Asymptotics of maximum likelihood parameter estimates for Gaussian processes: the Ornstein–Uhlenbeck prior

Toni Karvonen, Filip Tronarp, and Simo Särkkä

This article has been published in *Proceedings of 29th IEEE International Workshop on Machine Learning for Signal Processing*, held in Pittsburgh, PA, USA, on October 13–16, 2019. When citing this article, use the following reference:

T. Karvonen, F. Tronarp, and S. Särkkä, "Asymptotics of maximum likelihood parameter estimates for Gaussian processes: the Ornstein–Uhlenbeck prior," in *29th IEEE International Workshop on Machine Learning for Signal Processing*, 2019.

DOI: 10.1109/MLSP.2019.8918767

ASYMPTOTICS OF MAXIMUM LIKELIHOOD PARAMETER ESTIMATES FOR GAUSSIAN PROCESSES: THE ORNSTEIN–UHLENBECK PRIOR

Toni Karvonen, Filip Tronarp, and Simo Särkkä

Aalto University
Department of Electrical Engineering and Automation
Espoo, Finland

ABSTRACT

This article studies the maximum likelihood estimates of magnitude and scale parameters for a Gaussian process of Ornstein–Uhlenbeck type used to model a deterministic function that does not have to be a realisation of an Ornstein–Uhlenbeck process. Specifically, we derive explicit expressions for the limiting values of the maximum likelihood estimates as the number of observations increases. The results demonstrate that the function typically needs to be sufficiently similar to a sample path of an Ornstein–Uhlenbeck process or have discontinuities if the variance of the model is to remain non-zero. Numerical examples illustrate the behaviour of the estimates when the function is not a sample path of any Ornstein–Uhlenbeck process.

***Index Terms*—** Gaussian process regression, Ornstein–Uhlenbeck process, maximum likelihood estimation, probabilistic numerics

1. INTRODUCTION

Gaussian processes are often used to model deterministic functions in computer experiments [13] and probabilistic numerical analysis [10, 5, 2]. Despite being advertised as non-parametric, Gaussian process models usually contain covariance kernel hyperparameters that need to be selected carefully to obtain useful posterior estimates and meaningful quantification of uncertainty. The hyperparameters are often selected by maximising the likelihood of the function evaluations given the hyperparameter values [12, Section 5.4.1]. Marginalisation and cross-validation are popular alternatives.

In probabilistic numerics, the Gaussian process regression posterior variance quantifies the uncertainty associated to a numerical approximation. If the Gaussian process model is fixed, the posterior variance does not depend on the observations and is hence not useful for uncertainty quantification. The most popular approach to make uncertainty quantification meaningful, in the sense that it to some degree reflects the true numerical error, is to fit the kernel parameters using maximum likelihood (see, e.g., [8, 1, 19]). However, even when this is done it is possible that, for functions that are not well modelled by the Gaussian process model, the posterior variance is not representative of the true error. Unfortunately, little work has been done to analyse how the maximum likelihood estimates behave for different functions. The only results we are aware of are by Xu and Stein [21] for the Gaussian covariance kernel and either constant or linear functions. Note that much work has been done in a slightly different setting [20, 17, 18, 3].

In this article we consider *arbitrary* functions on $[0, 1]$ and analyse the asymptotic behaviour of the maximum likelihood estimates of magnitude and scale parameters of a Gaussian process model with the Ornstein–Uhlenbeck kernel (3) as the number of function evaluations on equispaced points increases (implications to uncertainty quantification are left future research). This particular model, related to the more conventional zero-mean Matérn $1/2$ model, is chosen here because inference and maximum likelihood estimation associated to it are highly tractable due to its representation as a Ornstein–Uhlenbeck process (1). The main results are contained in Propositions 3.2 to 3.4 that provide expression for the limiting values of the maximum likelihood estimates. Note that maximum likelihood estimation for the Ornstein–Uhlenbeck process has been studied for a long time [9, 11]. Our analysis is distinguished from the previous work in that we do *not* require that the observations come from a realisation of an Ornstein–Uhlenbeck process; instead, much of our interest is in the misspecified case when the function *cannot* be a sample path of the Gaussian process used to model it. This case is discussed in Section 3.3 and illustrated by a numerical example in Section 4.

2. SETTING

Let $f\colon [0,1] \to \mathbb{R}$ be a deterministic function that is evaluated exactly at $N + 1 \geq 2$ equispaced points

$$\{t_n\}_{n=0}^N \coloneqq \{0, h, 2h, \ldots, 1-h, 1\}, \qquad h = \frac{1}{N}.$$

This work was supported by the Aalto ELEC Doctoral School and the Academy of Finland.

978-1-7281-0824-7/19/$31.00

The function evaluations are $f_n \coloneqq f(t_n) = f(nh)$ and the full set of data is

$$\mathcal{D}_N = \{(nh, f(nh))\}_{n=0}^N = \{(t_n, f_n)\}_{n=0}^N.$$

2.1. An Ornstein–Uhlenbeck Model

We model f with the Ornstein–Uhlenbeck process X, the solution of the stochastic differential equation (SDE)

$$\mathrm{d}X(t) = -\lambda X(t)\,\mathrm{d}t + \sqrt{2\lambda}\sigma\,\mathrm{d}W(t), \tag{1}$$

where $\lambda > 0$ is a *scale parameter*, $\sigma > 0$ a *magnitude parameter*, and W the standard Wiener process. We use the deterministic initial condition $X(0) = f(0)$. This means that X is a Gaussian process with the mean

$$\mathbb{E}[X(t)] = f(0)\mathrm{e}^{-\lambda t} \tag{2}$$

and the covariance kernel

$$k(t,t') = \mathrm{Cov}[X(t), X(t')] = \sigma^2\big(\mathrm{e}^{-\lambda|t-t'|} - \mathrm{e}^{-\lambda(t+t')}\big). \tag{3}$$

The specific parametrisation of the Ornstein–Uhlenbeck process that we use is motivated by a connection to the Matérn $1/2$ (or exponential) kernel

$$k_{1/2}(t,t') = \sigma^2\mathrm{e}^{-\lambda|t-t'|}. \tag{4}$$

This is the covariance kernel of the stationary version of X, which is obtained by using the Gaussian initial condition $X(0) \sim \mathcal{N}(0, \sigma^2)$.

We are interested in analysing the behaviour of the maximum likelihood estimates σ^2_{ML} and λ_{ML} of the model parameters. Specifically, we study the asymptotic behaviour of the maximum likelihood estimates as $N \to \infty$.

2.2. On sample paths of Ornstein–Uhlenbeck processes

A function $f\colon [0,1] \to \mathbb{R}$ is said to be *Hölder continuous* with parameter $\alpha > 0$ if

$$|f(t) - f(t')| \leq C\,|t-t'|^\alpha$$

for some constant $C > 0$ and all $t, t' \in [0,1]$. If $\alpha = 1$, the function is Lipschitz continuous and hence continuously differentiable; if $\alpha > 1$, the function must be constant. Almost all sample paths of the Brownian motion or an Ornstein–Uhlenbeck process are non-differentiable and (locally) Hölder continuous with any $\alpha < 1/2$ [7, Proposition 3.4].

The *quadratic variation*

$$V^2(f) \coloneqq \lim_{N\to\infty} \sum_{n=1}^N (f_n - f_{n-1})^2$$

of f can be non-zero only if f is Hölder continuous with $\alpha \leq 1/2$. To see this, note that

$$\sum_{n=1}^N (f_n - f_{n-1})^2 \leq C \sum_{n=1}^N h^{2\alpha} = CN^{1-2\alpha},$$

which vanishes as $N \to \infty$ if $\alpha > 1/2$. In particular, almost all sample paths of the standard Brownian motion on $[0,1]$ have quadratic variation 1, while the quadratic variation of almost any sample path of the Ornstein–Uhlenbeck process X, as defined by (1), is $2\lambda\sigma$ (e.g., [22, Equation (3)]).

3. ASYMPTOTICS OF MAXIMUM LIKELIHOOD ESTIMATES

This section contains the main results of this article on the asymptotic behaviour of λ_{ML} and σ^2_{ML}. We also discuss the interpretation of the results. Note that the results do not require that f is a sample path of an Ornstein–Uhlenbeck process.

3.1. Maximum likelihood estimation

Because the Ornstein–Uhlenbeck process is a Markov process, the negative log-likelihood function of the function evaluations given the parameters can be conveniently factorised:

$$\begin{aligned}\ell(\lambda, \sigma^2) &\coloneqq -\log p(f_0 \ldots, f_N \mid \lambda, \sigma^2) \\ &= -\log\Bigg[\prod_{n=1}^N p(f_n \mid f_{n-1}, \lambda, \sigma^2)\Bigg] \\ &= -\sum_{n=1}^N \log p(f_n \mid f_{n-1}, \lambda, \sigma^2),\end{aligned}$$

where the transition densities are

$$p(f_n \mid f_{n-1}, \lambda, \sigma^2) = \mathcal{N}\big(f_n \mid f_{n-1}\mathrm{e}^{-\lambda h}, \sigma^2[1-\mathrm{e}^{-2\lambda h}]\big).$$

The negative log-likelihood function is therefore

$$\ell(\lambda, \sigma^2) = \frac{1}{2}\sum_{n=1}^N \Bigg[\log\big(2\pi\sigma^2[1-\mathrm{e}^{-2\lambda h}]\big) + \frac{(f_n - f_{n-1}\mathrm{e}^{-\lambda h})^2}{\sigma^2(1-\mathrm{e}^{-2\lambda h})}\Bigg].$$

By using suitable substitutions it is fairly straightforward to derive the maximum likelihood estimates [14, Section 11.3] (see also [6, Section 3.1.1])

$$\lambda_{\text{ML}} = -\frac{1}{h}\log\Bigg[\frac{\sum_{n=1}^N f_n f_{n-1}}{\sum_{n=1}^N f_{n-1}^2}\Bigg], \tag{5}$$

$$\sigma^2_{\text{ML}} = \frac{1}{N(1-\mathrm{e}^{-2\lambda_{\text{ML}} h})}\sum_{n=1}^N (f_n - f_{n-1}\mathrm{e}^{-\lambda_{\text{ML}} h})^2. \tag{6}$$

Denote

$$b_N = \sum_{n=1}^N f_n f_{n-1}, \quad c_N = \sum_{n=1}^N f_{n-1}^2, \quad d_N = \sum_{n=1}^N f_n^2.$$

Upon insertion of the expression for λ_{ML} the magnitude estimate becomes

$$\begin{aligned}\sigma_{\text{ML}}^2 &= \frac{1}{N}\left(d_N - \frac{b_N^2}{c_N}\right)\frac{c_N^2}{c_N^2 - b_N^2} \\ &= \frac{c_N}{N} \times \frac{d_N c_N - b_N^2}{c_N^2 - b_N^2}.\end{aligned}$$

3.2. Asymptotic analysis

We now analyse the asymptotic behaviour of the maximum likelihood estimates (5) and (6).

Lemma 3.1. *Suppose that f is Riemann integrable and $V^2(f) < \infty$. Then*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N f_n^2 = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N f_n f_{n-1} = \int_0^1 f(t)^2 \,\mathrm{d}t.$$

and

$$\lim_{N\to\infty} \sum_{n=1}^N f_{n-1}(f_n - f_{n-1}) = \frac{1}{2}\big[f(1)^2 - f(0)^2 - V^2(f)\big].$$

Proof. The first limit in the first statement follows from the Riemann integrability of f. The other limits follow from the expansion

$$\sum_{n=1}^N (f_n - f_{n-1})^2 = \sum_{n=1}^N (f_n^2 + f_{n-1}^2) - 2\sum_{n=1}^N f_n f_{n-1}.$$

That $\frac{1}{N}\sum_{n=1}^N f_n f_{n-1} \to \int_0^1 f(t)^2 \,\mathrm{d}t$ follows from arranging this equation as

$$\sum_{n=1}^N f_n f_{n-1} = \frac{1}{2}\left[\sum_{n=1}^N (f_n^2 + f_{n-1}^2) - \sum_{n=1}^N (f_n - f_{n-1})^2\right] \quad (7)$$

and recalling that the quadratic variation has been assumed to be finite. To prove the second statement, note that

$$\sum_{n=1}^N f_n f_{n-1} = \sum_{n=1}^N f_{n-1}^2 + \sum_{n=1}^N f_{n-1}(f_n - f_{n-1}),$$

which, when combined with (7), yields

$$\begin{aligned}&\sum_{n=1}^N f_{n-1}(f_n - f_{n-1}) \\ &\qquad = \frac{1}{2}\left[\sum_{n=1}^N (f_n^2 - f_{n-1}^2) - \sum_{n=1}^N (f_n - f_{n-1})^2\right] \\ &\qquad = \frac{1}{2}\left[f(1)^2 - f(0)^2 - \sum_{n=1}^N (f_n - f_{n-1})^2\right] \\ &\qquad \to \frac{1}{2}\big[f(1)^2 - f(0)^2 - V^2(f)\big]\end{aligned}$$

as $N \to \infty$. □

Proposition 3.2. *Suppose that f is Riemann integrable, $V^2(f) < \infty$, and $f(0)^2 - f(1)^2 + V^2(f) \neq 0$. Then*

$$\lim_{N\to\infty} \sigma_{\text{ML}}^2 = \frac{V^2(f)\int_0^1 f(t)^2 \,\mathrm{d}t}{f(0)^2 - f(1)^2 + V^2(f)}.$$

Proof. First, observe that

$$\begin{aligned}c_N - b_N &= -\sum_{n=1}^N f_{n-1}(f_n - f_{n-1}) \\ &\to \frac{1}{2}\big[f(0)^2 - f(1)^2 + V^2(f)\big]\end{aligned}$$

as $N \to \infty$ by Lemma 3.1. By using the identity

$$d_N = c_N + f(1)^2 - f(0)^2$$

we can write

$$d_N c_N - b_N^2 = c_N^2 - b_N^2 + c_N[f(1)^2 - f(0)^2].$$

Therefore

$$\begin{aligned}\sigma_{\text{ML}}^2 &= \frac{c_N}{N} \times \frac{d_N c_N - b_N^2}{c_N^2 - b_N^2} \\ &= \frac{c_N}{N}\left(1 + \frac{c_N[f(1)^2 - f(0)^2]}{(c_N + b_N)(c_N - b_N)}\right) \\ &\to \int_0^1 f(t)^2 \,\mathrm{d}t \left(1 + \frac{f(1)^2 - f(0)^2}{f(0)^2 - f(1)^2 + V^2(f)}\right) \\ &= \frac{V^2(f)\int_0^1 f(t)^2 \,\mathrm{d}t}{f(0)^2 - f(1)^2 + V^2(f)}\end{aligned}$$

as $N \to \infty$. □

Proposition 3.3. *Suppose that f is Riemann integrable, $V^2(f) < \infty$, and $\int_0^1 f(t)^2 \,\mathrm{d}t > 0$. Then*

$$\lim_{N\to\infty} \lambda_{\text{ML}} = \frac{f(0)^2 - f(1)^2 + V^2(f)}{2\int_0^1 f(t)^2 \,\mathrm{d}t}.$$

Proof. Observe that

$$\begin{aligned}\lambda_{\text{ML}} &= -\frac{1}{h}\log\left[\frac{\sum_{n=1}^N f_{n-1}^2 + \sum_{n=1}^N f_{n-1}(f_n - f_{n-1})}{\sum_{n=1}^N f_{n-1}^2}\right] \\ &= -\frac{1}{h}\log\left[1 + \frac{\sum_{n=1}^N f_{n-1}(f_n - f_{n-1})}{\sum_{n=1}^N f_{n-1}^2}\right].\end{aligned}$$

Because it holds that $\frac{1}{N}\sum_{n=1}^N f_{n-1}^2 \to \int_0^1 f(t)^2 \,\mathrm{d}t > 0$ and $\frac{1}{N}\sum_{n=1}^N f_{n-1}(f_n - f_{n-1}) = \mathcal{O}(N^{-1})$, we can expand the logarithm and write

$$\begin{aligned}\lambda_{\text{ML}} &= -N\left[\frac{\sum_{n=1}^N f_{n-1}(f_n - f_{n-1})}{\sum_{n=1}^N f_{n-1}^2} + \mathcal{O}(N^{-2})\right] \\ &= -\frac{\sum_{n=1}^N f_{n-1}(f_n - f_{n-1})}{\frac{1}{N}\sum_{n=1}^N f_{n-1}^2} + \mathcal{O}(N^{-1}).\end{aligned}$$

The claim now follows from Lemma 3.1. □

The two preceding propositions immediately yield the following result.

Proposition 3.4. *Suppose that f is Riemann integrable, $V^2(f) < \infty$, and $\int_0^1 f(t)^2 \,\mathrm{d}t > 0$. Then*

$$\lim_{N \to \infty} \lambda_{ML} \sigma_{ML}^2 = \frac{V^2(f)}{2}.$$

This result means that the variance of the Wiener process in the SDE (1) converges to the quadratic variation of f.

3.3. Interpretation of the results

We then briefly discuss the effects different properties of f have on the asymptotic maximum likelihood estimates.

Quadratic variation. Let us first discuss the effect of the quadratic variation $V^2(f)$. If f is differentiable, then $V^2(f) = 0$. Consequently,

$$\lambda_{\mathrm{ML}} \to \frac{f(0)^2 - f(1)^2}{2 \int_0^1 f(t)^2 \,\mathrm{d}t} = - \frac{\int_0^1 f(t) f'(t) \,\mathrm{d}t}{\int_0^1 f(t)^2 \,\mathrm{d}t} \tag{8}$$

and the SDE (1) becomes the deterministic ordinary differential equation $X'(t) = -\lambda_{\mathrm{ML}} X(t)$. This is reasonable since $-\lambda_{\mathrm{ML}} f$ is the projection of f' onto f in $L^2([0,1])$. Moreover, $\sigma_{\mathrm{ML}}^2 \to 0$, which is explained by the fact that a function with zero quadratic variation is a sample path of the Ornstein–Uhlenbeck process with probability zero.

Decay of f. For the limits of λ_{ML} and σ_{ML}^2 to be positive it is necessary that

$$f(0)^2 - f(1)^2 + V^2(f) > 0,$$

which is to say that unless the quadratic variation is sufficiently large, the magnitude of the initial value of f must exceed that at the end point. As (1) for $\lambda > 0$ describes a contractive model, this is an intuitive requirement. Irregularity (or "randomness") of f, as measured by the quadratic variation, allows for some deviation from the strict decay condition $|f(1)| < |f(0)|$. If $f(0)^2 - f(1)^2 + V^2(f) = 0$ and $V^2(f) > 0$, the model reverts to the Wiener model

$$\mathrm{d}X_t = \sqrt{V^2(f)} \,\mathrm{d}W_t$$

with the Brownian motion covariance kernel

$$k_{\mathrm{BM}}(t, t') = V^2(f) \min\{t, t'\}. \tag{9}$$

Because the posterior mean of a Gaussian process with the kernel (9) is the linear spline interpolant, this is reminiscent of the observation in the scattered data approximation literature that at the limit $\lambda \to 0$ kernel interpolants for finitely smooth kernels convergence to certain polyharmonic spline interpolants [16].

4. EXAMPLES

This section numerically illustrates the behaviour of the maximum likelihood estimates in (a) a *misspecified setting* with a differentiable function that is a sample path of the Ornstein–Uhlenbeck process with probability zero and (b) a more ambiguous setting with a fairly regular function with discontinuities. Furthermore, we comment on a function for which the estimates are available in very simple form for any $N \geq 1$. All three example functions satisfy the assumption of the theoretical results in Section 3.

4.1. Differentiable function

We first consider the differentiable function

$$f(t) = 1 - \sqrt{t} \sin(\tfrac{9}{4}\pi t) + \tfrac{1}{10} \sin(64\pi t). \tag{10}$$

Because the function is differentiable, $V^2(f) = 0$, which implies that $\sigma_{\mathrm{ML}}^2 \to 0$. The limiting value of the scale parameter is

$$\lim_{N \to \infty} \lambda_{\mathrm{ML}} \approx 0.4212.$$

The function, as well as the maximum likelihood estimates for N up to 10,000, are displayed in Figure 1.

4.2. Discontinuous function

As a second example we use the discontinuous function

$$f(t) = \begin{cases} 1 - 3\,|t - \frac{1}{6}| & \text{if} \quad 0 \leq t < \frac{1}{3} \\ \frac{1}{10} + 10(t - \frac{1}{2})^2 & \text{if} \quad \frac{1}{3} \leq t \leq \frac{2}{3} \\ 1 - 3\,|t - \frac{5}{6}| & \text{if} \quad \frac{2}{3} < t \leq 1 \end{cases}. \tag{11}$$

Due to the discontinuities at $\frac{1}{3}$ and $\frac{2}{3}$, this function has the non-zero quadratic variation

$$V^2(f) = 2(\tfrac{1}{10} + \tfrac{10}{36} - \tfrac{1}{2})^2 = \tfrac{121}{4050} \approx 0.0299.$$

The limiting values for the maximum likelihood estimates are

$$\lim_{N \to \infty} \lambda_{\mathrm{ML}} \approx 0.0370 \quad \text{and} \quad \lim_{N \to \infty} \sigma_{\mathrm{ML}}^2 \approx 0.4035.$$

The function and the maximum likelihood estimates are displayed in Figure 2.

4.3. Exponential function

Let $f(t) = \mathrm{e}^{-\theta t}$ for $\theta > 0$. Then it is easy to compute that

$$\lambda_{\mathrm{ML}} = \theta \quad \text{and} \quad \sigma_{\mathrm{ML}}^2 = 0$$

for any $N \geq 1$. The probability of X deviating from f vanishes as $\sigma \to 0$ in the SDE (1). Consequently, if the observations appear to come from an exponential function, the most likely parameter values are those giving rise to the ordinary differential equation $X'(t) = -\theta X(t)$, solved by $X(t) = f(t)$.

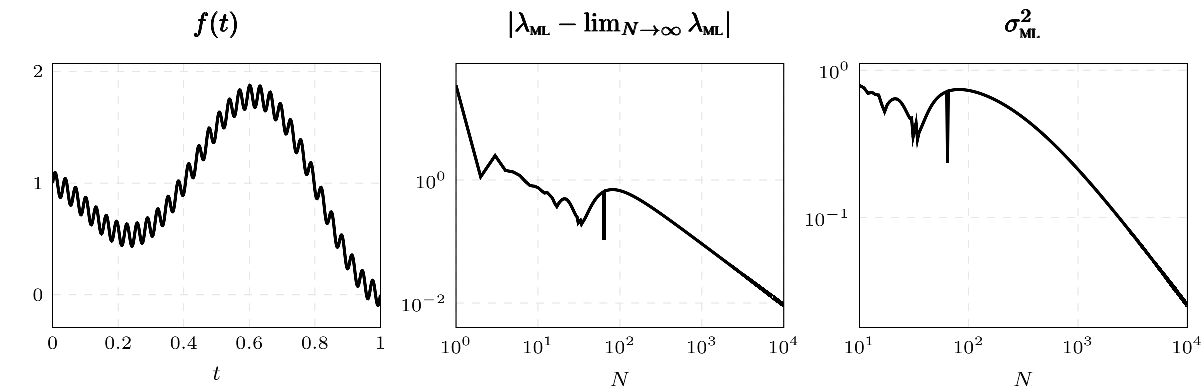

Fig. 1: The differentiable function (10) on $[0, 1]$ and the behaviour of the maximum likelihood estimates λ_{ML} and σ^2_{ML} in (5) and (6) as functions of N.

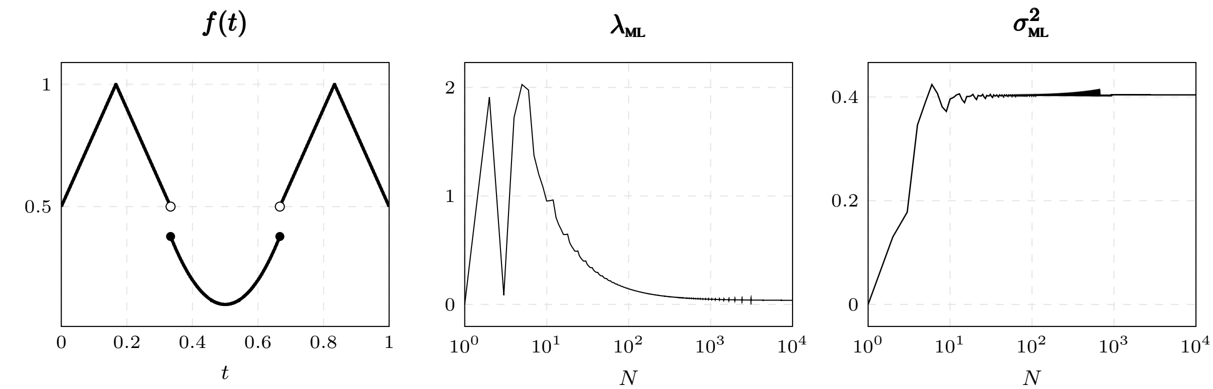

Fig. 2: The discontinuous function (11) on $[0, 1]$ and the behaviour of the maximum likelihood estimates λ_{ML} and σ^2_{ML} in (5) and (6) as functions of N.

5. CONCLUSIONS AND DISCUSSION

We have considered a Gaussian process model of Ornstein–Uhlenbeck type, defined by (1) or, equivalently, by the mean function (2) and covariance kernel (3), for a deterministic function $f\colon [0,1] \to \mathbb{R}$. Propositions 3.2 to 3.4 provide expressions for the limits of the maximum likelihood estimates of the magnitude and scale parameters σ and λ of the Gaussian process model as the number of evaluations increases.

The most important practical conclusions concern stochasticity of the model, defined by the SDE (1). If $\lim_{N\to\infty} \lambda_{\text{ML}}\sigma^2_{\text{ML}} = 0$, the Wiener process term vanishes and the model becomes a deterministic differential equation. Stochasticity of the model is typically retained if the function being modelled either (a) is sufficiently rough and similar to a sample path of an Ornstein–Uhlenbeck process in the sense that its quadratic variation is non-zero or (b) has a finite number of discontinuities. The latter condition is perhaps surprising because elsewhere the function does not have to bear any resemblance to sample paths of an Ornstein–Uhlenbeck process. For example, the function can be infinitely smooth outside the discontinuities.

To a Gaussian process user the model we have studied is fairly atypical. We have focussed on this particular model only because its analytical simplicity. It is easy to see that the maximum likelihood estimate of σ for the more conventional zero-mean Matérn $1/2$ model specified by the kernel (4) is

$$\sigma^2_{\text{ML}} = \frac{1}{N+1}\left[f_0^2 + \sum_{n=1}^{N} \frac{(f_n - f_{n-1}\mathrm{e}^{-\lambda h})^2}{1-\mathrm{e}^{-2\lambda h}} \right].$$

However, the scale parameter is only available as a solution to the cubic equation

$$\begin{aligned} N'(c_N - f_0^2)a^3 + (1-2N')b_N a^2 \\ + (N' f_0^2 + N' d_N - c_N - d_N)a + b_N = 0, \end{aligned} \tag{12}$$

where $N' = N/(N+1)$ and $a = \mathrm{e}^{-\lambda_{\text{ML}} h}$. As mentioned in Section 2, the only difference between the Matérn $1/2$ model and the one analysed here is the initialisation: the Matérn model corresponds to (1) with $X(0) \sim \mathcal{N}(0, \sigma^2)$ while we have used $X(0) = f(0)$. Because the effect of the prior ought to diminish as more data is obtained, we expect that Propositions 3.2 to 3.4 hold also for the maximum likelihood estimates of the Matérn $1/2$ model. Proving this requires careful analysis of limiting behaviour, as $N \to \infty$, of the appropriate solution of (12).

An obvious further generalisations would be to consider a general Matérn model of smoothness $\nu > 0$, defined by the covariance function

$$k_\nu(t, t') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \big(\lambda \, |t - t'|\big)^\nu \mathrm{K}_\nu\big(\lambda \, |t - t'|\big), \qquad (13)$$

where Γ is the Gamma function and K_ν a modified Bessel function of the second kind. When $\nu = p - 1/2$ for positive integer p, the Gaussian process defined by (13) corresponds to a pth order linear time-invariant SDE [4, 15]. Although this correspondence leads to computational speed-ups, the likelihood function becomes more complicated and theoretical analysis accordingly more involved.

6. REFERENCES

[1] F.-X. Briol, C. J. Oates, M. Girolami, M. A. Osborne, and D. Sejdinovic. Probabilistic integration: A role in statistical computation? *Statistical Science*, 34(1):1–22, 2019.

[2] J. Cockayne, C. Oates, T. Sullivan, and M. Girolami. Bayesian probabilistic numerical methods. *SIAM Review*, 2019. To appear.

[3] A. Hadji and B. Szabó. Can we trust Bayesian uncertainty quantification from Gaussian process priors with squared exponential covariance kernel? *arXiv:1904.01383*, 2019.

[4] J. Hartikainen and S. Särkkä. Kalman filtering and smoothing solutions to temporal Gaussian process regression models. In *20th IEEE International Workshop on Machine Learning for Signal Processing*, pages 379–384, 2010.

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

[6] S. M. Iacus. *Simulation and Inference for Stochastic Differential Equations: With R Examples*. Springer Series in Statistics. Springer, 2008.

[7] T. Kaarakka and P. Salminen. On fractional Ornstein-Uhlenbeck processes. *Communications on Stochastic Analysis*, 5(1):121–133, 2011.

[8] T. Karvonen, C. J. Oates, and S. Särkkä. A Bayes–Sard cubature method. In *Advances in Neural Information Processing Systems*, volume 31, pages 5882–5893, 2018.

[9] K. Koncz. On the parameter estimation of diffusional type processes with constant coefficients (Elementary Gaussian Processes). *Analysis Mathematica*, 13(1):75–91, 1987.

[10] F. M. Larkin. Gaussian measure in Hilbert space and applications in numerical analysis. *Rocky Mountain Journal of Mathematics*, 2(3):379–422, 1972.

[11] A. R. Pedersen. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22(1):55–71, 1995.

[12] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, 2006.

[13] J. Sacks, Welch W. J., T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423, 1989.

[14] S. Särkkä and A. Solin. *Applied Stochastic Differential Equations*. Cambridge University Press, 2019.

[15] S. Särkkä, A. Solin, and J. Hartikainen. 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, 2013.

[16] G. Song, J. Riddle, G. E. Fasshauer, and F. J. Hickernell. Multivariate interpolation with increasingly flat radial basis functions of finite smoothness. *Advances in Computational Mathematics*, 36(3):485–501, 2012.

[17] B. Szabó, A. W. van der Vaart, and J. H. van Zanten. Empirical Bayes scaling of Gaussian priors in the white noise model. *Electronic Journal of Statistics*, 7:991–1018, 2013.

[18] B. Szabó, A. W. van der Vaart, and J. H. van Zanten. Frequentist coverage of adaptive nonparametric Bayesian credible sets. *The Annals of Statistics*, 43(4):1391–1428, 2015.

[19] F. Tronarp, H. Kersting, S. Särkkä, and P. Hennig. Probabilistic solutions to ordinary differential equations as non-linear Bayesian filtering: A new perspective. *Statistics and Computing*, 2019. To appear.

[20] A. W. van der Vaart and J. H. van Zanten. Information rates of nonparametric Gaussian process methods. *Journal of Machine Learning Research*, 12:2095–2119, 2011.

[21] W. Xu and M. L. Stein. Maximum likelihood estimation for a smooth Gaussian random field model. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):138–175, 2017.

[22] L. Zhang, P. A Mykland, and Y. Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411, 2005.