Path-wise, spectral, and geometric perspectives on Gaussian processes

Alexander Terenin
Imperial College London

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

Talk for University of Oxford

October 5th, 2020

HTTPS://AVT.IM/ · @AVT_IM

Imperial College London

UCL

*Equal contribution

Efficiently sampling functions from Gaussian process posteriors

James T. Wilson,* Viacheslav Borovitskiy,* Alexander Terenin,*
Peter Mostowsky,* and Marc Deisenroth

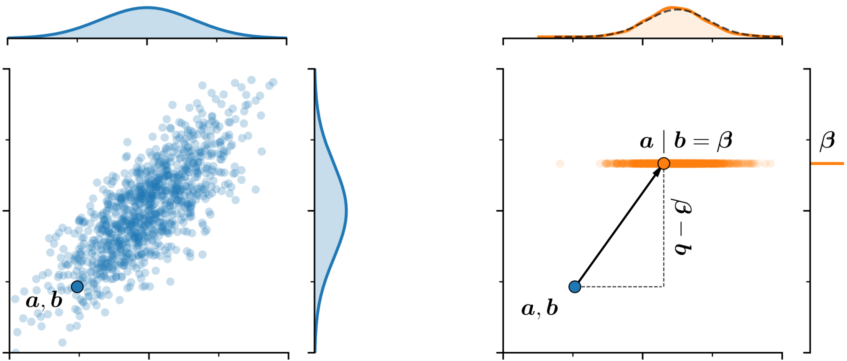



ICML 2020
Honorable Mention for Outstanding Paper Award

*Equal contribution

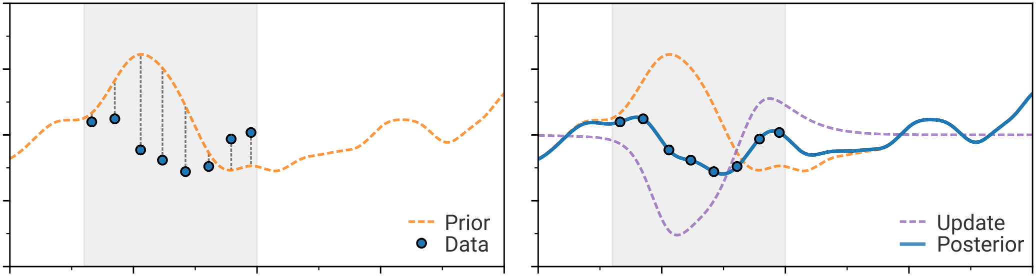
Matheron's update rule

$$\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{a}} \\ \boldsymbol{\mu}_{\boldsymbol{b}} \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{\boldsymbol{aa}} & \mathbf{K}_{\boldsymbol{ab}} \\ \mathbf{K}_{\boldsymbol{ba}} & \mathbf{K}_{\boldsymbol{bb}} \end{bmatrix} \right) \implies (\boldsymbol{a} \mid \boldsymbol{b} = \boldsymbol{\beta}) \overset{\mathrm{d}}{=} \boldsymbol{a} + \mathbf{K}_{\boldsymbol{ab}} \mathbf{K}_{\boldsymbol{bb}}^{-1} (\boldsymbol{\beta} - \boldsymbol{b})$$



Path-wise sampling for Gaussian processes

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



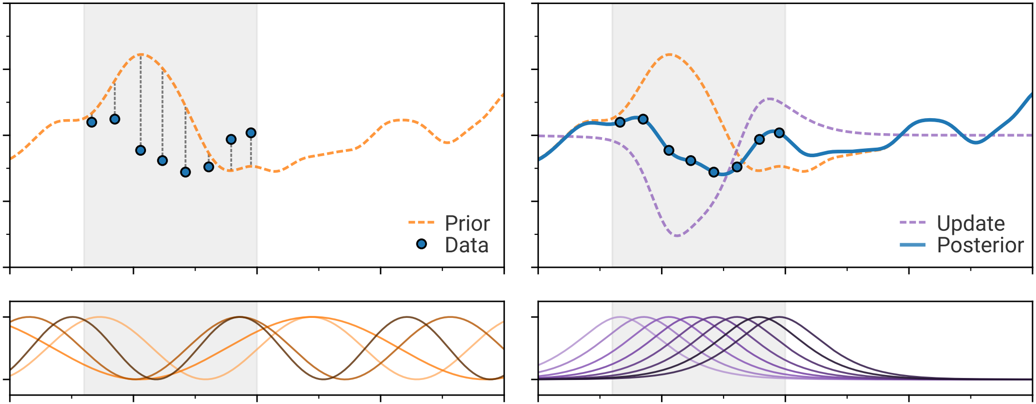
Cubic training costs: $\mathbf{K}_{\boldsymbol{x}\boldsymbol{x}}^{-1}$ — Cubic evaluation costs: $f(\cdot)$

A sparse path-wise approximation

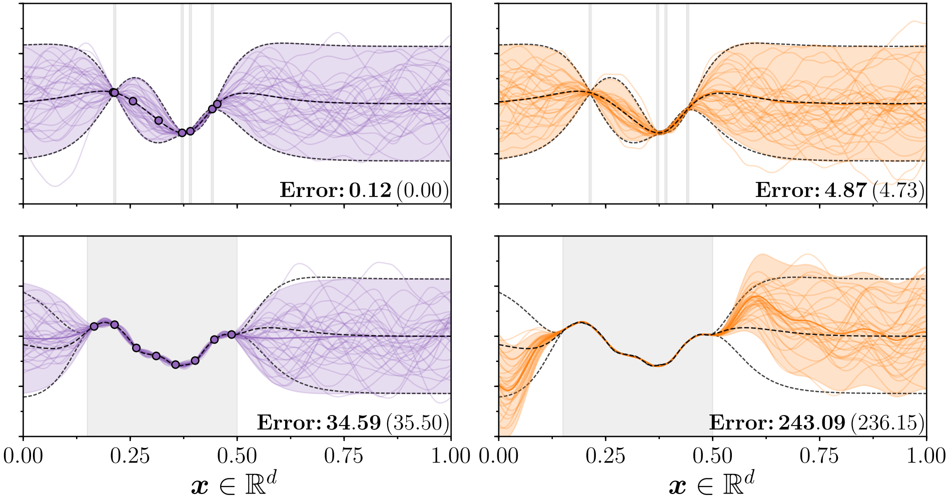
$$\underbrace{(f \mid \boldsymbol{y})(\cdot)}_{\text{approximate posterior}} \overset{\mathrm{d}}{\approx} \underbrace{\sum_{i=1}^{\ell} w_i \phi_i(\cdot)}_{\text{RFF basis for stationary prior}} + \underbrace{\sum_{i=1}^{m} v_i k(\cdot, z_i)}_{\text{canonical basis for sparse update}}$$

$$\boldsymbol{v} = \mathbf{K}_{\boldsymbol{z}\boldsymbol{z}}^{-1}(\boldsymbol{u} - \boldsymbol{\Phi}^T \boldsymbol{w})$$

Linear evaluation costs



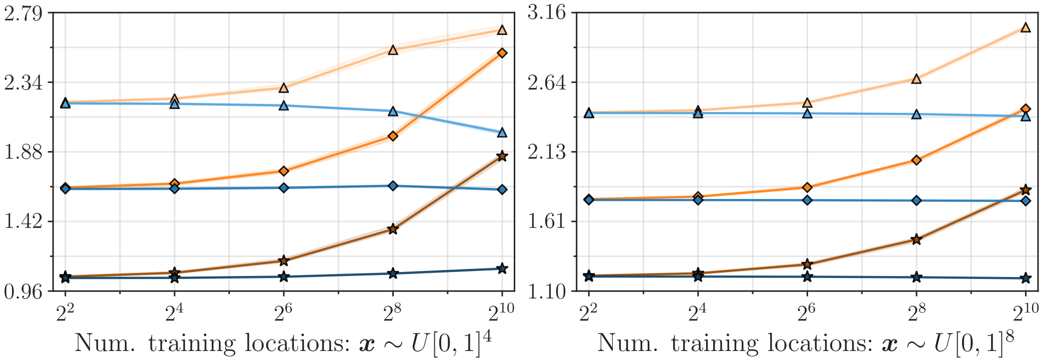
Uncertainty



✓ Avoids variance starvation

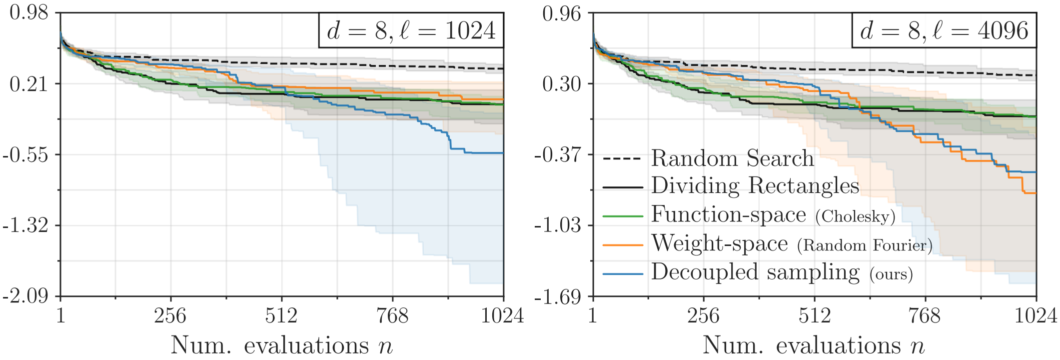
Error analysis

$$\underbrace{W_{2,L^2(\mathcal{X})}(f^{(d)}, f \mid \boldsymbol{y})}_{\text{total approximation error}} \leq \underbrace{W_{2,L^2(\mathcal{X})}(f^{(s)}, f \mid \boldsymbol{y})}_{\text{error in sparse posterior}} + \underbrace{C\, W_{2,L^2(\mathcal{X})}(f^{(w)}, f)}_{\text{error in approximate prior}}$$



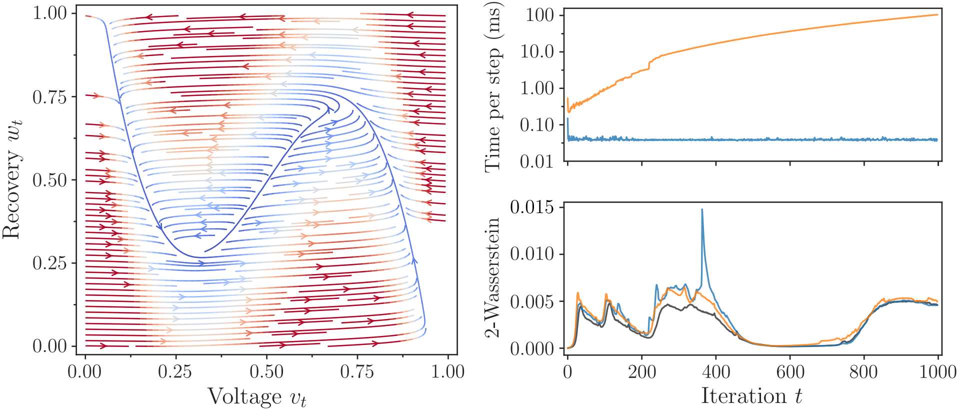
✓ Empirical Wasserstein error smaller than for RFF

Bayesian optimization: Thompson sampling



✓ Improved performance owing to smaller error

FitzHugh-Nagumo model neuron dynamical system



✓ Low approximation error

✓ Significantly more efficient time-stepping

Matérn Gaussian processes on Riemannian manifolds

Viacheslav Borovitskiy,* Alexander Terenin,* Peter Mostowsky,* and Marc Deisenroth

*Equal contribution

NeurIPS 2020

Matérn Gaussian processes on Riemannian manifolds

$$k(x,x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{\|x-x'\|}{\kappa}\right)^{\nu} K_\nu \left(\sqrt{2\nu}\frac{\|x-x'\|}{\kappa}\right)$$

σ^2: variance κ: length scale ν: smoothness

$\nu \to \infty$: recovers square exponential kernel

This defines GPs $f : \mathbb{R}^d \to \mathbb{R}$

What about $f : M \to \mathbb{R}$ where M is a Riemannian manifold?

A candidate generalization for $\nu \to \infty$ via geodesics

$$k_{\text{naïve}}(x, x') = \sigma^2 \exp\left(-\frac{d_g(x, x')^2}{2\kappa^2}\right)$$

Theorem. (Feragen et al.) Let M be a complete Riemannian manifold without boundary. If $k_{\text{naïve}}$ is a positive semi-definite kernel for all κ, then M is isometric to a Euclidean space.

Need a different candidate generalization

Matérn Gaussian processes as solutions of stochastic PDEs

$$\underbrace{\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{\frac{\nu}{2}+\frac{d}{4}} f = \mathcal{W}}_{\text{Matérn}} \qquad \underbrace{e^{-\frac{\kappa^2}{4}\Delta} f = \mathcal{W}}_{\text{Squared Exponential}}$$

Δ : Laplacian $\quad \mathcal{W}$: (rescaled) white noise $\quad e^{-\frac{\kappa^2}{4}\Delta}$: (rescaled) heat semigroup

- GP analog of $\boldsymbol{f} = \mathbf{L}\boldsymbol{z}$ where $\boldsymbol{z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$

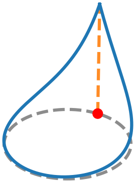
✓ Generalizes well to the Riemannian setting

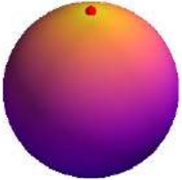
☒ Not very constructive, requires solving SPDEs

This work: compute the kernel, enable training via inducing point methods

A candidate generalization via stochastic PDEs

What do these kernels look like? $(\nu = 1/2)$



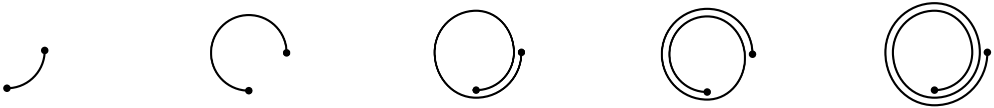


What's wrong with the geodesic definition?

Torus:

$$k(x, x') = \sum_{n \in \mathbb{Z}^d} \sigma^2 \exp\left(-\frac{\|x - x' + n\|^2}{2\kappa^2}\right)$$

(up to a pair of constants)



Similar to naïve generalization, but with extra terms

Riemannian Matérn kernels on compact spaces

General case: $$k_\nu(x,x') = \frac{\sigma^2}{C_\nu} \sum_{n=0}^{\infty} \left(\frac{2\nu}{\kappa^2} - \lambda_n \right)^{\nu - \frac{d}{2}} f_n(x) f_n(x')$$ (λ_n, f_n) : Laplace–Beltrami eigenpairs

Sphere: $$k_\nu(x,x') = \frac{\sigma^2}{C_\nu} \sum_{n=0}^{\infty} c_{n,d}\, \rho_\nu(n)\, \mathcal{C}_n^{(d-1)/2}\Big(\cos\big(d_g(x,x')\big)\Big)$$

$c_{n,d}$: explicit constants

C_ν : normalization constant

$\mathcal{C}_n^{(\cdot)}$: Gegenbauer polynomials

$\rho_\nu(n)$: generalized spectral measure

Posterior samples from Riemannian Matérn GPs

(a) Ground Truth (b) Posterior Mean (c) Standard Deviation (d) One posterior sample

✓ Train by sampling from the prior and using pathwise formula

✓ Does not require repeated numerical SPDE solves

Concluding remarks

Thank you for your attention!

HTTPS://AVT.IM/

@AVT_IM

Imperial College London

UCL

J. T. Wilson,* V. Borovitskiy,* A. Terenin,* P. Mostowsky,* M. P. Deisenroth. Efficiently Sampling Functions from Gaussian Process Posteriors. International Conference on Machine Learning, 2020. *Equal contribution.

V. Borovitskiy,* A. Terenin,* P. Mostowsky,* M. P. Deisenroth. Matérn Gaussian Processes on Riemannian Manifolds. Advances in Neural Information Processing Systems, 2020. *Equal contribution.