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

Imperial College Mathematics Virtual Symposium

July 29th, 2020

Imperial College London

HTTPS://AVT.IM/ · @AVT_IM

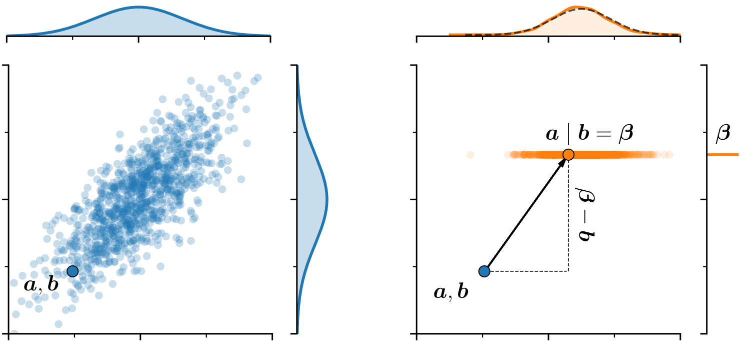
UCL

*Equal contribution

Matheron's update rule

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

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



Path-wise sampling with sparse GPs

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

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

Prior term: discretize with random Fourier features

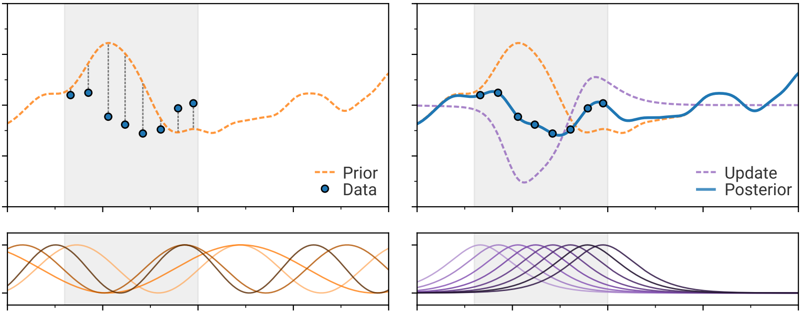
Data term: approximate with sparse GPs

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

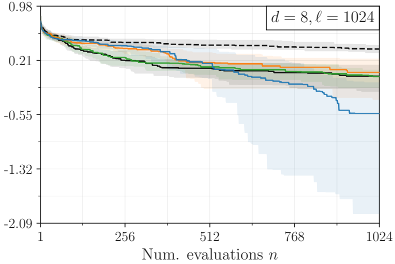
Visualizing path-wise sampling

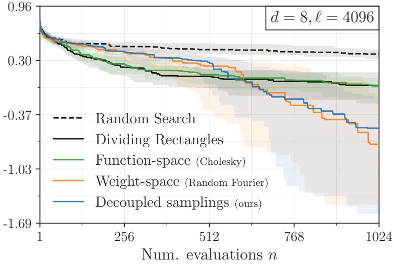
$$(f \mid \boldsymbol{y})(\cdot) \overset{\mathrm{d}}{\approx} \underbrace{\sum_{i=1}^{\ell} w_i \phi_i(\cdot)}_{\text{RFF basis for stationary prior}} + \underbrace{\sum_{i=1}^{m} v_i k(\cdot, z_i)}_{\text{canonical basis for sparse update}} \qquad \boldsymbol{v} = \mathbf{K}_{\boldsymbol{z}\boldsymbol{z}}^{-1}(\boldsymbol{u}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

approximate posterior



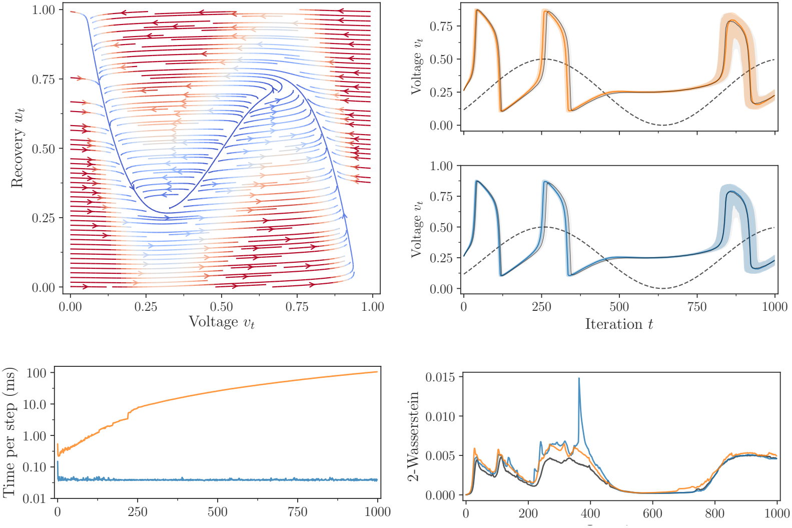
Bayesian optimization: Thompson sampling





Improved performance owing to smaller error

FitzHugh-Nagumo model neuron dynamical system



Significantly more efficient time-stepping