Pathwise, spectral, and geometric perspectives on Gaussian processes

Alexander Terenin
Imperial College London

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

Talk for Vector Institute

February 5th, 2021

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

Pathwise Conditioning of Gaussian Processes

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

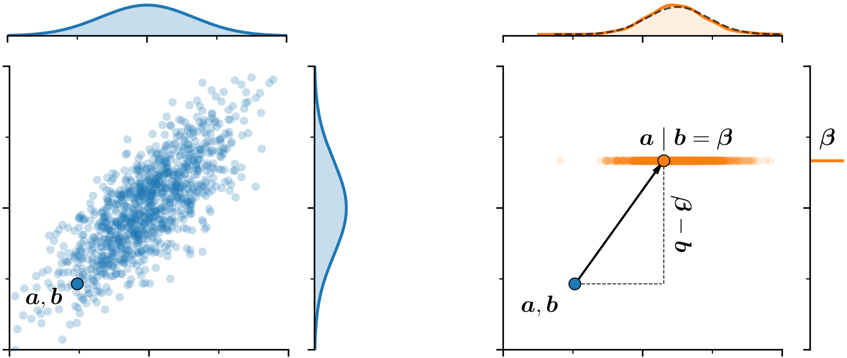



*Equal contribution

JMLR 2021

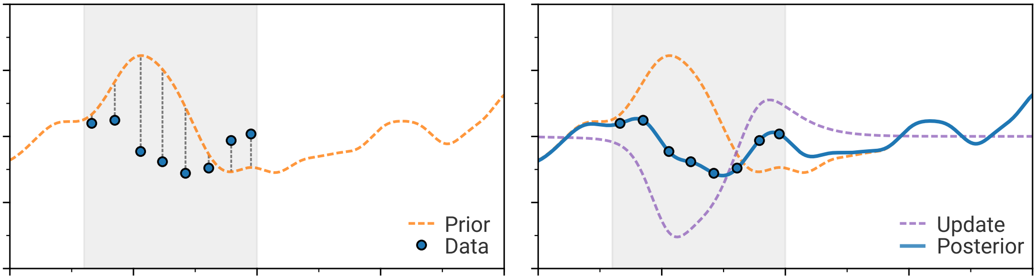
Matheron's update rule

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



Pathwise 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: $f(\cdot)$

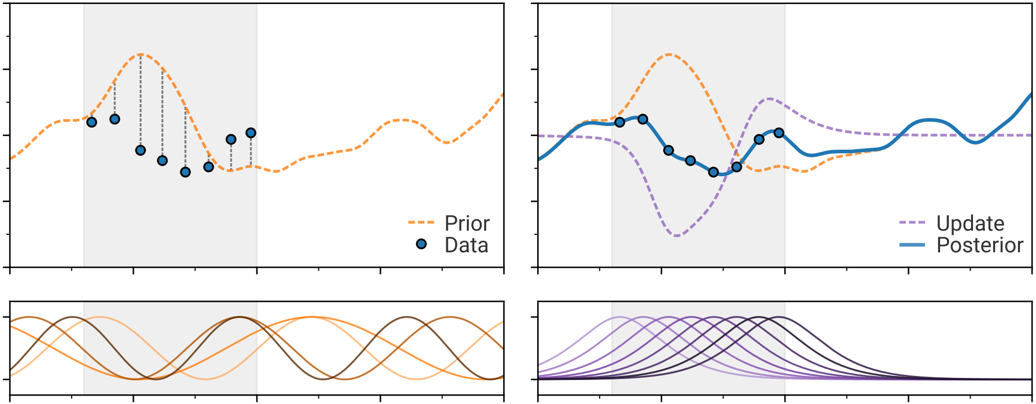
A sparse pathwise approximation

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

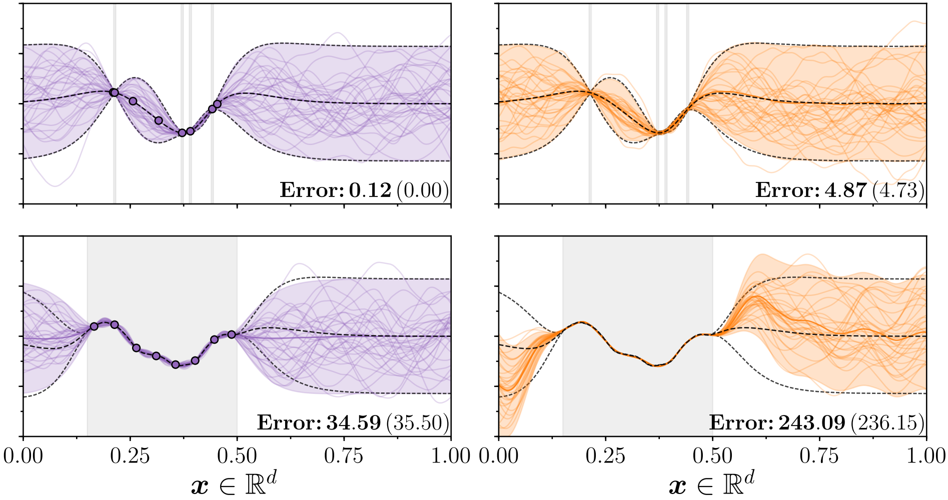
approximate posterior

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

Linear evaluation costs



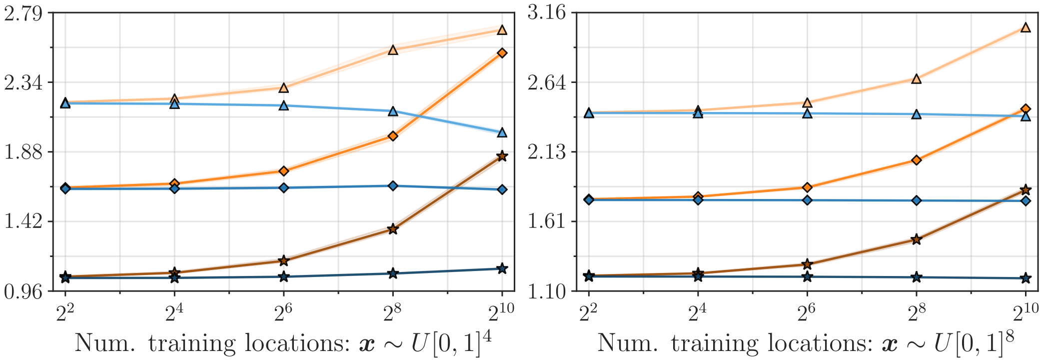
Uncertainty



✓ Avoids variance starvation

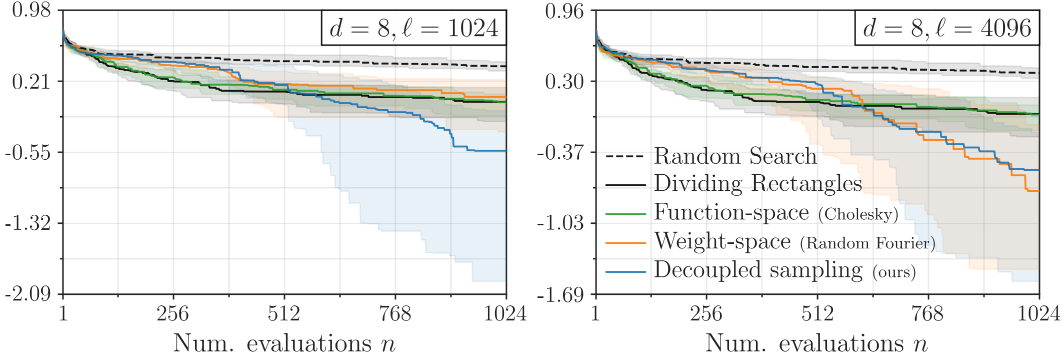
Error analysis

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



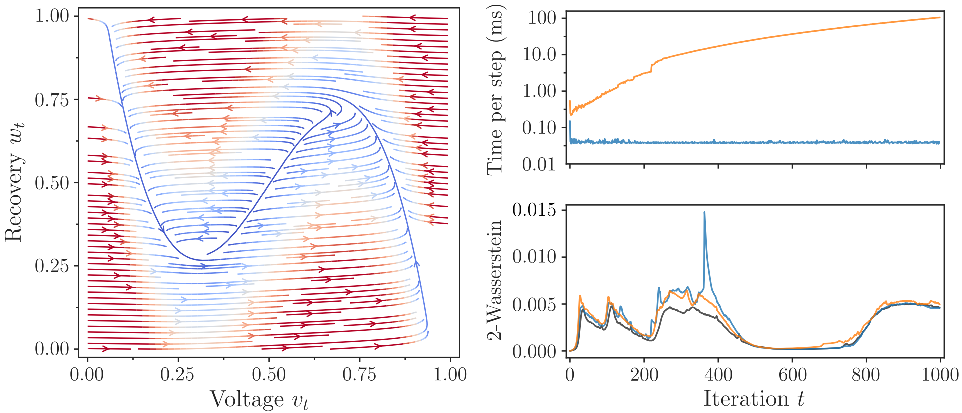
✓ Empirical Wasserstein error smaller than for RFF

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 $\boldsymbol{z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$

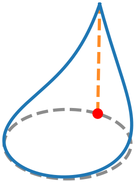
✓ Generalizes well to the Riemannian setting

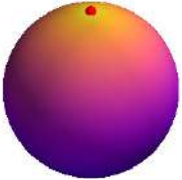
☒ Not very constructive, requires solving SPDEs

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

A candidate generalization via stochastic PDEs

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



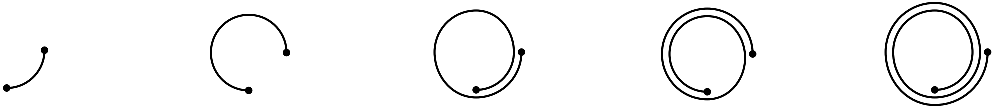


What's wrong with the geodesic definition?

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)



$\|x-x'\|$ $\|x-x'-1\|$ $\|x-x'+1\|$ $\|x-x'-2\|$ $\|x-x'+2\|$

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

$c_{n,d}$: 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

Matérn Gaussian processes on Graphs

Viacheslav Borovitskiy,* Iskander Azangulov,* Alexander Terenin,*
Peter Mostowsky, Marc Deisenroth, and Nicolas Durrande





AISTATS 2021
Oral Presentation

*Equal contribution

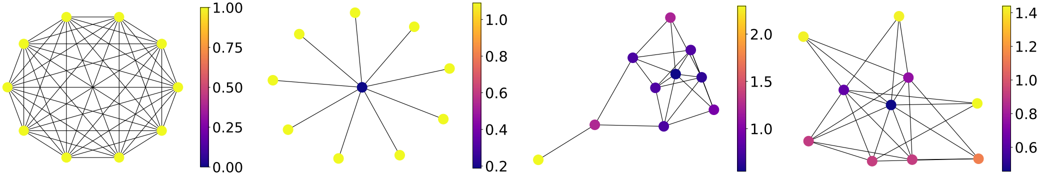
Matérn Gaussian processes via the graph Laplacian

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

$\mathbf{\Delta}$: graph Laplacian $\mathcal{W}$: IID Gaussian $e^{-\frac{\kappa^2}{2}\mathbf{\Delta}}$: (rescaled) graph heat semigroup

✓ Generalizes well to the graph setting

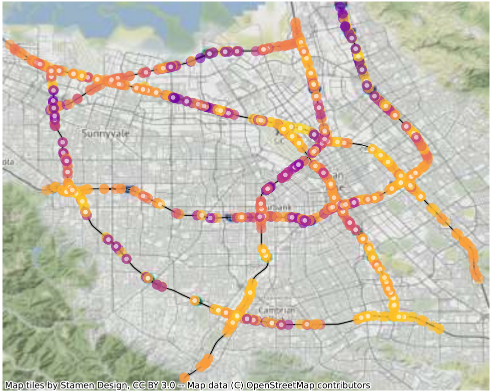
Prior variance



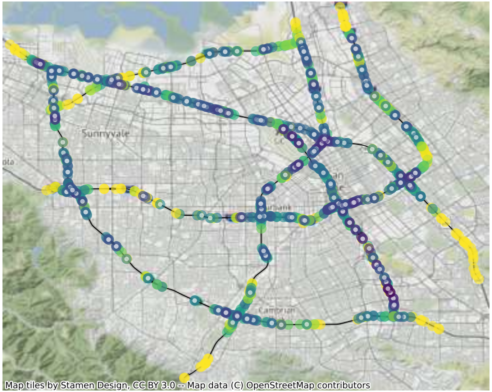
(a) Complete graph (b) Star graph (b) Random graph (b) Random graph

Prior variance depends on geometry

Trained models



(a) Predictive mean



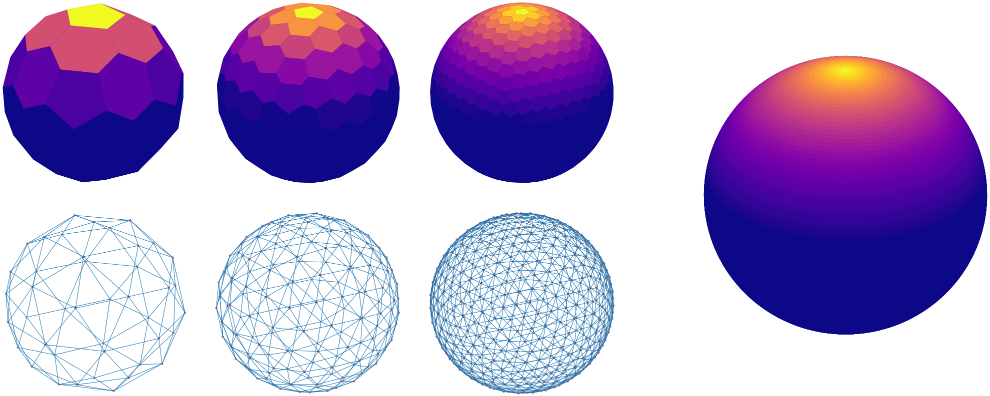
(b) Standard deviation

Trained models

(a) Predictive mean

(b) Standard deviation

Convergence to manifold limit



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.

J. T. Wilson,* V. Borovitskiy,* A. Terenin,* P. Mostowsky,* M. P. Deisenroth. Pathwise Conditioning of Gaussian Processes. Journal of Machine Learning Research, 2021. *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.

V. Borovitskiy,* I. Azangulov,*, A. Terenin,* P. Mostowsky, M. P. Deisenroth, N. Durrande. Matérn Gaussian Processes on Graphs. Artificial Intelligence and Statistics, 2021. *Equal contribution.