



Aalto University
School of Electrical
Engineering

State-space Gaussian Process for Drift Estimation in Stochastic Differential Equations

Zheng Zhao, Filip Tronarp, Roland Hostettler, and Simo Särkkä

Aalto University
zheng.zhao@aalto.fi

In 45th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2020)

Introduction

Let us consider an SDE

$$dX_t = a(X_t, t) dt + b(X_t, t) dW_t, \quad (1)$$

to solution $X_k \triangleq X_{t_k}$, where $W_k \triangleq W_{t_k}$ is a Wiener process with spectral density q , and suppose that there is a weakly unique solution.

Now we encounter a problem that how do we select those SDE coefficients (i.e., the drift function a and dispersion b)?

More practically, how to learn from **observations/data**,
 $X_{1:N} = \{X_1, X_2, \dots, X_N\}$?

Parametric Approaches

One way to proceed is the parametric approach, assuming that, for example, $a := a_\theta$.

See e.g., early studies by Kessler (1997), Dacunha-Castelle & Florens-Zmirou (1986), Lamouroux & Lehnertz (2009).

MLE, MAP (Papaspiliopoulos et al. 2012, Friedrich et al. 2011), and sampling-based (e.g., MCMC) (Papaspiliopoulos et al. 2012) methods are often used to estimate the parameters.

However, parametrization is not flexible, and the sampling-based inferencer are computationally expensive.

GP Approach

An alternative approach is to postulate the Gaussian process (GP) prior on the unknown SDE coefficients (Papaspiliopoulos et al. 2012, Ruttor et al. 2013, Rasmussen & Williams 2006).

Let us focus on the drift function a and assume that b is constant. We model a as a GP,

$$a \sim \mathcal{GP}(0, K(X, X')), \quad (2)$$

where K is a covariance function. We target at the posterior

$$p(a \mid X_{1:N}). \quad (3)$$

GP Approach

Next, to build the measurement model, we first need to discretize the SDE. A typical example is the Euler–Maruyama scheme, given by

$$X_{k+1} = X_k + a(X_k) \Delta t + b \Delta W_k, \quad (4)$$

where Δt is the time interval and $\Delta W_k = W_{k+1} - W_k \sim \mathcal{N}(0, q \Delta t)$. For notation convenience, we rewrite the above (4) as

$$Y_k = a(X_k) \Delta t + b \Delta W_k, \quad (5)$$

where $Y_k = X_{k+1} - X_k$.

GP Approach

Now the GP regression model for drift function a estimation is

$$\begin{aligned} a &\sim \mathcal{GP}(0, K(X, X')), \\ Y_k &= a(X_k) \Delta t + \epsilon_k, \end{aligned} \tag{6}$$

where $\epsilon_k \sim \mathcal{N}(0, q \Delta t)$. This is a vanilla GP regression problem, of which the posterior admits a closed-form (still Gaussian) (Rasmussen & Williams 2006).

This GP method has already been discussed by, for example, Papaspiliopoulos et al. (2012), Ruttor et al. (2013), García et al. (2017), Batz et al. (2018).

Problems

The regression on the above model (6) has many problem:

- ▶ The computation of batch GP regression scales cubically $\mathcal{O}(N^3)$.
- ▶ The measurement model by Euler–Maruyama only works well when Δt is small enough, as its strong order of convergence is merely $\mathcal{O}(\Delta t^{0.5})$ (Kloeden & Platen 1992).

We propose to solve by

- ▶ Perform GP regression using Bayesian smoothing, scales linearly $\mathcal{O}(N)$.
- ▶ Well then, use higher order convergence model, for example, Itô–Taylor strong order 1.5 (Itô-1.5).

Itô–Taylor Expansion

Let us replace the measurement model $Y_k = a(X_k) \Delta t + \epsilon_k$ in (6) with a more accurate formulation

$$Y_k = g(X_k) + \epsilon_k(X_k), \quad (7)$$

where function g random variable ϵ_k are obtained by approximating the solution

$$X_{k+1} = X_k + \int_{t_k}^{t_{k+1}} a(X_s) ds + \int_{t_k}^{t_{k+1}} b dW_s, \quad (8)$$

with higher order Itô–Taylor expansion (Kloeden & Platen 1992). The Euler–Maruyama is the case when the above integrals are simply estimated with rectangle (i.e., $g(X_k) = a(X_k) \Delta t$ and $\epsilon_k(X_k) = b \Delta W_k$).

Itô–Taylor Expansion

Here we use a particular useful expansion: Itô-1.5, such that

$$\begin{aligned}g(X_k) &= a(X_k) \Delta t + \frac{1}{2} \left(\frac{da(X_k)}{dX_k} a(X_k) + \frac{b^2 q}{2} \frac{d^2 a(X_k)}{dX_k^2} \right) \Delta t^2, \\ \epsilon_k(X_k) &= b \Delta \omega_k + \frac{da(X_k)}{dX_k} b \Delta \gamma_k,\end{aligned}\tag{9}$$

where

$$\begin{bmatrix} \Delta \gamma_k \\ \Delta \omega_k \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} q \Delta t^3/3 & q \Delta t^2/2 \\ q \Delta t^2/2 & q \Delta t \end{bmatrix} \right).\tag{10}$$

Not limited to more higher weak order expansions (see, e.g., Kloeden & Platen 1992, Zhao et al. 2020).

Itô–Taylor GP Regression Model

Now the GP regression model is

$$\begin{aligned} a &\sim \mathcal{GP}(0, K(X, X')), \\ Y_k &= g(X_k) + \epsilon_k(X_k). \end{aligned} \tag{11}$$

Unfortunately, the measurement model is non-linear, and contains the derivatives of GP: a' , a'' . This makes the GP regression even more complicated and expensive.

Our strategy is to transform the GP a into a state-space representation $\alpha = [a \quad a' \quad a'' \quad \dots]^T$, and thus the regression is done by non-linear Kalman filters and smoothers.

State-space Representation of GP

Knowing that a linear SDE characterizes a GP, thus we transform the GP a with certain covariance function by:

$$\begin{aligned} a &\sim \mathcal{GP}(0, K(X, X')) \\ &\Updownarrow \\ d\alpha(X) &= \mathbf{F} \alpha(X) dX + \mathbf{L} d\mathbf{W}(X) \end{aligned} \tag{12}$$

where $\alpha = [a \quad a' \quad a'' \quad \dots]^\top$, $\mathbf{W}(x)$ is a Wiener process with diffusion matrix \mathbf{Q} . The initial condition is $\alpha(X_0) \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_\infty)$, where \mathbf{P}_∞ is the stationary covariance obtained from Lyapunov equation (Särkkä & Solin 2019).

References for this connection: Lindgren et al. (2011), Särkkä & Solin (2019), Särkkä et al. (2013).

State-space Representation of GP

A typical example is the Matérn kernel

$$K(X, X') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}|X - X'|}{\ell} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}|X - X'|}{\ell} \right),$$

for which we have

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & \vdots & \ddots & \\ -\phi_1 \lambda^D & -\phi_2 \lambda^{D-1} & \cdots & -\phi_D \lambda \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

where $\lambda = \sqrt{2\nu}/\ell$, $D = \nu + 1/2$, and $\phi_i = \binom{D}{i-1}$ are the binomial coefficients.

Itô–Taylor GP State-space Model

Overall, our state-space Itô-1.5 GP regression model for the drift function estimation is

$$\begin{aligned}d\boldsymbol{\alpha}(X) &= \mathbf{F} \boldsymbol{\alpha}(X) dX + \mathbf{L} d\mathbf{W}(X), \\ Y_k &= g(\boldsymbol{\alpha}(X_k)) + \epsilon_k(\boldsymbol{\alpha}(X_k)).\end{aligned}\tag{13}$$

where the values of the parameters, for example, \mathbf{F} and \mathbf{L} depends on the GP covariance function. The estimation of the posterior

$$p(\boldsymbol{\alpha}(X) \mid Y_1, Y_2, \dots, Y_N),\tag{14}$$

is clearly a Bayesian filtering and smoothing problem (Särkkä 2013).

Itô–Taylor GP State-space Model

When the measurement model is linear (i.e., use Euler–Maruyama), the posterior

$$p(\boldsymbol{\alpha}(X) \mid Y_1, Y_2, \dots, Y_N) = \mathcal{N}(\boldsymbol{\alpha}(X); \mathbf{m}^s(X), \mathbf{P}^s(X)), \quad (15)$$

at X , is easily obtained by using Kalman filter and smoother, and is exactly the same as of the full batch GP regression in linear computational complexity.

Although our measurement model is non-linear (Itô-1.5), there are still plenty of good solvers e.g., sigma-point (unscented, cubature, Gauss–Hermite) filters and smoothers, and particle filters and smoothers.

Numerical Experiments

Consider two synthetic toy models

$$dx = 3(x - x^3) dt + dW, \quad x_0 = 1, \quad (16a)$$

$$dx = \tanh(x) dt + 0.01 dW, \quad x_0 = 0, \quad (16b)$$

and we want to estimate their drift functions. The samples are generated with small enough time intervals by Euler–Maruyama. We calculate the root mean square error (RMSE) and CPU time with 1000 independent Monte Carlo trials.

We choose Matérn $\nu = 5/2$ kernel for GPs.

Numerical Experiments

The following methods are chosen for comparison with naming convention.

- ▶ Full batch GP (GP).
- ▶ Fully independent conditional sparse GP (FIC).
- ▶ Deterministic training conditional sparse GP (DTC).
- ▶ Kalman filter and RTS smoother (KF-RTS)
- ▶ Unscented Kalman filter and RTS smoother (UKF-RTS)
- ▶ Iterated posterior linearization filter and smoother (IPLS) and RTS smoother (IPLF-RTS).

More experiment setting details shall be found in the paper.

Numerical Experiments

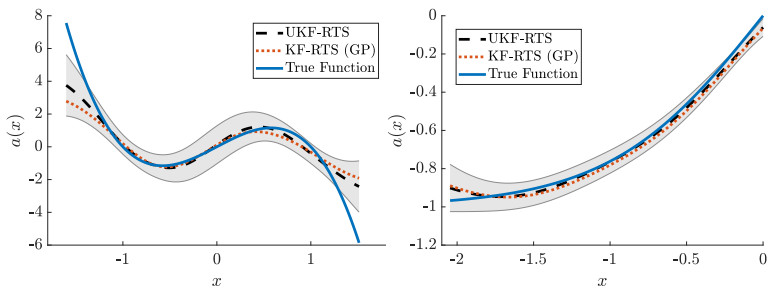


Figure 1: Demonstration of drift function estimation ($\Delta t = 0.1$ s) for model (16a) and (16b), respectively. Shaded area stands for 0.95 confidence interval of UKF-RTS.

Numerical Experiments

Compare RMSE and CPU time

	RMSE	CPU Time (s)
Kalman	1.42	0.245
GP	1.42	330.034
Sparse GP (FIC)	1.49	0.225
Sparse GP (DTC)	1.49	0.044

Table 1: RMSE and CPU time of drift estimation for model (16a) with Euler–Maruyam and $\Delta t = 0.002$ s.

Numerical Experiments

Compare RMSE with different Δt

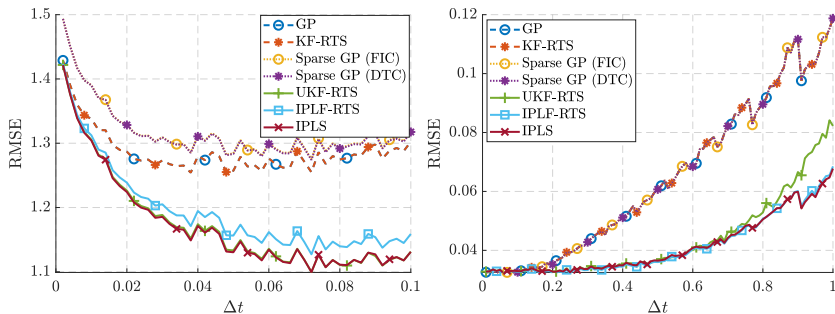


Figure 2: RMSE of drift estimation for model (16a) (left) and (16b) (right) as function of Δt .

Demonstration on Real-world Data

Monthly sunspot activity dataset by WDCSILSO, Royal Observatory of Belgium, Brussels.

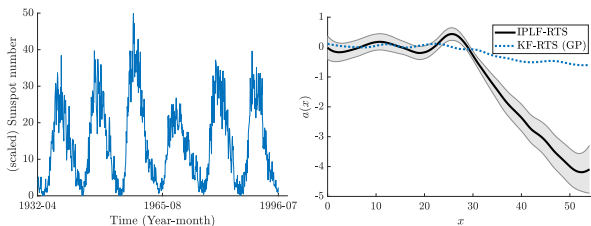


Figure 3: Drift estimation results (right column) for sunspot (top left) data. Shaded area stands for 0.95 confidence interval of IPLF-RTS.

The estimate seems reasonable, because the drift function for sunspot activity was modeled with a linear plus bias form by physicians (Allen & Huff 2010).

Demonstration on Real-world Data

Electrocardiogram (ECG) motion dataset by Aalto university and Helsinki University Central Hospital

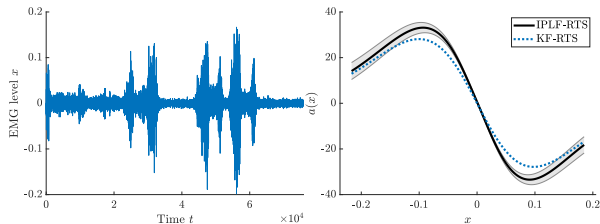


Figure 4: Drift estimation results (right column) for EMG motion (bottom left) data. GP gives out of memory error, as too many samples.

The estimated drift function has peak values around -0.1 mV and 0.1 mV, which seems to correspond to the EMG significant magnitudes due to human motion.

Conclusion

Conclusion:

- ▶ The work is concerned with estimation of unknown drift functions of SDEs from observations.
- ▶ We use non-parametric GP regression approach.
- ▶ We use more accurate measurement model by Itô–Taylor expansion, where the vanilla GP regression is not applicable.
- ▶ We use Bayesian smoothing instead of batch GP regression resorting to a linear computational complexity.

References

- Allen, E. J. & Huff, C. (2010), 'Derivation of stochastic differential equations for sunspot activity', *Astronomy & Astrophysics* **516**.
- Batz, P., Ruttor, A. & Opper, M. (2018), 'Approximate Bayes learning of stochastic differential equations', *Physical Review E* **98**.
- Dacunha-Castelle, D. & Florens-Zmirou, D. (1986), 'Estimation of the coefficients of a diffusion from discrete observations', *Stochastics* **19**(4), 263–284.
- Friedrich, R., Peinke, J., Sahimi, M. & Tabar, M. R. R. (2011), 'Approaching complexity by stochastic methods: From biological systems to turbulence', *Physics Reports* **506**(5), 87–162.
- García, C. A., Otero, A., Félix, P., Presedo, J. & Márquez, D. G. (2017), 'Nonparametric estimation of stochastic differential equations with sparse Gaussian processes', *Physical Review E* **96**.
- Kessler, M. (1997), 'Estimation of an ergodic diffusion from discrete observations', *Scandinavian Journal of Statistics* **24**(2), 211–229.

References (cont.)

- Kloeden, P. E. & Platen, E. (1992), *Numerical Solution of Stochastic Differential Equations*, Springer–Verlag Berlin Heidelberg.
- Lamberton, D. & Lapeyre, B. (2007), *Introduction to Stochastic Calculus Applied to Finance*, Chapman and Hall/CRC.
- Lamouroux, D. & Lehnertz, K. (2009), ‘Kernel-based regression of drift and diffusion coefficients of stochastic processes’, *Physics Letters A* **373**(39), 3507–3512.
- Lindgren, F., Rue, H. & Lindström, J. (2011), ‘An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(4), 423–498.
- Øksendal, B. (2003), *Stochastic Differential Equations: An Introduction with Applications*, 6 edn, Springer.

References (cont.)

- Papaspiliopoulos, O., Pokern, Y., Roberts, G. O. & Stuart, A. M. (2012), 'Nonparametric estimation of diffusions: a differential equations approach', *Biometrika* **99**(3), 511–531.
- Rasmussen, C. E. & Williams, C. K. I. (2006), *Gaussian Processes for Machine Learning*, The MIT Press.
- Rutter, A., Batz, P. & Opper, M. (2013), Approximate Gaussian process inference for the drift function in stochastic differential equations, in 'Advances in Neural Information Processing Systems 26', MIT Press, pp. 2040–2048.
- Särkkä, S. (2013), *Bayesian Filtering and Smoothing*, Institute of Mathematical Statistics Textbooks, Cambridge University Press.
- Särkkä, S. & Solin, A. (2019), *Applied Stochastic Differential Equations*, Institute of Mathematical Statistics Textbooks, Cambridge University Press.

References (cont.)

Särkkä, S., Solin, A. & Hartikainen, J. (2013), 'Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing: A look at Gaussian process regression through Kalman filtering', *IEEE Signal Processing Magazine* **30**(4), 51–61.

Zhao, Z., Karvonen, T., Hostettler, R. & Särkkä, S. (2020), 'Taylor moment expansion for continuous-discrete Gaussian filtering and smoothing', *arXiv preprint arXiv:2001.02466* .