Posterior Contraction Rates for Matérn Gaussian Processes on Riemannian Manifolds

Paul Rosa (University of Oxford), Viacheslav Borovitskiy (ETH Zurich), Alexander Terenin (University of Cambridge), Judith Rousseau (University of Oxford)

Problem

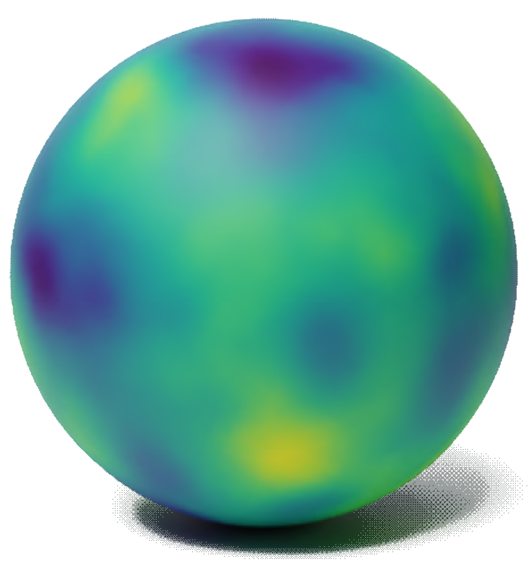
- $\mathcal{M} \subset \mathbb{R}^D$ d-dimensional smooth & compact submanifold
- $f_0 : \mathcal{M} \to \mathbb{R}$ regression function
- Regression model : $y_i = f_0(x_i) + \varepsilon_i$ with $\varepsilon_i \backsim \mathrm{N}(0,\sigma^2)$ iid, $x_i \sim p_0 \cdot \mu$ iid where $\sigma > 0, \mu$ d-dimensional Hausdorff measure on $\mathcal{M}$ and p_0 a lower bounded density
- **Question** : do we need to incorporate geometric informations when designing estimators of f_0 in order to prove good convergence properties ? In the Gaussian process framework, are processes/kernels defined on the whole space $\mathbb{R}^D$ "good enough" ?

Sobolev & Hölder-Zygmund spaces

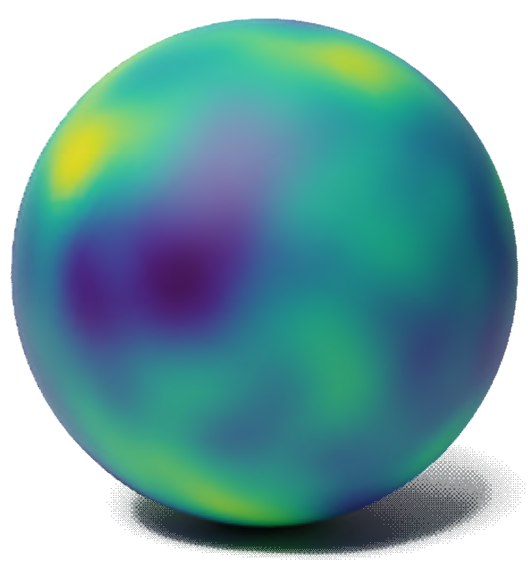
- $\boldsymbol{\Delta} : \boldsymbol{L^2}(\mathcal{M}) \mapsto \boldsymbol{L^2}(\mathcal{M})$ Laplace-Beltrami operator
- $(\lambda_j, \phi_j)_{j\geq 1}$ eigendecomposition of Δ
- $\Theta : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$smooth compactly on [0,2] with $\Theta \equiv 1$ on [0,1]
- $\Theta_j(\sqrt{\Delta})f_0 = \sum_{j=1}^{\infty} \Theta(2^{-j}\sqrt{\lambda_j})\langle\phi_j|f_0\rangle\phi_j$
- Sobolev space : $\|f_0\|^2_{H^\beta(\mathcal{M})} := \sum_{j=1}^{\infty} 2^{2j\beta}\left\|\Theta_j(\sqrt{\Delta})f_0 - f_0\right\|^2_{L^2(\mathcal{M})}$
- Hölder-Zygmund type space : $\|f_0\|^2_{B^\beta_{\infty\infty}(\mathcal{M})} := sup_{j\geq 1} 2^{j\beta}\left\|\Theta_j(\sqrt{\Delta})f_0 - f_0\right\|_{L^\infty(\mathcal{M})}$
- Connections with more classical definitions of Besov spaces on $\mathcal{M}$ exist

Prior processes

- $f = \frac{\sigma_f^2}{C_{\nu,\kappa}} \sum_{j=1}^{J} \left(\frac{2\nu}{\kappa^2} + \lambda_j\right)^{-\frac{\nu+d/2}{2}} Z_j\phi_j, Z_j \sim \mathrm{N}(0,1)$iid the "intrinsic" Matérn process
- Truncation level $J \in \{1, \dots, \infty\}$
- $\tilde{f} \sim GP(0, K_{\nu,\kappa,\sigma_f^2})$ a standard "extrinsic" Matérn process on $\mathbb{R}^D$ i.e. with $K_{\nu,\kappa,\sigma_f^2}(x,y) = \sigma_f^2 \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\frac{\|x-y\|}{\kappa}\right)^\nu K_\nu\left(\sqrt{2\nu}\frac{\|x-y\|}{\kappa}\right)$ where K_ν modified Bessel function of the 2nd kind ; and then we consider the restriction $f = \tilde{f}_{|\mathcal{M}}$



Extrinsic Matérn process sample on S^2



Intrinsic process sample on S^2

Questions :

- Are the intrinsic Matérn processes "better" than the extrinsic ones ?
- For this we study the **asymptotic behaviour of the posterior distributions** of the processes under the **frequentists assumptions** of the nonparametric regression model
- In the Euclidean setting, Van der Vaart & Van Zanten showed that (Euclidean) Matern processes achieve a minimax optimal rates of convergence $n^{-\frac{\beta}{2\beta+d}}$ provided that prior and real smoothness match
- We prove here an analog of this result on the manifold $\mathcal{M}$

Kernel $k(\cdot, y)$ for the extrinsic process on the dragon manifold

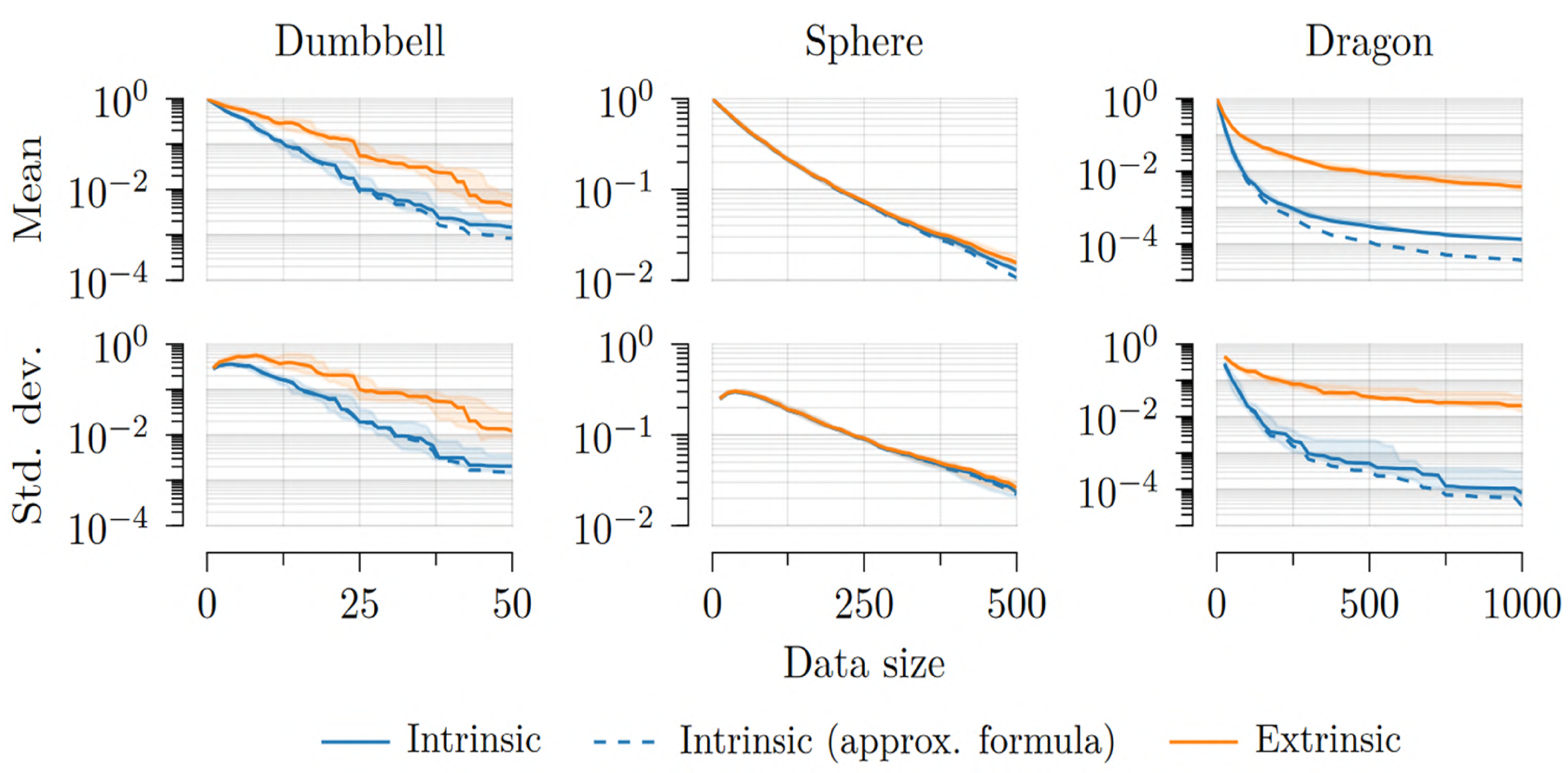
Kernel $k(\cdot, y)$ for the intrinsic process on the dragon manifold

Theorem

- Suppose that $f_0 \in H^\beta(\mathcal{M}) \cap B^\beta_{\infty\infty}(\mathcal{M}), \beta > \frac{d}{2}$
- Take f either an intrinsic or an extrinsic Matérn process with $\nu > d/2$
- We allow a truncation parameter $J \gtrsim n^{\frac{d\min(1,\nu/\beta)}{2\nu+d}}$
- Then for all $q \geq 1, \mathbb{E}_{(x,y)}\Pi\left[\|f - f_0\|^q_{L^2(p_0)}|(x,y)\right] \leq C\left(n^{-\frac{\beta\wedge\nu}{2\nu+d}}\right)^q$ for some $C > 0$
- i.e. an extrinsic Matérn process has the same asymptotic performance as the intrinsic one

Key elements of the proof :

- The key : in both cases the RKHS of the process is norm equivalent to the ("projected") Sobolev space $H^{\nu+d/2}(\mathcal{M}) \cap \{\phi_1, \dots, \phi_J\}$
- For the extrinsic process this follows from the work of Große & al 2013 relating manifold and ambient Sobolev spaces by trace and extension operators
- For the intrinsic process this follows almost directly from its definition and the definition of the Sobolev space
- We prove first an intermediate contraction result with respect to the empirical L^2 distance $\|f - f_0\|_n^2 = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - f_0(x_i))^2$ (i.e. in the *fixed design* setting) and we extrapolate the result using Hölder continuity of the prior process samples
- While this regularity property is known for the extrinsic process, we believe the result is **new** for the intrinsic process



Practical results

- We run experiments computing an approximation of the pointwise worst case expected error for the posterior mean estimator over the unit ball of the RKHS of the process
- We compare three manifolds : the dumbbell, the sphere and the dragon
- We find significant differences between extrinsic and intrinsic processes when the manifold has regions where the geodesic distance is not well approximated by the Euclidean distance (here the dumbbell and the dragon manifolds) ; in particular the error decays non-uniformly in space
- However when the geodesic distance is well approximated by the Euclidean distance uniformly in space (the sphere in our case) then the error decays uniformly in space and the extrinsic or intrinsic processes have similar performances

References

- Castillo & al, *Thomas Bayes' walk on manifolds*, 2013
- Große & al, *Sobolev spaces on Riemannian manifolds with bounded geometry: General coordinates and traces*, 2013
- Van der Vaart & al, *Information Rates of Nonparametric Gaussian Process Methods*, 2011

paul.rosa@stats.ox.ac.uk

paulrosa.owlstown.net

European Research Council
Established by the European Commission

UNIVERSITY OF OXFORD

Jesus College OXFORD