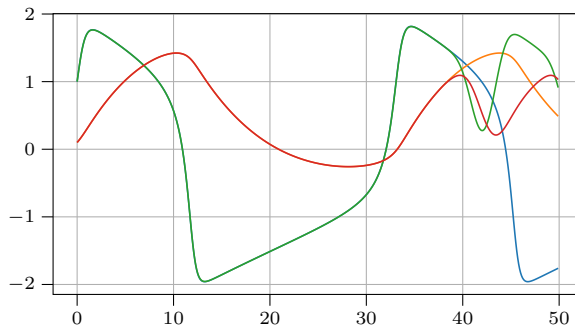

Figure 2. Hybrid Filter on FitzHugh–Nagumo ODE, Equation (31). Blue and yellow line are true curves x_1 and x_2. Red and green line are hybrid filter mean with prediction from $T_p = 37.5$.

5. Experiments

We try a Gaussian ODE filter with hybrid Taylor-Fourier SSM on two standard oscillating ODEs: the Van der Pol oscillator

$$\begin{cases} \dot{x}_1(t) & = \mu(x_1(t) - \frac{1}{3}x_1(t)^3 - x_2(t)), \\ \dot{x}_2(t) & = \frac{1}{\mu}x_1(t), \end{cases} \quad (30)$$

with $\mu = 5$, $x(0) = [1, -1]$, $T = 50$, and the FitzHugh–Nagumo model

$$\begin{cases} \dot{x}_1(t) & = x_1(t) - \frac{x_1(t)^3}{3} - x_2(t) + I, \\ \dot{x}_2(t) & = \frac{1}{\tau}\left(x_1(t) + a - x_2(t)\right), \end{cases} \quad (31)$$

with parameters $(I, a, b, \tau) = (0.5, 0.7, 0.8, 10.0)$, $x(0) = [1., 0.1]$ and $T = 50$.

5.1. Experimental Set-Up

We set the parameters of the Taylor model as follows: $q = 1$ and $\sigma^2 = 1$. Moreover, we choose the following parameters of the Fourier model: $l = 3$, $w_0 = 1$, $\sigma^2 = 1$, $J = 3$ $p = 2\pi$ and $R = 0$. Note that these parameters are not fine-tuned. We define the prediction time T_p for both ODEs to be $T_p = \frac{3}{4}T = 37.5$.

5.2. Results

The plots in Figures 1 and 2 show that the hybrid ODE filter works in principle. It picks up some structure from the trajectories on $[0, T_p]$ and can extrapolate forward by a sum of harmonic oscillators. The quality of the extrapolation is, however, not good enough yet. We suspect that this is due to our ad-hoc choice of parameters. Since our state space model is by Solin & Särkkä (2014) an approximation of GP regression with periodic kernel and derivative observations, it should be possible to make the extrapolation as accurate as periodic GP regression (at least for $J \to \infty$).

In particular, it should be possible to choose the angular

velocity w_0 in a more principled way and thereby match the period of the oscillator more precisely. Our choice of $w_0 = 1$ is particularly off in Figure 2. We also believe that a more accurate extrapolation could be achieved if a larger J is chosen. This will probably only work well once we have found a suitable way to choose w_0. Moreover, future research should examine which $J + 1$ data points from the Taylor model should be used for the Fourier model—which is an active learning task with Gaussian processes (Seo et al., 2000).

6. Conclusion

We examined how Fourier state space models can be employed in Gaussian ODE filtering, to solve oscillating ODEs. To this end, we developed a novel Fourier sate space model that is applicable to ODEs. We reasoned that it might outperform Taylor methods on global extrapolation tasks. Since Fourier expansions are not locally accurate enough to serve as a practical ODE solver on its own, we have developed the hybrid ODE Solver which combines Taylor and Fourier expansions. It first uses a Taylor SSM to compute up to a certain time T_p while training a Fourier SSM 'on the fly', and then uses the so-trained Fourier SSM to predict forward. We demonstrated that, in principle, this can work—even if we are not yet satisfied with the quality of the Fourier prediction.

Future research should examine how the Fourier coefficients can be learned better—e.g. by finding ways to choose the Fourier parameters (w_0, J) better or to employ smart active learning (Seo et al., 2000) for the Fourier coefficients (a_j, b_j). Since the desired Fourier coefficients are integrals, maybe ideas from Bayesian quadrature (Briol et al., 2019) can be borrowed for this purpose; see Equations (17) and (18). We hope that this might pave the way to almost cost-free predictions of oscillating systems which could come in useful in settings where ODEs have to be solved over a long time horizon with very limited budget or where

a (reinforcement learning) system has to make sudden decisions in the context of ODE dynamics (Deisenroth & Rasmussen, 2011).

If so, then exciting new ideas unknown to classical numerical analysis—such as quasi-periodic extrapolations (Solin & Särkkä, 2014, Section 3.5)—could be introduced to ODE solvers. Such a development would also benefit probabilistic numerical methods for boundary value problems (John et al., 2019), PDEs (Oates et al., 2019), and ODE inverse problems (Kersting et al., 2020).

In machine learning, such advances could provide better uncertainty quantification (of the numerical error) for continuous normalizing flows with ODEs, see (Chen et al., 2018, Section 4)—where a free-form ODE, potentially an oscillator (Grathwohl et al., 2019), has to be numerically solved to approximate the transformed state and the numerical error is, to date, not accounted for.

References

Briol, F.-X., Oates, C. J., Girolami, M., Osborne, M. A., and Sejdinovic, D. Probabilistic integration: A role for statisticians in numerical analysis? (with discussion and rejoinder). *Statistical Sciences*, 34(1):1–22 (Rejoinder on p38–42), 2019.

Chen, R., Rubanova, Y., Bettencourt, J., and Duvenaud, D. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

Deisenroth, M. and Rasmussen, C. PILCO: A Model-Based and Data-Efficient Approach to Policy Search. In *International Conference on Machine Learning (ICML)*, 2011.

Grathwohl, W., Chen, R. T. Q., Bettencourt, J., and Duvenaud, D. Scalable reversible generative models with free-form continuous dynamics. In *International Conference on Learning Representations*, 2019.

Hairer, E., Nørsett, S., and Wanner, G. *Solving Ordinary Differential Equations I – Nonstiff Problems*. Springer, 1987.

Hennig, P., Osborne, M. A., and Girolami, M. Probabilistic numerics and uncertainty in computations. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 471(2179):20150142, 2015.

John, D., Heuveline, V., and Schober, M. GOODE: A Gaussian off-the-shelf ordinary differential equation solver. In *International Conference on Machine Learning (ICML)*, 2019.

Kersting, H., Sullivan, T. J., and Hennig, P. Convergence rates of Gaussian ODE filters. *arXiv:1807.09737v2 [math.NA]*, 2019.

Kersting, H., Krämer, N., Schiegg, M., Daniel, C., Tiemann, M., and Hennig, P. Differentiable likelihoods for fast inversion of 'likelihood-free' dynamical systems. In *International Conference on Machine Learning (ICML)*, 2020.

Oates, C., Cockayne, J., Aykroyd, R., and Girolami, M. Bayesian probabilistic numerical methods in time-dependent state estimation for industrial hydrocyclone equipment. *Journal of the American Statistical Association*, 114(528):1518–1531, 2019.

Rackauckas, C., Ma, Y., Dixit, V., Guo, X., Innes, M., Revels, J., Nyberg, J., and Ivaturi, V. A comparison of automatic differentiation and continuous sensitivity analysis for derivatives of differential equation solutions. *arXiv:1812.01892 [math.NA]*, 2018.

Rezende, D. and Mohamed, S. Variational inference with normalizing flows. In *International Conference on Machine Learning (ICML)*, 2015.

Schober, M., Särkkä, S., and Hennig, P. A probabilistic model for the numerical solution of initial value problems. *Statistics and Computing*, 29(1):99–122, 2019.

Seo, S., Wallat, M., Graepel, T., and Obermayer, K. Gaussian process regression: Active data selection and test point rejection. *Mustererkennung*, 2000.

Solin, A. and Särkkä, S. Explicit link between periodic covariance functions and state space models. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, volume 33 of *Proceedings of Machine Learning Research*, pp. 904–912. PMLR, 2014.

Tronarp, F., Kersting, H., Särkkä, S., and Hennig, P. Probabilistic solutions to ordinary differential equations as nonlinear Bayesian filtering: a new perspective. *Statistics and Computing*, 29(6):1297–1315, 2019.

Tronarp, F., Kersting, H., Särkkä, S., and Hennig, P. Bayesian ode solvers: The maximum a posteriori estimate. *arXiv:2004.00623 [math.NA]*, 2020.