

Efficiently Sampling Functions from Gaussian Process Posteriors

Alexander Terenin*
Imperial College London

Joint work with James T. Wilson*, Viacheslav Borovitskiy*,
Peter Mostowsky*, and Marc Deisenroth

Honorable Mention for Outstanding Paper Award
at the International Conference on Machine Learning 2020

Second Symposium on Machine Learning
and Dynamical Systems, Fields Institute

September 23rd, 2020

[HTTPS://AVT.IM/](https://AVT.IM/) ·  @AVT_IM



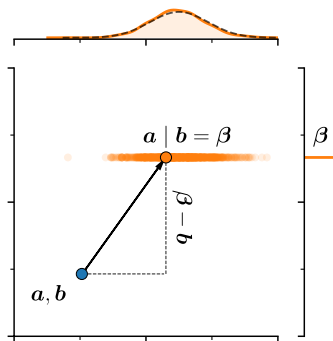
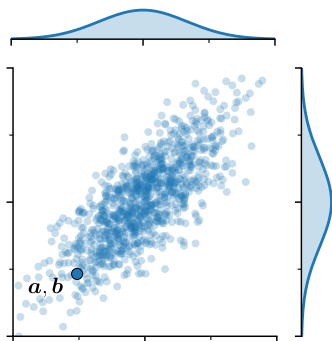
**Imperial College
London**

*Equal contribution

Matheron's update rule

For $\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix}\right)$ we have

$$(\mathbf{f}_1 \mid \mathbf{f}_2 = \mathbf{u}) \stackrel{d}{=} \mathbf{f}_1 + \mathbf{K}_{12}\mathbf{K}_{22}^{-1}(\mathbf{u} - \mathbf{f}_2)$$



Path-wise sampling with sparse GPs

This expression *lifts* to a path-wise characterization of posterior GPs

$$\underbrace{(f \mid \mathbf{y})(\cdot)}_{\text{posterior}} \stackrel{d}{=} \underbrace{f(\cdot)}_{\text{prior}} + \underbrace{\mathbf{K}_{(\cdot)x} \mathbf{K}_{xx}^{-1} (\mathbf{y} - f(\mathbf{x}))}_{\text{update}}$$

Prior term: discretize with random Fourier features

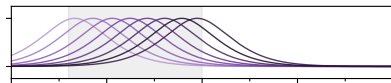
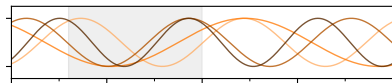
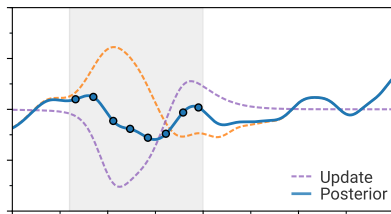
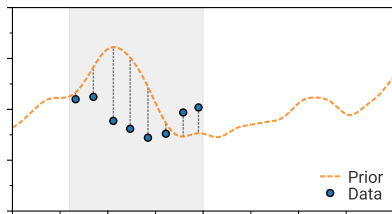
Data term: approximate with sparse GPs

$$\underbrace{(f \mid \mathbf{y})(\cdot)}_{\text{approximate posterior}} \approx \underbrace{\sum_{i=1}^{\ell} w_i \phi_i(\cdot)}_{\text{RFF basis for stationary prior}} + \underbrace{\sum_{i=1}^m v_i k(\cdot, z_i)}_{\text{canonical basis for sparse update}} \quad \mathbf{v} = \mathbf{K}_{zz}^{-1} (\mathbf{u} - \Phi^T \mathbf{w})$$

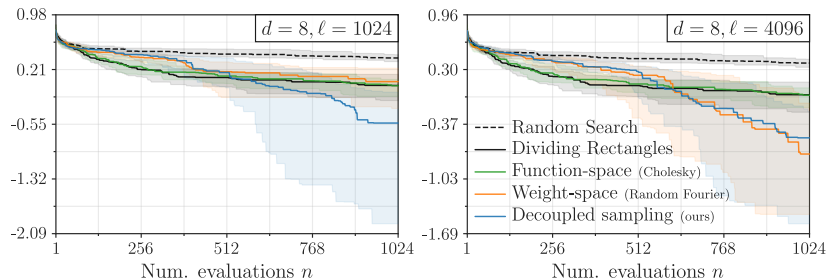
Visualizing path-wise sampling

$$(f | \mathbf{y})(\cdot) \stackrel{d}{\approx} \underbrace{\sum_{i=1}^{\ell} w_i \phi_i(\cdot)}_{\text{RFF basis for stationary prior}} + \underbrace{\sum_{i=1}^m v_i k(\cdot, z_i)}_{\text{canonical basis for sparse update}} \quad \mathbf{v} = \mathbf{K}_{zz}^{-1}(\mathbf{u} - \Phi^T \mathbf{w})$$

approximate posterior

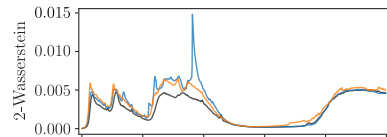
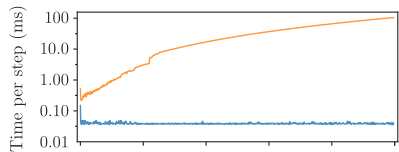
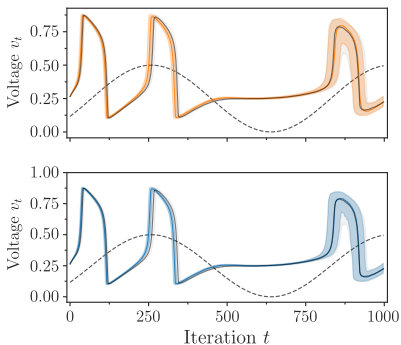
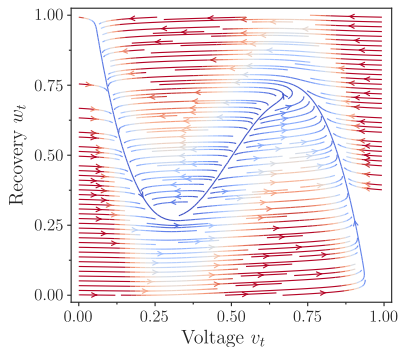


Bayesian optimization: Thompson sampling



Improved performance owing to smaller error

FitzHugh-Nagumo model neuron dynamical system



Significantly more efficient time-stepping