Infinite-Horizon Gaussian Processes

Arno Solin
Aalto University
arno.solin@aalto.fi

James Hensman
PROWLER.io
james@prowler.io

Richard E. Turner
University of Cambridge
ret26@cam.ac.uk

Aalto University

PROWLER.io the decision company

UNIVERSITY OF CAMBRIDGE

Introduction

- Gaussian process models provide a **plug & play** interpretable approach to **probabilistic modelling**
- Naïve implementations of GPs require the construction and decomposition of a kernel matrix at cost $\mathcal{O}(n^3)$, where n is the number of data
- We consider **GP time series** (one input dimension)
- We exploit the (approximate) Markov structure of the process and re-write the model as a linear Gaussian state space model
- Inference by Kalman filtering costs $\mathcal{O}(m^3 n)$, where m is the dimension of the state space
- We propose the **Infinite-Horizon GP** approximation (IHGP) which reduces the cost to $\mathcal{O}(m^2 n)$
- We further extend the model to run on **streams of data** and learn the kernel hyperparameters **on the fly**

State Space GPs

- Consider a GP model admitting the form:

$$f(t) \sim \mathrm{GP}(0, \kappa(t,t')) \qquad \text{prior}$$
$$\mathbf{y} \mid \mathbf{f} \sim \textstyle\prod_{i=1}^{n} p(y_i \mid f(t_i)) \qquad \text{likelihood}$$

where (t_i, y_i) are the n input–output pairs

- A naïve solution would scale as $\mathcal{O}(n^3)$
- For **Markovian** covariance functions, $\kappa(t,t')$, an equivalent formulation can be given in terms of **stochastic differential equations** (SDE, see [2]):

$$\dot{\mathbf{f}}(t) = \mathbf{F}\,\mathbf{f}(t) + \mathbf{L}\,\mathbf{w}(t) \qquad \text{prior}$$
$$y_i \sim p(y_i \mid \mathbf{h}^\mathsf{T}\,\mathbf{f}(t_i)) \qquad \text{likelihood}$$

- Can be written as a discrete-time **state space model**:

$$\mathbf{f}_i \sim \mathrm{N}(\mathbf{A}_{i-1}\mathbf{f}_{i-1}, \mathbf{Q}_{i-1}) \qquad \text{prior}$$
$$y_i \sim p(y_i \mid \mathbf{h}^\mathsf{T}\mathbf{f}_i) \qquad \text{likelihood}$$

- This model can be solved by Kalman filtering in $\mathcal{O}(m^3 n)$, where m is the dimensionality of $\mathbf{f}_i$ (see [2, 3])
- The state dimension m is typically small, but grows quickly if the GP prior is complicated—especially when involving **sums and products** of several kernels

Infinite-Horizon GPs

- We leverage the idea of **steady-state filtering**, where the solution filter is seen to reach a steady state when $t \to \infty$
- The steady state is solved by **Discrete Algebraic Riccati Equations** (DAREs)
- After the initial setup cost, the Infinite-Horizon GP scales as $\mathcal{O}(m^2 n)$
- The memory scaling is **linear** in the number of data and state dimension, $\mathcal{O}(mn)$
- The infinite-horizon approximation introduces **biases near the boundaries** (first/last samples) of data

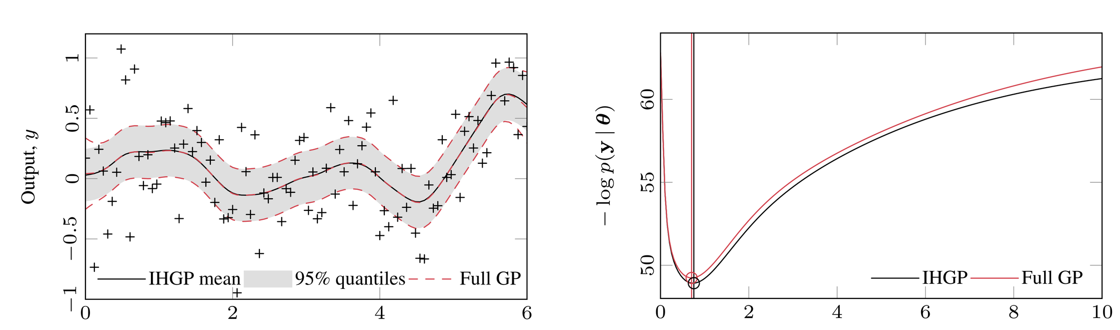


Figure: (Left) GP regression with $n = 100$ observations and a Matérn covariance function. The IHGP is close to exact far from boundaries, where the constant marginal variance assumption shows. (Right) The negative marginal likelihood curves as a function of length-scale.

Non-Gaussian Likelihoods

- For non-Gaussian likelihoods, we leverage **Single-sweep Expectation propagation** (EP) [4]
- Also known as **Assumed density filtering**
- Only requires visiting each data point once
- In IHGP, the computational efficiency comes from matching a likelihood variance parameter by moment matching
- The matched parameters are used for finding the corresponding steady state by cubic convolutional interpolation
- Directly applicable to **streaming** applications

Codes

Watch video

- Codes: `https://github.com/AaltoML/IHGP`
- Video abstract: `https://youtu.be/myCvUT3XGPc`

Online Learning of Hyperparameters

- **Hyperparameter learning** as incremental gradient descent
- Resembling **stochastic gradient descent** without the assumption of finding a stationary optimum
- The 'mini-batches' are windows of recent data
- The infinite-horizon method guarantees that there are **no boundary effects** related to choosing the batch

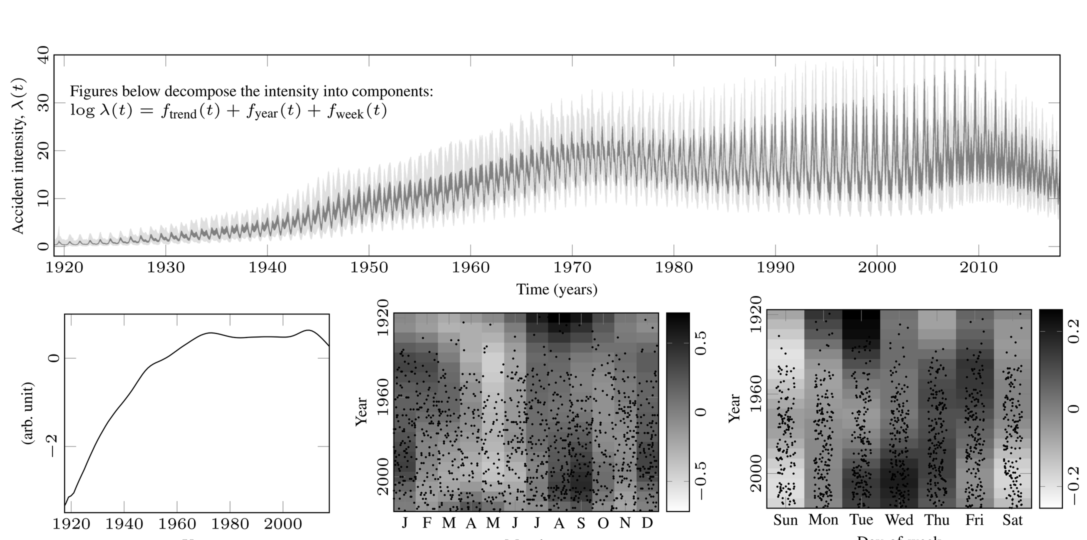
References

[1] A. Solin, J. Hensman, and R. E. Turner (2018). Infinite-horizon Gaussian processes. *Advances in Neural Information Processing Systems (NeurIPS).*

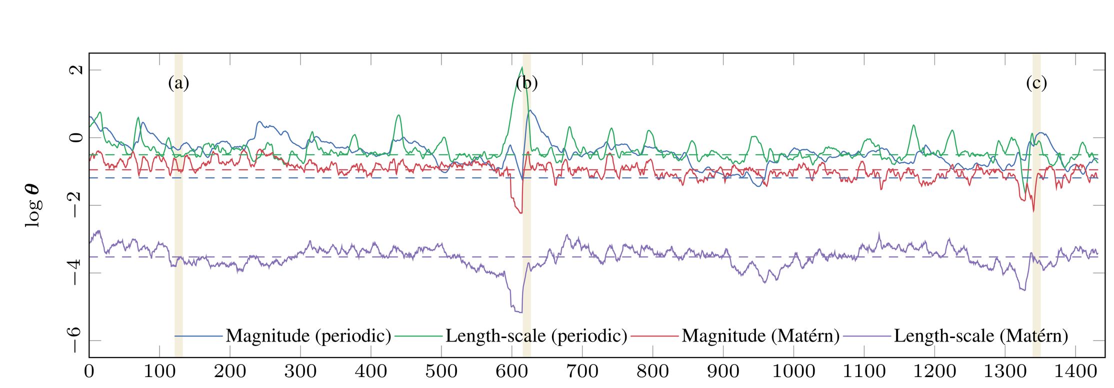
[2] S. Särkkä and A. Solin (in press). *Applied Stochastic Differential Equations.* Cambridge University Press, Cambridge.

[3] H. Nickisch, A. Solin, and A. Grigorievskiy (2018). State space Gaussian processes with non-Gaussian likelihood. *International Conference on Machine Learning (ICML).*

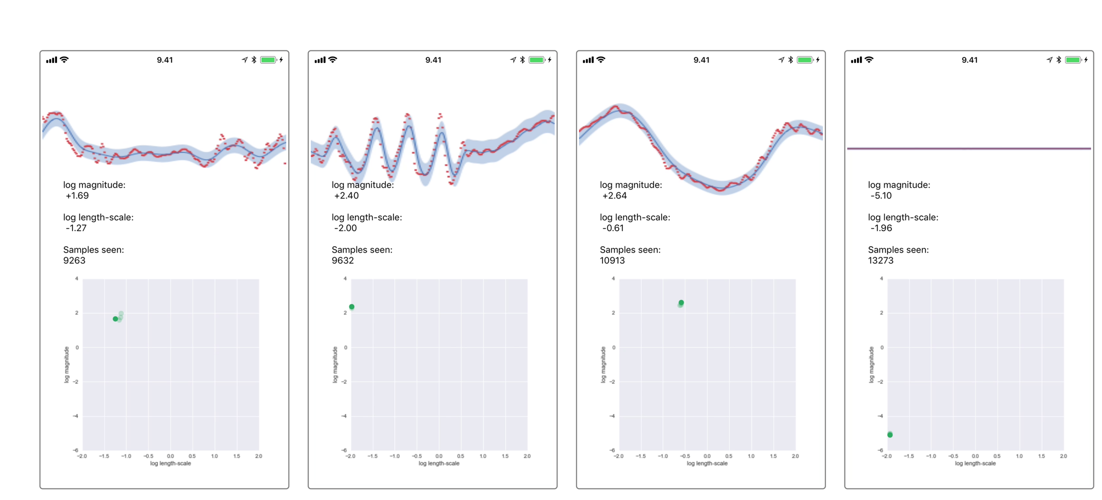
[4] L. Csató and M. Opper (2002). Sparse on-line Gaussian processes. *Neural Computation*, 14(3):641–668.



Experiment 1: Explanatory analysis of the aircraft accident intensity data set ([3], 1210 accidents predicted in $n = 35{,}959$ daily bins) between years 1919–2018 by a log-Gaussian Cox process (Poisson likelihood). We recover a slow trend, and time-varying periodic yearly and weekly variation.



Experiment 2: Results for explorative analysis of electricity consumption data over 1,442 days with one-minute resolution ($n > 2$M). The batch optimized hyperparameters values shown by dashed lines, the results for IHGP with adaptation (solid) adapt to changing circumstances. The model adapts; e.g. in (b) the periodic component is turned off when the house is vacant for a long time.



Experiment 3: Screenshots of online adaptive IHGP running in real-time on an iPhone. The lower plot shows current hyperparameters (measurement noise is fixed to $\sigma_n^2 = 1$ for easier visualization) of the prior covariance function, with a trail of previous hyperparameters. The top part shows the last 2 seconds of accelerometer data (red), the GP mean, and 95% quantiles.

Contact details: arno.solin@aalto.fi, web address: http://arno.solin.fi, NeurIPS/NIPS 2018.