Streaming Sparse Gaussian Process Approximations

Thang D. Bui, Cuong V. Nguyen, and Richard E. Turner

Department of Engineering, University of Cambridge

UNIVERSITY OF CAMBRIDGE

Summary

- New framework for GP learning and inference in streaming setting using pseudo-data
- Principled methods for online hyperparameter learning and pseudo-point location optimization
- Natural framework to leverage recent advances in batch GP approximate inference for continual learning and time-series data

Batch Sparse GP Regression

GP Regression: data $\{\mathbf{x}_n, y_n\}_{n=1}^N$

$$y_n = f(\mathbf{x}_n) + \epsilon_n, \qquad \epsilon_n \sim \mathcal{N}(0, \sigma_y^2)$$

VFE Sparse GP Approximation [1]

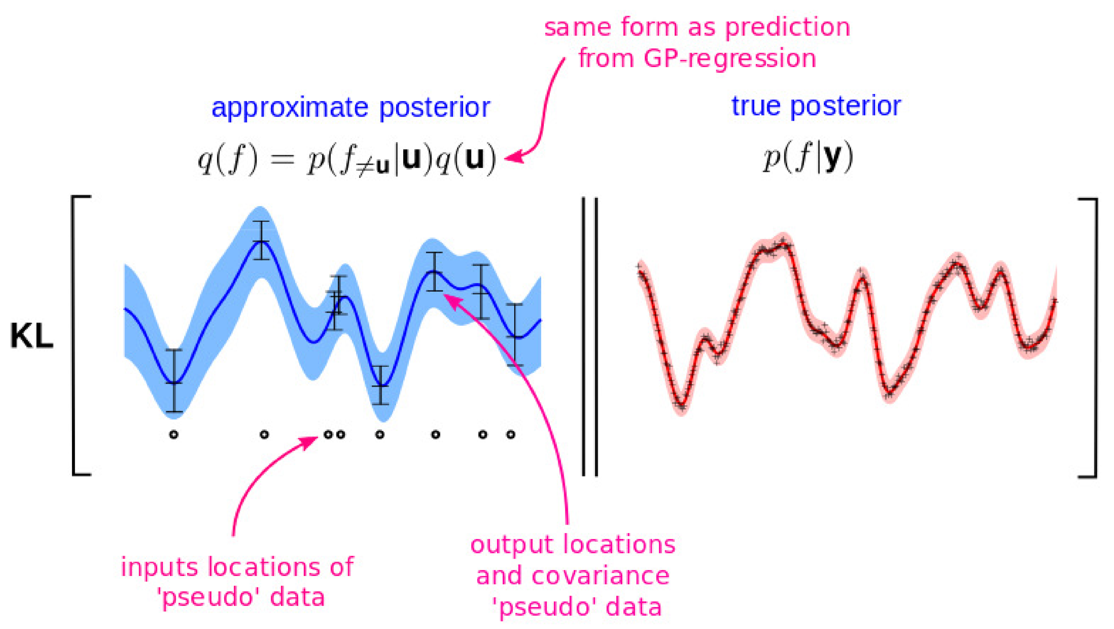
- Parameterize approximate posterior q by pseudo-points $\mathbf{u}$:

$$f = \{f_{\neq \mathbf{u}}, \mathbf{u}\} \qquad q(f) = p(f_{\neq \mathbf{u}}|\mathbf{u}, \theta) q(\mathbf{u})$$

- Lower bound data marginal log-likelihood:

$$\log p(\mathbf{y}|\theta) = \log \int \mathrm{d}f\ p(\mathbf{y}, f|\theta)$$
$$\geq \int \mathrm{d}f\ q(f) \log \frac{p(\mathbf{y}, f|\theta)}{q(f)} = \mathcal{F}_{\mathrm{vfe}}(q, \theta)$$

$$\textbf{Goal:} \max_{q,\theta} \mathcal{F}_{\mathrm{vfe}}(q,\theta) \leftrightarrow \min_{q,\theta} \mathrm{KL}[q(f) \,\|\, p(f|\mathbf{y}, \theta)]$$



Streaming Sparse GP (SSGP)

- Data arrive sequentially
- Can only access current data points $\mathbf{y}_{\mathrm{new}}$
- At each iteration:

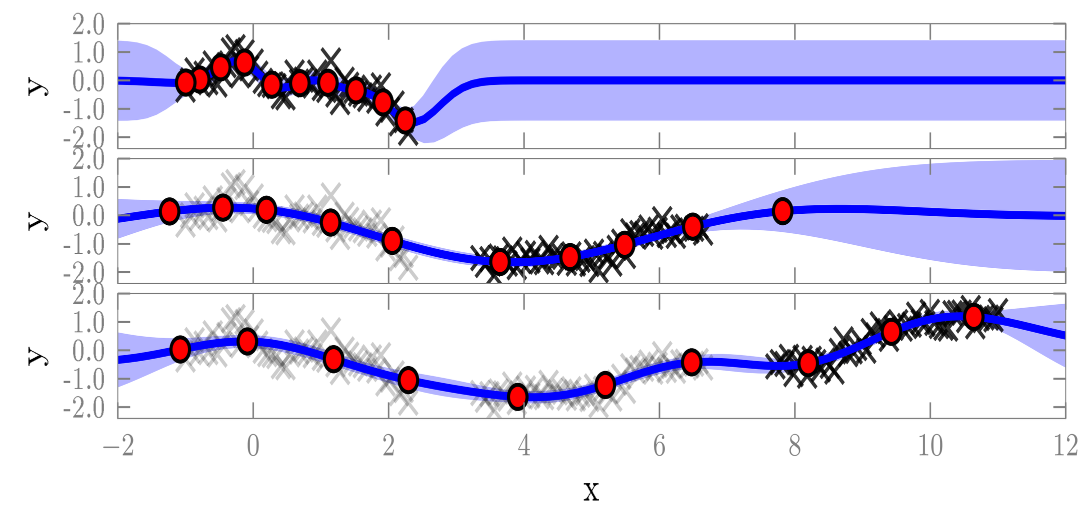
$$q_{\mathrm{old}}(f) \approx p(f|\mathbf{y}_{\mathrm{old}}) \propto p(f|\theta_{\mathrm{old}}) p(\mathbf{y}_{\mathrm{old}}|f)$$
$$q_{\mathrm{new}}(f) \approx p(f|\mathbf{y}_{\mathrm{old}}, \mathbf{y}_{\mathrm{new}}) \propto p(f|\theta_{\mathrm{new}}) p(\mathbf{y}_{\mathrm{old}}|f) p(\mathbf{y}_{\mathrm{new}}|f)$$
$$\hat{p}(f|\mathbf{y}_{\mathrm{old}}, \mathbf{y}_{\mathrm{new}}) \propto p(f|\theta_{\mathrm{new}}) p(\mathbf{y}_{\mathrm{new}}|f) \frac{q_{\mathrm{old}}(f)}{p(f|\theta_{\mathrm{old}})}$$

- $q_{\mathrm{old}}(f)$ and $q_{\mathrm{new}}(f)$ parameterized by pseudo-points

$$\textbf{Goal:} \min_{q_{\mathrm{new}}, \theta_{\mathrm{new}}} \mathrm{KL}[q_{\mathrm{new}}(f) \,\|\, \hat{p}(f|\mathbf{y}_{\mathrm{old}}, \mathbf{y}_{\mathrm{new}})]$$

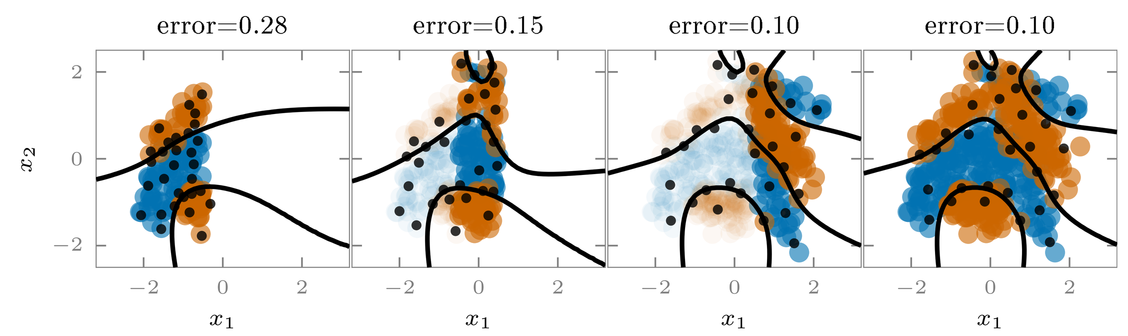
- KL and analytic minimum available at $O(N_{\mathrm{batch}} M^2)$ cost

Behaviour on time-series data



- Pseudo-point locations (red) automatically optimized when observing new data
- Automatic online adaptation of hyperparameters

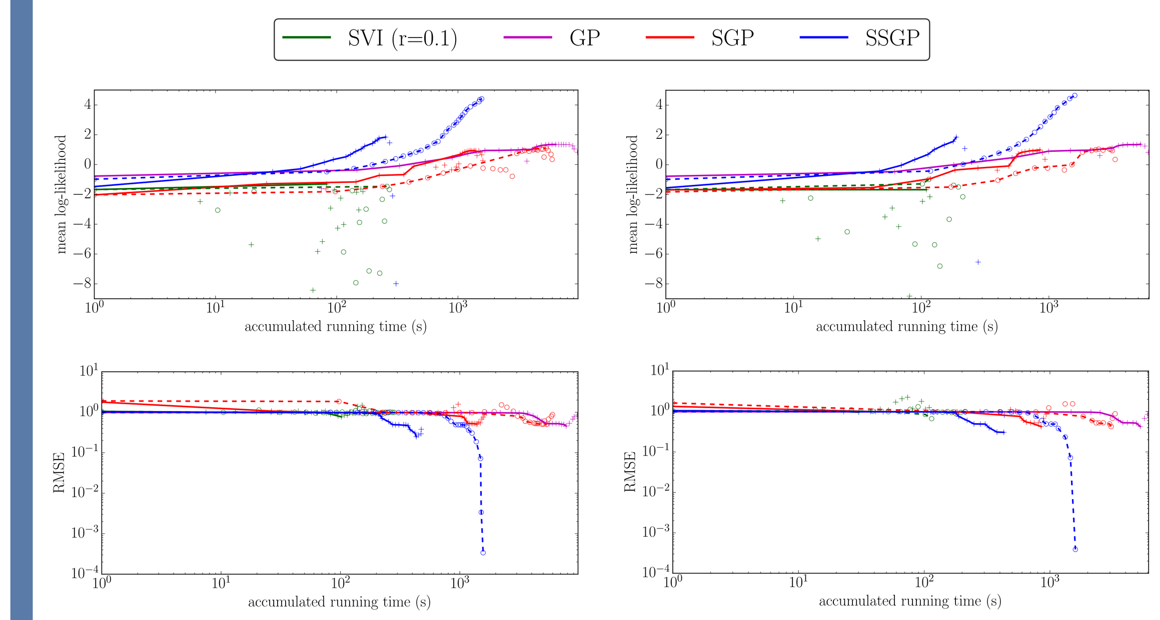
Behaviour on classification problem



Experiments

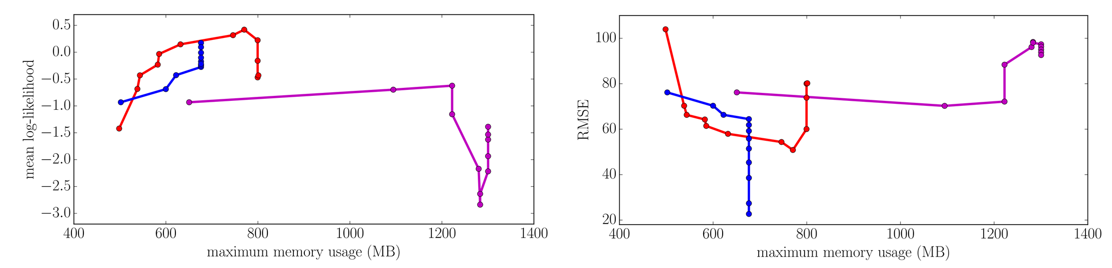
RMSE and log-likelihood vs. running time on audio time-series

- 18,000 data points
- Interleaved training/testing sets
- $M = 100, 200$ pseudo-points



batch size = 300   batch size = 500

Comparison of memory usage

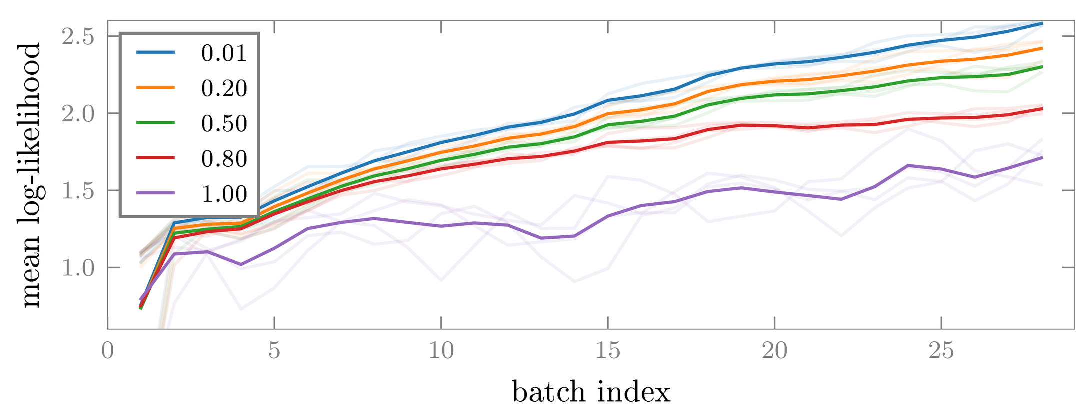


More datasets/experiments in paper

Extension to Power-EP

- Replace KL by α-divergence and optimize locally

Comparing different α values



Discussions

- Smaller α values are better (VFE: $\alpha \to 0$)
  - EP tends to clump pseudo-points together
  - Clumping effect accumulates over time
- Streaming α-divergence approach = run a forward filtering pass of Power-EP
- For fixed hyperparameters & pseudo-points:
  - Streaming α-divergence = batch Power-EP
  - $\alpha = 1$: recover batch ADF [2]
  - $\alpha \to 0$: recover batch VFE [3]
  - Equivalent to streaming VB/online VI [4, 5, 6]

Extensions to Bayesian Deep Neural Networks [7]

- Variational Continual Learning (VCL)
  - Combine online VI with Monte Carlo VI
  - Include episodic memory enhancement
  - General and flexible: applicable to both deep discriminative and deep generative models
- Visit our poster at Bayesian DL workshop!

References

[1] Titsias. Variational learning of inducing variables in sparse Gaussian processes. In *AISTATS*, 2009.

[2] Opper. A Bayesian approach to online learning. In *On-Line Learning in Neural Networks*. 1999.

[3] Csató. *Gaussian Processes – Iterative Sparse Approximations*. PhD thesis, Aston University, 2002.

[4] Broderick et al. Streaming variational Bayes. In *NIPS*, 2013.

[5] Ghahramani and Attias. Online variational Bayesian learning. In *NIPS Workshop*, 2000.

[6] Sato. Online model selection based on the variational Bayes. *Neural Computation*, 2001.

[7] Nguyen et al. Variational continual learning. *arXiv:1710.10628*, 2017.

Acknowledgements This work is supported by EPSRC and Google.