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
ret20@cam.ac.uk

**INTRODUCTION**

- Gaussian process models provide a plug & play interpretable approach to probabilistic modelling.
- Naïve implementations of GPs require construction and decomposition of a kernel matrix at cost $O(n^3)$, 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 $O(mn^2)$, where $m$ is the dimension of the state space.
- We propose the Infinite-Horizon GP approximation (IHGP) which reduces to cost $O(mn)$.
- We further extend the model to run on streams of data and learn the kernel hyperparameters on the fly.

**STATE SPACE GPPS**

- Consider a GP model admitting the form:
  \[ f(t) \sim \text{GP}(0, k(t, t')) \quad \text{prior} \]
  \[ y \sim p(y|f(t)) \quad \text{likelihood} \]
  where $(t, y)$ are the $n$ input-output pairs.
- A naïve solution would scale as $O(n^2)$.
- For Markovian covariance functions, $k(t, t')$, an equivalent formulation can be given in terms of stochastic differential equations (SDE, see [2]):
  \[ f(t) = f(t) + Lw(t) \quad \text{prior} \]
  \[ y(t) \sim p(y|f(t)) \quad \text{likelihood} \]
- Can be written as a discrete-time state space model:
  \[ f_t \sim N(Af_{t-1}, Q) \quad \text{prior} \]
  \[ y_t \sim p(y_t|Hf_t) \quad \text{likelihood} \]
- This model can be solved by Kalman filtering in $O(mn^2)$, where $m$ is the dimensionality of $f$. (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).
- Alter the initial setup cost, the Infinite-Horizon GP scales as $O(mn)$.
- The memory scaling is linear in the number of data and state dimension, $O(mn)$.
- The infinite-horizon approximation introduces biases near the boundaries (first/last samples) of data.

**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.

---

**REFERENCES**


---

**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.

---

Experimt 1: Explanatory analysis of the aircraft accident intensity data set [3], 1,210 accidents predicted in $n \approx 35,959$ 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 ($n \approx 2M$). 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 $\sigma_y^2 = 1$ 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.