Manifold adaptive regression : insights from the behaviours of Matern processes

joint work with Viacheslav Borovitskiy, Alexander Terenin, Judith Rousseau

September 18, 2023

Geometrical setting

- $\mathcal{M} \subset \mathbb{R}^D$ = smooth compact submanifold without boundary
  $\rightsquigarrow$ **support of our data**

Geometrical setting

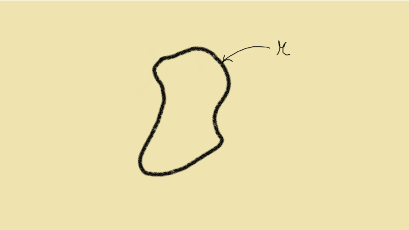
- $\mathcal{M} \subset \mathbb{R}^D$ = smooth compact submanifold without boundary $\rightsquigarrow$ **support of our data**
- $dim(\mathcal{M}) = d < D$: goal = dimension reduction

Geometrical setting

- $\mathcal{M} \subset \mathbb{R}^D$ = smooth compact submanifold without boundary
  $\rightsquigarrow$ **support of our data**
- $dim(\mathcal{M}) = d < D$: goal = dimension reduction
- μ = d-dim Hausdorff measure (normalization : $\mu(\mathcal{M}) = 1$)
  $\rightsquigarrow$ notion of densities wrt μ on $\mathcal{M}$

Geometrical setting

- $\mathcal{M} \subset \mathbb{R}^D$ = smooth compact submanifold without boundary
  $\rightsquigarrow$ **support of our data**
- $dim(\mathcal{M}) = d < D$: goal = dimension reduction
- μ = d-dim Hausdorff measure (normalization : $\mu(\mathcal{M}) = 1$)
  $\rightsquigarrow$ notion of densities wrt μ on $\mathcal{M}$



Statistical model

Statistical model

Regression with random design

$$y_i = f_0(x_i) + \epsilon_i, \epsilon_i \overset{i.i.d}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$

$$x_i \overset{i.i.d}{\sim} p_0 \cdot \mu, i = 1, \ldots, n, p_0 \text{ density wrt } \mu$$

Statistical model

Regression with random design

$$y_i = f_0(x_i) + \epsilon_i, \epsilon_i \overset{i.i.d}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$

$$x_i \overset{i.i.d}{\sim} p_0 \cdot \mu, i = 1, \dots, n, p_0 \text{ density wrt } \mu$$

$\rightsquigarrow$ here for simplicity $\sigma > 0$ is known

Statistical model

Regression with random design

$$y_i = f_0(x_i) + \epsilon_i, \epsilon_i \overset{i.i.d}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$

$$x_i \overset{i.i.d}{\sim} p_0 \cdot \mu, i = 1, \ldots, n, p_0 \text{ density wrt } \mu$$

$\rightsquigarrow$ here for simplicity $\sigma > 0$ is known

question : how can we efficiently estimate f_0 ? In which sense ? How fast ?

General strategy

General strategy

- $\mathcal{X}$ = input space, $f \sim \Pi$ = prior distribution stochastic process = random function

General strategy

- $\mathcal{X}$ = input space, $f \sim \Pi$ = prior distribution stochastic process = random function
- conditional on observing

$$y_i = f_0(x_i) + \epsilon_i \iff y_i - f_0(x_i) \overset{i.i.d}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$

this gives a posterior distribution $\Pi\left[\cdot|\mathbb{X}^n\right]$ on f, $\mathbb{X}^n = (x_i, y_i)_{i=1}^n$

General strategy

- $\mathcal{X}$ = input space, $f \sim \Pi$ = prior distribution stochastic process = random function
- conditional on observing

$$y_i = f_0(x_i) + \epsilon_i \iff y_i - f_0(x_i) \overset{i.i.d}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$

  this gives a posterior distribution $\Pi\left[\cdot|\mathbb{X}^n\right]$ on f, $\mathbb{X}^n = (x_i, y_i)_{i=1}^n$
- bonus : if Π = Gaussian process then $\Pi\left[\cdot|\mathbb{X}^n\right]$ is also Gaussian with explicit parameters

General strategy

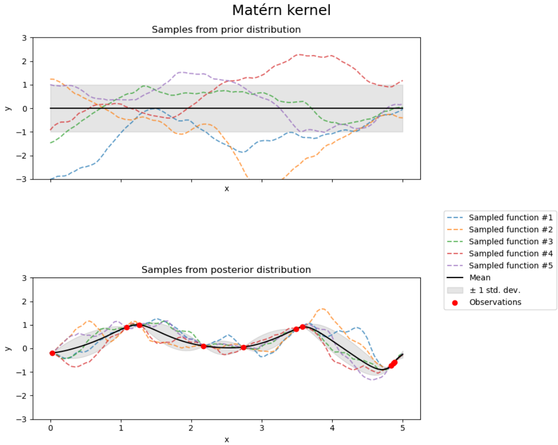
- $\mathcal{X}$ = input space, $f \sim \Pi$ = prior distribution stochastic process = random function
- conditional on observing

$$y_i = f_0(x_i) + \epsilon_i \iff y_i - f_0(x_i) \overset{i.i.d}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$

  this gives a posterior distribution $\Pi\left[\cdot|\mathbb{X}^n\right]$ on f, $\mathbb{X}^n = (x_i, y_i)_{i=1}^n$
- bonus : if Π = Gaussian process then $\Pi\left[\cdot|\mathbb{X}^n\right]$ is also Gaussian with explicit parameters
- as $n \to \infty$ then we expect "contraction" of the posterior $\Pi\left[\cdot|\mathbb{X}^n\right]$ around f_0 (in some sense)

Illustration in 1D

Figure: source : scikit-learn.org



Gaussian process regression

Gaussian process regression

- GPs are powerful Bayesian nonparametric priors for regression

Gaussian process regression

- GPs are powerful Bayesian nonparametric priors for regression
- as Bayesian methods they are especially interesting because they come with a natural notion of uncertainty quantification (sort of)

Gaussian process regression

- GPs are powerful Bayesian nonparametric priors for regression
- as Bayesian methods they are especially interesting because they come with a natural notion of uncertainty quantification (sort of)
- their statistical properties are now well understood through an elegant theory

Gaussian process regression

- GPs are powerful Bayesian nonparametric priors for regression
- as Bayesian methods they are especially interesting because they come with a natural notion of uncertainty quantification (sort of)
- their statistical properties are now well understood through an elegant theory
- there has been recent developments on the construction of GPs on non-Euclidean spaces such as graphs or manifolds

Construction of stochastic processes on $\mathcal{M}$

Construction of stochastic processes on $\mathcal{M}$

For our story :

- if $\mathcal{M}$ = known : take a prior $f \sim \Pi$ defined on $\mathcal{X} = \mathcal{M}$ and condition on the observations

Construction of stochastic processes on $\mathcal{M}$

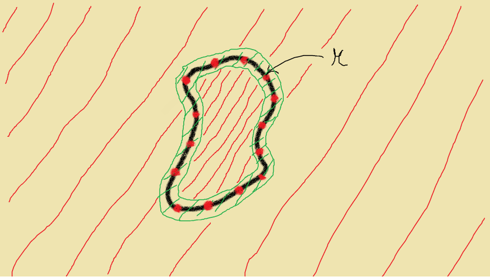
For our story :

- if $\mathcal{M}$ = known : take a prior $f \sim \Pi$ defined on $\mathcal{X} = \mathcal{M}$ and condition on the observations
- if $\mathcal{M} \subset \mathbb{R}^D$ = unknown : take a prior $f \sim \Pi$ defined on $\mathcal{X} = \mathbb{R}^D$ and condition on the observations

Construction of stochastic processes on $\mathcal{M}$

For our story :

- if $\mathcal{M}$ = known : take a prior $f \sim \Pi$ defined on $\mathcal{X} = \mathcal{M}$ and condition on the observations
- if $\mathcal{M} \subset \mathbb{R}^D$ = unknown : take a prior $f \sim \Pi$ defined on $\mathcal{X} = \mathbb{R}^D$ and condition on the observations



$\rightsquigarrow$ we still expect contraction, but on $\mathcal{M}$ only

How to quantify contraction

How to quantify contraction

$\rightsquigarrow$ we want to find posterior contraction rates

$$\mathbb{E}_{(x_i,y_i)\overset{i.i.d}{\sim}P_0} \Pi\left[\|f - f_0\|^2_{L^2(p_0)} \,|\mathbb{X}^n\right] = \mathcal{O}\left(\varepsilon_n^2\right)$$

where

How to quantify contraction

$\rightsquigarrow$ we want to find posterior contraction rates

$$\mathbb{E}_{(x_i,y_i)\overset{i.i.d}{\sim}P_0}\Pi\left[\|f - f_0\|^2_{L^2(p_0)}\,|\mathbb{X}^n\right] = \mathcal{O}\left(\varepsilon_n^2\right)$$

where

- ε_n = some rate of convergence

How to quantify contraction

$\rightsquigarrow$ we want to find posterior contraction rates

$$\mathbb{E}_{(x_i,y_i)\overset{i.i.d}{\sim}P_0}\Pi\left[\|f-f_0\|^2_{L^2(p_0)}\,|\mathbb{X}^n\right]=\mathcal{O}\left(\varepsilon_n^2\right)$$

where

- $\varepsilon_n =$ some rate of convergence
- $P_0(dxdy) = p_0(x)\mu(dx)\mathcal{N}(dy|f_0(x),\sigma^2)$ true frequentist distribution

How to quantify contraction

$\rightsquigarrow$ we want to find posterior contraction rates

$$\mathbb{E}_{(x_i,y_i)\overset{i.i.d}{\sim}P_0}\Pi\left[\|f-f_0\|^2_{L^2(p_0)}\,|\mathbb{X}^n\right]=\mathcal{O}\left(\varepsilon_n^2\right)$$

where

- $\varepsilon_n =$ some rate of convergence
- $P_0(dxdy) = p_0(x)\mu(dx)\mathcal{N}(dy|f_0(x),\sigma^2)$ true frequentist distribution
- $\|f-f_0\|^2_{L^2(p_0)} = \int_{\mathcal{M}} |f(x)-f_0(x)|^2\,p_0(x)\mu(dx)$ choice of metric

Proof of contraction for Gaussian priors

Proof of contraction for Gaussian priors

- mean zero : $f \sim GP(0, K)$, $K =$ a kernel over $\mathcal{M}$ or $\mathbb{R}^D \rightsquigarrow K(x, y) = \mathbb{E}(f(x)f(y))$

Proof of contraction for Gaussian priors

- mean zero : $f \sim GP\,(0, K)\,, K =$ a kernel over $\mathcal{M}$ or $\mathbb{R}^D \rightsquigarrow K(x, y) = \mathbb{E}(f(x)f(y))$
- we usually need extra functional analytic properties on the process in order to prove asymptotic properties $\rightsquigarrow$ "gaussian random element" : fine in what follows

Proof of contraction for Gaussian priors

- mean zero : $f \sim GP(0, K)$, $K =$ a kernel over $\mathcal{M}$ or $\mathbb{R}^D \rightsquigarrow K(x, y) = \mathbb{E}(f(x)f(y))$
- we usually need extra functional analytic properties on the process in order to prove asymptotic properties $\rightsquigarrow$ "gaussian random element" : fine in what follows
- key : with f comes an RKHS $\mathbb{H} =$ completion of $\left\{\sum_{i=1}^{p} a_i K(x_i, \cdot) : p \geq 1, a_i \in \mathbb{R}, x_i \in \mathcal{X}\right\}$ with $\langle K(x, \cdot), K(y, \cdot)\rangle_{\mathbb{H}} = K(x, y)$
  together with f_0, $\mathbb{H}$ essentially dictates the contraction rate

idea : if f_0 is "well approximated" by elements of $\mathbb{H}$ with controlled norms + $\mathbb{H}$ is not too "funky" then contraction happens with good rate

idea : if f_0 is "well approximated" by elements of $\mathbb{H}$ with controlled norms + $\mathbb{H}$ is not too "funky" then contraction happens with good rate

Two parts in the proof (inspired from Van der Vaart & Van Zanten [4]

idea : if f_0 is "well approximated" by elements of $\mathbb{H}$ with controlled norms + $\mathbb{H}$ is not too "funky" then contraction happens with good rate

Two parts in the proof (inspired from Van der Vaart & Van Zanten [4])

1. we prove a contraction rate wrt

$$\|f - f_0\|_n^2 = \frac{1}{n}\sum_{i=1}^{n} |f(x_i) - f_0(x_i)|^2$$

idea : if f_0 is "well approximated" by elements of $\mathbb{H}$ with controlled norms + $\mathbb{H}$ is not too "funky" then contraction happens with good rate

Two parts in the proof (inspired from Van der Vaart & Van Zanten [4])

1. we prove a contraction rate wrt

$$\|f - f_0\|_n^2 = \frac{1}{n}\sum_{i=1}^{n} |f(x_i) - f_0(x_i)|^2$$

2. assuming Holder continuity for f_0 + proving that our prior processes are a posteriori essentially supported on functions with Holder norms "not too big" + a concentration inequality we extrapolate

$$\frac{1}{n}\sum_{i=1}^{n} |f(x_i) - f_0(x_i)|^2 = \mathcal{O}(\varepsilon_n^2) \rightsquigarrow \|f - f_0\|_{L^2(p_0)}^2 = \mathcal{O}(\varepsilon_n^2)$$

Intrinsic Matérn process : Laplace-Beltrami operator

Intrinsic Matérn process : Laplace-Beltrami operator

- Riemannian metric on $\mathcal{M} \rightsquigarrow$ notion of gradient ∇f for every $f \in \mathcal{D}(\mathcal{M})$

Intrinsic Matérn process : Laplace-Beltrami operator

- Riemannian metric on $\mathcal{M}$ $\rightsquigarrow$ notion of gradient ∇f for every $f \in \mathcal{D}(\mathcal{M})$
- define the Laplacian operator :

$$\forall f \in \mathcal{D}(\mathcal{M}), \int_{\mathcal{M}} \Delta(f) f d\mu = \int_{\mathcal{M}} |\nabla f|^2 d\mu$$

$\rightsquigarrow$ generalizes $-\Delta(f)$ in $\mathbb{R}^D$

Intrinsic Matérn process : Laplace-Beltrami operator

- Riemannian metric on $\mathcal{M}$ $\rightsquigarrow$ notion of gradient ∇f for every $f \in \mathcal{D}(\mathcal{M})$
- define the Laplacian operator :

$$\forall f \in \mathcal{D}(\mathcal{M}), \int_{\mathcal{M}} \Delta(f) f d\mu = \int_{\mathcal{M}} |\nabla f|^2 d\mu$$

$\rightsquigarrow$ generalizes $-\Delta(f)$ in $\mathbb{R}^D$

- Then :

$$L^2 = L^2(\mathcal{M}, \mu) = \bigoplus_{j \geq 1} \mathcal{H}_j, \mathcal{H}_j = ker(\Delta - \lambda_j I_{L^2})$$

$$\lambda_j \geq 0, \mathcal{H}_j \subset \mathcal{D}(\mathcal{M}), dim(\mathcal{H}_j) < \infty$$

$\rightsquigarrow$ notion of frequencies/Laplace-Fourier transform

Intrinsic Matern process

Intrinsic Matern process

$\mathcal{M}$ = known

Intrinsic Matern process

$\mathcal{M}$ = known

$$f = \sum_{j \geq 0} (1 + \lambda_j)^{-\frac{s+d/2}{2}} Z_j u_j, Z_j \overset{i.i.d}{\sim} \mathcal{N}(0,1) \rightsquigarrow \text{a Gaussian process}$$

Intrinsic Matern process

$\mathcal{M}$ = known

$$f = \sum_{j \geq 0} (1 + \lambda_j)^{-\frac{s+d/2}{2}} Z_j u_j, Z_j \overset{i.i.d}{\sim} \mathcal{N}(0,1) \rightsquigarrow \text{a Gaussian process}$$

$$\mathbb{H} = H^{s+d/2}(\mathcal{M}) \subset L^2(\mathcal{M})$$

Intrinsic Matern process

$\mathcal{M}$ = known

$$f = \sum_{j \geq 0} (1+\lambda_j)^{-\frac{s+d/2}{2}} Z_j u_j, Z_j \overset{i.i.d}{\sim} \mathcal{N}(0,1) \rightsquigarrow \text{a Gaussian process}$$

$$\mathbb{H} = H^{s+d/2}(\mathcal{M}) \subset L^2(\mathcal{M})$$

$$\|g\|^2_{H^{s+d/2}(\mathcal{M})} = \sum_{j \geq 0} (1+\lambda_j)^{s+d/2} \langle u_j | g \rangle^2_{L^2(\mathcal{M})}$$

Intrinsic Matern process

$\mathcal{M}$ = known

$$f = \sum_{j\geq 0} (1+\lambda_j)^{-\frac{s+d/2}{2}} Z_j u_j, Z_j \overset{i.i.d}{\sim} \mathcal{N}(0,1) \rightsquigarrow \text{a Gaussian process}$$

$$\mathbb{H} = H^{s+d/2}(\mathcal{M}) \subset L^2(\mathcal{M})$$

$$\|g\|^2_{H^{s+d/2}(\mathcal{M})} = \sum_{j\geq 0} (1+\lambda_j)^{s+d/2} \langle u_j|g\rangle^2_{L^2(\mathcal{M})}$$

why "Matern" ? $\rightsquigarrow$ because the Matern GP in $\mathbb{R}^d$ also has an RKHS isometric to $H^{s+d/2}(\mathbb{R}^d)$ + same description as solutions of SPDEs

Posterior contraction

Posterior contraction

proof in action :

Posterior contraction

proof in action :

- $f_0 \in B^{\beta}_{\infty\infty}(\mathcal{M}) \cap H^{\beta}(\mathcal{M}) \implies$ we can control the rate of approximation of f_0 by elements of $\mathbb{H} = H^{s+d/2}(\mathcal{M})$ in sup norm

Posterior contraction

proof in action :

- $f_0 \in B^{\beta}_{\infty\infty}(\mathcal{M}) \cap H^{\beta}(\mathcal{M}) \implies$ we can control the rate of approximation of f_0 by elements of $\mathbb{H} = H^{s+d/2}(\mathcal{M})$ in sup norm
- the space $H^{s+d/2}(\mathcal{M})$ is not too "funky" : follows for instance by chart description of Sobolev spaces and known results on $\mathbb{R}^d$

Posterior contraction

proof in action :

- $f_0 \in B^{\beta}_{\infty\infty}(\mathcal{M}) \cap H^{\beta}(\mathcal{M}) \implies$ we can control the rate of approximation of f_0 by elements of $\mathbb{H} = H^{s+d/2}(\mathcal{M})$ in sup norm
- the space $H^{s+d/2}(\mathcal{M})$ is not too "funky" : follows for instance by chart description of Sobolev spaces and known results on $\mathbb{R}^d$
- we show in our work that f has a posteriori a Holder norm that is "not too big"

Posterior contraction

proof in action :

- $f_0 \in B^{\beta}_{\infty\infty}(\mathcal{M}) \cap H^{\beta}(\mathcal{M}) \implies$ we can control the rate of approximation of f_0 by elements of $\mathbb{H} = H^{s+d/2}(\mathcal{M})$ in sup norm
- the space $H^{s+d/2}(\mathcal{M})$ is not too "funky" : follows for instance by chart description of Sobolev spaces and known results on $\mathbb{R}^d$
- we show in our work that f has a posteriori a Holder norm that is "not too big"

in practice : the **sum needs to be truncated** (handled in our work) + **eigendecomposition needs to be computed** : either algebraically (for manifolds with symmetries : see Azangulov & al [2]) or numerically

Posterior contraction

proof in action :

- $f_0 \in B^{\beta}_{\infty\infty}(\mathcal{M}) \cap H^{\beta}(\mathcal{M}) \implies$ we can control the rate of approximation of f_0 by elements of $\mathbb{H} = H^{s+d/2}(\mathcal{M})$ in sup norm
- the space $H^{s+d/2}(\mathcal{M})$ is not too "funky" : follows for instance by chart description of Sobolev spaces and known results on $\mathbb{R}^d$
- we show in our work that f has a posteriori a Holder norm that is "not too big"

in practice : the **sum needs to be truncated** (handled in our work) + **eigendecomposition needs to be computed** : either algebraically (for manifolds with symmetries : see Azangulov & al [2]) or numerically $\rightsquigarrow$ final rate : $\forall \beta \wedge s > d/2, \varepsilon_n \simeq n^{-\frac{\beta \wedge s}{2s+d}}$

Extrinsic Matern process

Extrinsic Matern process

typically : $\mathcal{M} \subset \mathbb{R}^D =$ unknown

Extrinsic Matern process

typically : $\mathcal{M} \subset \mathbb{R}^D$ = unknown

- g = standard Matern GP on $\mathbb{R}^D \implies \mathbb{H}_g = H^{s+D/2}\left(\mathbb{R}^D\right)$

Extrinsic Matern process

typically : $\mathcal{M} \subset \mathbb{R}^D$ = unknown

- g = standard Matern GP on $\mathbb{R}^D \implies \mathbb{H}_g = H^{s+D/2}\left(\mathbb{R}^D\right)$
- $f = g_{|\mathcal{M}}$ = theoretical object

Extrinsic Matern process

typically : $\mathcal{M} \subset \mathbb{R}^D$ = unknown

- g = standard Matern GP on $\mathbb{R}^D \implies \mathbb{H}_g = H^{s+D/2}(\mathbb{R}^D)$
- $f = g_{|\mathcal{M}}$ = theoretical object

Theorem

$\mathbb{H}_f \simeq H^{s+d/2}(\mathcal{M})$ i.e. $\mathbb{H}_f \equiv H^{s+d/2}(\mathcal{M})$ and

$$\exists C \geq 1, \forall g \in \mathbb{H}_f, C^{-1} \|g\|_{\mathbb{H}_f} \leq \|f\|_{H^{s+d/2}(\mathcal{M})} \leq C \|g\|_{\mathbb{H}_f}$$

Theorem

$\mathbb{H}_f \simeq H^{s+d/2}(\mathcal{M})$ i.e. $\mathbb{H}_f \equiv H^{s+d/2}(\mathcal{M})$ and

$$\exists C \geq 1, \forall g \in \mathbb{H}_f, C^{-1} \|g\|_{\mathbb{H}_f} \leq \|f\|_{H^{s+d/2}(\mathcal{M})} \leq C \|g\|_{\mathbb{H}_f}$$

Proof.

1) Fact :

Theorem

$\mathbb{H}_f \simeq H^{s+d/2}(\mathcal{M})$ i.e. $\mathbb{H}_f \equiv H^{s+d/2}(\mathcal{M})$ and

$$\exists C \geq 1, \forall g \in \mathbb{H}_f, C^{-1} \|g\|_{\mathbb{H}_f} \leq \|f\|_{H^{s+d/2}(\mathcal{M})} \leq C \|g\|_{\mathbb{H}_f}$$

Proof.

1) Fact :

$$\mathbb{H}_f = \left\{g = h_{|\mathcal{M}} : h \in \mathbb{H}_g\right\}, \|g\|_{\mathbb{H}_f} = \inf_{g=h_{|\mathcal{M}}, h \in \mathbb{H}_g} \|h\|_{\mathbb{H}_g}$$

Theorem

$\mathbb{H}_f \simeq H^{s+d/2}(\mathcal{M})$ i.e. $\mathbb{H}_f \equiv H^{s+d/2}(\mathcal{M})$ and

$$\exists C \geq 1, \forall g \in \mathbb{H}_f, C^{-1} \|g\|_{\mathbb{H}_f} \leq \|f\|_{H^{s+d/2}(\mathcal{M})} \leq C \|g\|_{\mathbb{H}_f}$$

Proof.

1) Fact :

$$\mathbb{H}_f = \left\{ g = h_{|\mathcal{M}} : h \in \mathbb{H}_g \right\}, \|g\|_{\mathbb{H}_f} = \inf_{g = h_{|\mathcal{M}}, h \in \mathbb{H}_g} \|h\|_{\mathbb{H}_g}$$

$$\iff \mathbb{H}_f = \left\{ g = h_{|\mathcal{M}} : h \in H^{s+D/2}\left(\mathbb{R}^D\right) \right\}$$

$$\|g\|_{\mathbb{H}_f} = \inf_{g = h_{|\mathcal{M}}, h \in H^{s+D/2}\left(\mathbb{R}^D\right)} \|h\|_{H^{s+D/2}\left(\mathbb{R}^D\right)}$$

□

Proof.

2) But actually by Grosse & Schneider [3]

$$Tr : f \in H^{s+D/2}\left(\mathbb{R}^D\right) \to f_{|\mathcal{M}} \in H^{s+d/2}(\mathcal{M}) = \text{bounded}$$

and we can construct a bounded right inverse $Tr \circ Ex = I_{H^{s+d/2}(\mathcal{M})}$

$$Ex : g \in H^{s+d/2}(\mathcal{M}) \mapsto Ex(g) \in H^{s+D/2}\left(\mathbb{R}^D\right)$$

□

Consequence

Consequence

1. $\Longrightarrow$ the RKHS of $f_{|\mathcal{M}}$ is norm equivalent to the one of the intrinsic process !

Consequence

1. $\Longrightarrow$ the RKHS of $f_{|\mathcal{M}}$ is norm equivalent to the one of the intrinsic process !
2. $\Longrightarrow$ they have the same contraction properties !
$\rightsquigarrow \varepsilon_n \sim n^{-\frac{\beta \wedge s}{2s+d}}$

Consequence

1. $\Longrightarrow$ the RKHS of $f_{|\mathcal{M}}$ is norm equivalent to the one of the intrinsic process !
2. $\Longrightarrow$ they have the same contraction properties ! $\rightsquigarrow \varepsilon_n \sim n^{-\frac{\beta \wedge s}{2s+d}}$

in particular : the restriction of a Euclidean Matern process to a submanifold $\mathcal{M} \subset \mathbb{R}^D$ of dimension $d < D$ has a contraction rate depending exponentially in d (not D !)

Extension to Besov priors

- sometimes : Holder/Sobolev assumptions are too restrictive

Extension to Besov priors

- sometimes : Holder/Sobolev assumptions are too restrictive
- spatially inhomogeneous functions are known to be more adequatly represented by some Besov spaces $B^s_{pp}(\mathcal{M}), p \in [1,2) \rightsquigarrow$ inverse problems, imaging..

Extension to Besov priors

- sometimes : Holder/Sobolev assumptions are too restrictive
- spatially inhomogeneous functions are known to be more adequately represented by some Besov spaces $B^s_{pp}(\mathcal{M}), p \in [1,2) \rightsquigarrow$ inverse problems, imaging..
- actually Grosse & Schneider [3] give $Tr : B^{s+D/p}_{pp}(\mathbb{R}^D) \to B^{s+d/p}_{pp}(\mathcal{M}), Ex : B^{s+d/p}_{pp}(\mathcal{M}) \to B^{s+D/p}_{pp}(\mathbb{R}^D)$

- we can mimick the approach of GPs and RKHS using the "p-exponential priors" of Agapiou & al [1] :

$$f = \sum_{j\geq 1} a_j Z_j u_j, Z_j \overset{i.i.d}{\sim} f_p$$

where $f_p(x) \propto e^{-|x|^p/p}, a_j \in \mathbb{R}, (u_j)_{j\geq 1}$ = Schauder basis of $\mathcal{C}(\mathcal{X})$ (here $\mathcal{X} = [0,1]^D$)

- we can mimick the approach of GPs and RKHS using the "p-exponential priors" of Agapiou & al [1] :

$$f = \sum_{j \geq 1} a_j Z_j u_j, Z_j \overset{i.i.d}{\sim} f_p$$

where $f_p(x) \propto e^{-|x|^p/p}, a_j \in \mathbb{R}, (u_j)_{j \geq 1}$ = Schauder basis of $\mathcal{C}(\mathcal{X})$ (here $\mathcal{X} = [0,1]^D$)

- problem : the restriction of a p-exponential prior is not necessarily p-exponential $\rightsquigarrow$ to conclude we consider $(u_j) \iff (\psi_{jk})_{j \geq 1, k \leq 2^{jD}}$ = regular wavelet basis

Theorem

In the fixed design regression model, if
$f_0 \in B^s_{pp}(\mathcal{M}), s > d/p, p \in [1,2]$ and

$$f = \left(n\epsilon_n^2\right)^{-1/p} \sum_{j \geq 1} 2^{-j(s-d/p+D/2)} \sum_{k=1}^{2^{jD}} \xi_{jk}\psi_{jk}, \xi_{jk} \overset{i.i.d}{\sim} f_p$$

then

$$\Pi\left[\|f - f_0\|_n > M\epsilon_n | \mathbb{X}^n\right] \xrightarrow[n\to\infty]{P_0^\infty} 0$$

for $M > 0$ large enough and $\epsilon_n \propto n^{-\frac{s}{2s+d}}$.

idea : even if $f_{|\mathcal{M}}$ is not a p-exponential process, we can always consider

$$p_{\#} f = \left(n\epsilon_n^2\right)^{-1/p} \sum_{j\geq 1} 2^{-j(s-d/p+D/2)} \sum_{k\in I_j} \xi_{jk}\psi_{jk}$$

where $I_j = \#\left\{1 \leq k \leq 2^{jD} : supp(\psi_{jk}) \cap \mathcal{M} \neq \emptyset\right\}$, which is always p-exponentially distributed ; and using $\#I_j \asymp 2^{jd} << 2^{jD}$ + trace/extension theorem allows us to replace D by d in the rate

Conclusion & future work

Conclusion & future work

Take home message :

Conclusion & future work

Take home message :

- two ways to construct priors for functions on manifolds : **intrinsically** or **extrinsically**

Conclusion & future work

Take home message :

- two ways to construct priors for functions on manifolds : **intrinsically** or **extrinsically**
- as we saw : in somes cases the two methods may have **similar rates of contraction** : differences stem from the constants

Conclusion & future work

Take home message :

- two ways to construct priors for functions on manifolds : **intrinsically** or **extrinsically**
- as we saw : in somes cases the two methods may have **similar rates of contraction** : differences stem from the constants

However

Conclusion & future work

Take home message :

- two ways to construct priors for functions on manifolds : **intrinsically** or **extrinsically**
- as we saw : in somes cases the two methods may have **similar rates of contraction** : differences stem from the constants

However

- we do see **differences of performance in practice**

Conclusion & future work

Take home message :

- two ways to construct priors for functions on manifolds : **intrinsically** or **extrinsically**
- as we saw : in somes cases the two methods may have **similar rates of contraction** : differences stem from the constants

However

- we do see **differences of performance in practice**
- geometrical/intrinsic methods tend to perform better in the **low data regime**

Conclusion & future work

Take home message :

- two ways to construct priors for functions on manifolds : **intrinsically** or **extrinsically**
- as we saw : in somes cases the two methods may have **similar rates of contraction** : differences stem from the constants

However

- we do see **differences of performance in practice**
- geometrical/intrinsic methods tend to perform better in the **low data regime**

Possible extensions

Conclusion & future work

Take home message :

- two ways to construct priors for functions on manifolds : **intrinsically** or **extrinsically**
- as we saw : in somes cases the two methods may have **similar rates of contraction** : differences stem from the constants

However

- we do see **differences of performance in practice**
- geometrical/intrinsic methods tend to perform better in the **low data regime**

Possible extensions

- similar result for the **square exponential kernel** ?

Conclusion & future work

Take home message :

- two ways to construct priors for functions on manifolds : **intrinsically** or **extrinsically**
- as we saw : in somes cases the two methods may have **similar rates of contraction** : differences stem from the constants

However

- we do see **differences of performance in practice**
- geometrical/intrinsic methods tend to perform better in the **low data regime**

Possible extensions

- similar result for the **square exponential kernel** ?
- **adaptivity** : ok for intrinsic, what about extrinsic ?

Thank you !

UNIVERSITY OF OXFORD

European Research Council
Established by the European Commission

Jesus College
OXFORD

Sergios Agapiou, Masoumeh Dashti, and Tapio Helin.
Rates of contraction of posterior distributions based on p-exponential priors, 2020.

Iskander Azangulov, Andrei Smolensky, Alexander Terenin, and Viacheslav Borovitskiy.
Stationary kernels and gaussian processes on lie groups and their homogeneous spaces i: the compact case, 2022.

Grosse Nadine and Cornelia Schneider.
Sobolev spaces on riemannian manifolds with bounded geometry: General coordinates and traces, 2013.

Aad van der Vaart and Harry van Zanten.
Information rates of nonparametric gaussian process methods.
Journal of Machine Learning Research, 12(60):2095–2119, 2011.