DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF COPENHAGEN

CENTRE FOR STOCHASTIC GEOMETRY AND ADVANCED BIOIMAGING

Wrapped Gaussian Process Regression on Riemannian Manifolds

Anton Mallasto Aasa Feragen
mallasto@di.ku.dk *aasa@di.ku.dk*

Summary

- Gaussian process (GP) regression is a powerful tool in non-parametric regression providing uncertainty estimates. However, it is limited to data in vector spaces, thus not suitable for non-linear geometries (see Fig. 1).
- We tackle this problem by defining wrapped Gaussian processes (WGPs) on Riemannian manifolds, using the probabilistic setting to generalize GP regression to the context of manifold-valued targets.
- We experiment on diffusion weighted imaging (DWI) data, directional data on the sphere and in the Kendall shape space, endorsing WGP regression as an efficient and flexible tool for manifold-valued regression.

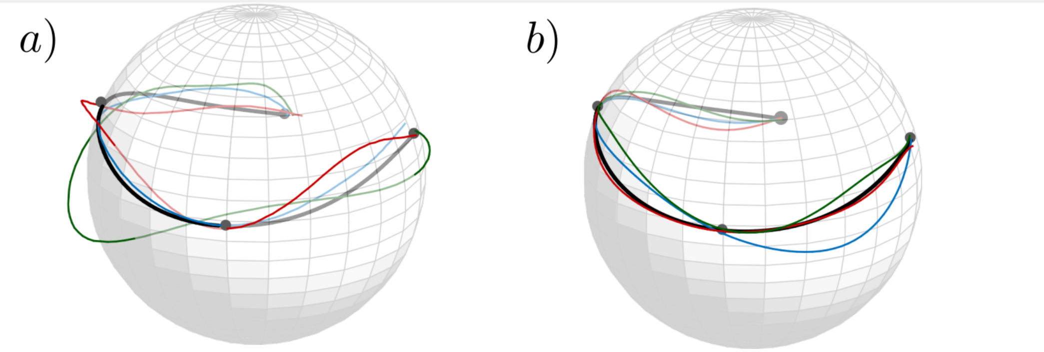

Figure 1: **a)** Euclidean GP regression on a 2-sphere and **b)** WGP regression on the same data set.

Wrapped Gaussian Distributions

- Let M be an n-dimensional Riemannian manifold and choose $\mu \in M$. Then, if for some multivariate Gaussian $Y \sim \mathcal{N}(0, K)$ living in the tangent space $T_\mu M$

$$X = \mathrm{Exp}_\mu(Y), \tag{1}$$

then X has a *wrapped Gaussian distribution*, denoted by $X \sim \mathcal{N}_M(\mu, K)$. Furthermore, we define the maps $\mu_{\mathcal{N}_M}(X) := \mu$ and $\mathrm{Cov}_{\mathcal{N}_M}(X) = K$. See Fig. 2 a).

- The random points $X_i \sim \mathcal{N}_{M_i}(\mu_i, K_i)$, $i = 1, 2$, are *jointly WGD*, if the random point (X_1, X_2) on $M_1 \times M_2$ is WGD, that is,

$$(X_1, X_2) \sim \mathcal{N}_{M_1 \times M_2}\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}K_1 & K_{12}\\ K_{21} & K_2\end{pmatrix}\right), \tag{2}$$

for some matrix $K_{12} = K_{21}^T$.

Theorem 1 *Assume X_1, X_2 are jointly WGD as in (2), then we have the conditional distribution*

$$X_1 | (X_2 = p_2) \sim \left(\mathrm{Exp}_{\mu_1}\right)_\# \left(\sum_{v \in A} \lambda_v \mathcal{N}(\mu_v, K_v)\right), \tag{3}$$

where

$$\begin{aligned}
\mu_v &= K_{12} K_2^{-1} v,\\
K_v &= K_1 - K_{12} K_2^{-1} K_{12}^T,\\
\lambda_v &= \frac{\mathcal{N}(v|\mathbf{0}, K_2)}{\mathbb{P}\{A\}},\\
A &= \{v \in T_{\mu_2} M \mid \mathrm{Exp}_{\mu_2}(v) = p_2\},\\
\mathbb{P}\{A\} &= \sum_{v \in A} \mathcal{N}(v|\mathbf{0}, K_2),
\end{aligned} \tag{4}$$

and $f_\# \mu(A) = \mu(f^{-1}(A))$ for measure μ and measurable f, A.

Wrapped Gaussian Processes

- A collection f of random points on a manifold M indexed over a set Ω is a *wrapped Gaussian process* (WGP), if every finite subcollection $(f(\omega_i))_{i=1}^N$, $\omega_i \in \Omega$, is jointly WGD on M^N. We define

$$m(\omega) := \mu_{\mathcal{N}_M}(f(\omega)) \tag{5}$$

$$k(\omega, \omega') := \mathrm{Cov}_{\mathcal{N}_M}(f(\omega), f(\omega')), \tag{6}$$

called the *basepoint function* and *tangent space covariance function* of f. See Fig. 2 b) for a visualization.

WGP Regression

- The learning goal is as follows: given data $\boldsymbol{D}_M = \{(x_i, p_i)\}_{i=1}^n$ and assuming $p_i = f(x_i)$ for some map $f : \mathbb{R}^d \to M$, infer f.
- Approach: define a prior distribution $f \sim \mathcal{GP}_M(m, k)$ and condition on the given data using Theorem 1. The process is given in Algorithm 1 and illustrated in Fig. 3.

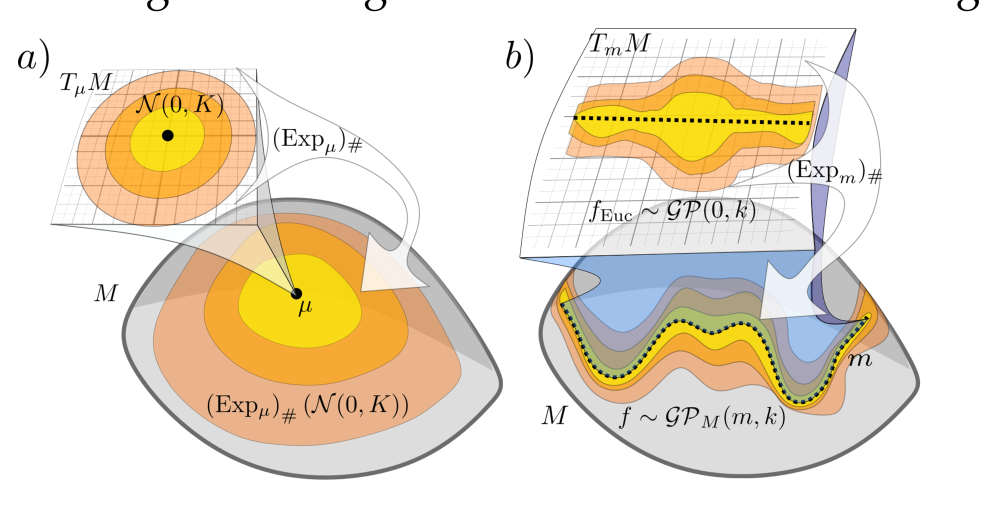

Figure 2: **a)** Illustration of a wrapped Gaussian distribution $\mathcal{N}_M(\mu, K)$ and **b)** of a wrapped Gaussian process $\mathcal{GP}_M(m, k)$.

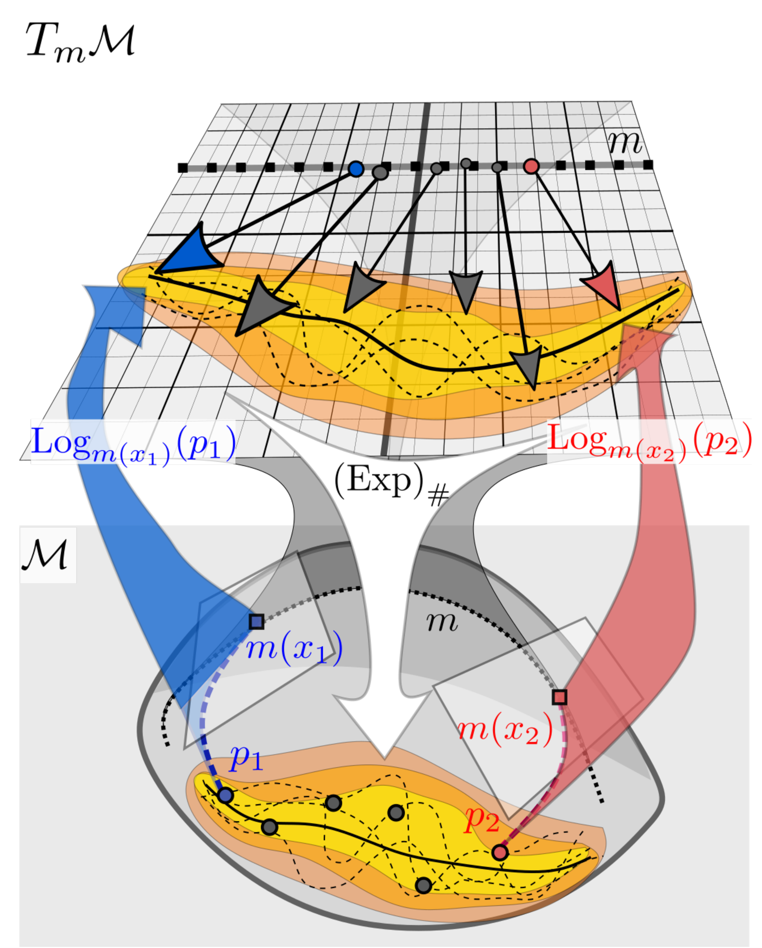

Figure 3: Visualization of WGP regression. First, a prior basepoint function m is chosen (dotted black). Given a data point (x_i, p_i), we compute $\mathrm{Log}_{m(x_i)}(p_i)$ (in blue and red). Then, ordinary GP regression is applied in $T_m M$. The resulting GP is then pushed-forward onto M by the Riemannian exponential Exp.

Algorithm 1: WGP Regression.
Data: $\boldsymbol{D}_M = \{(x_i, p_i)\}_{i=1}^n$.
Result: Predictive distribution for $\boldsymbol{p}_* | \boldsymbol{p}$ at $\boldsymbol{x}_*$.

i. Choose a prior basepoint function m.
ii. Transform $\boldsymbol{D}_{T_m M} \leftarrow \{(x_i, \mathrm{Log}_{m(x_i)}(p_i))\}_{i=1}^N$.
iii. Choose a prior tangent space covariance function k from a parametric family by optimizing the hyperparameters.
iv. Using GP prior $\mathcal{GP}(0, k)$, carry out Euclidean GP regression for the transformed data $\boldsymbol{D}_{T_m M}$, yielding the mean and covariance $(\boldsymbol{\mu}_*, \Sigma_*)$.
vi. End with the predictive distribution $\boldsymbol{p}_* | \boldsymbol{p} \sim (\mathrm{Exp}_{\boldsymbol{m}_*})_\# (\mathcal{N}(\boldsymbol{\mu}_*, \Sigma_*))$

Experiments

- We demonstrate WGP regression on three different manifolds; the 2-sphere, the space of 3-by-3 covariance matrices, and the *Kendall shape space*.
- In the sphere experiment, we use the RBF and periodic covariance functions (kernels). For the rest, we use the RBF kernel with hyperparameters chosen to maximize the likelihood of observed data.

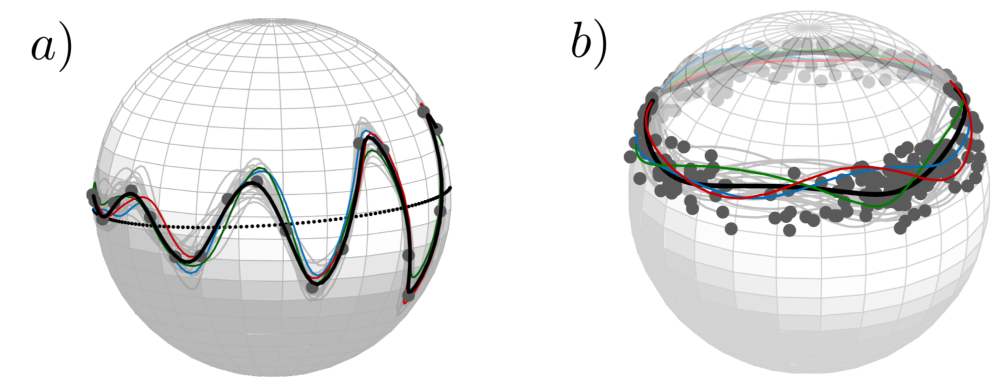

Figure 4: **a)** WGP with prior BPF given by geodesic regression (dotted black) on a toy data set (grey dots) on S^2. **b)** WGP regression on a motion capture dataset of the orientation of the left *femur* of a walking person, prior BPF is given by *principal curve analysis*.

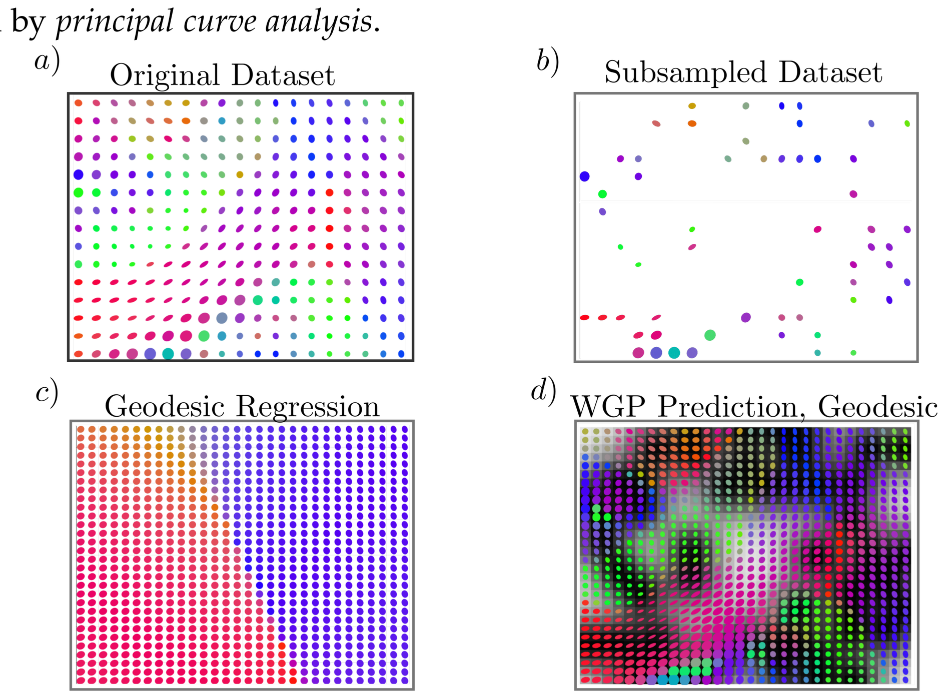

Figure 5: Upsampling DTI tensor field by WGP regression. Original data set is given in **a)**, which we subsample, yielding **b)**, with only 20% of the elements of the original tensor field are present. The regression is carried out using the prior BPF given by geodesic regression, shown in **c)**. The resulting predictive distribution is given in **d)**. The color of the tensors is given by their principal eigenvector, and in **d)**, white background means uncertain predictions.

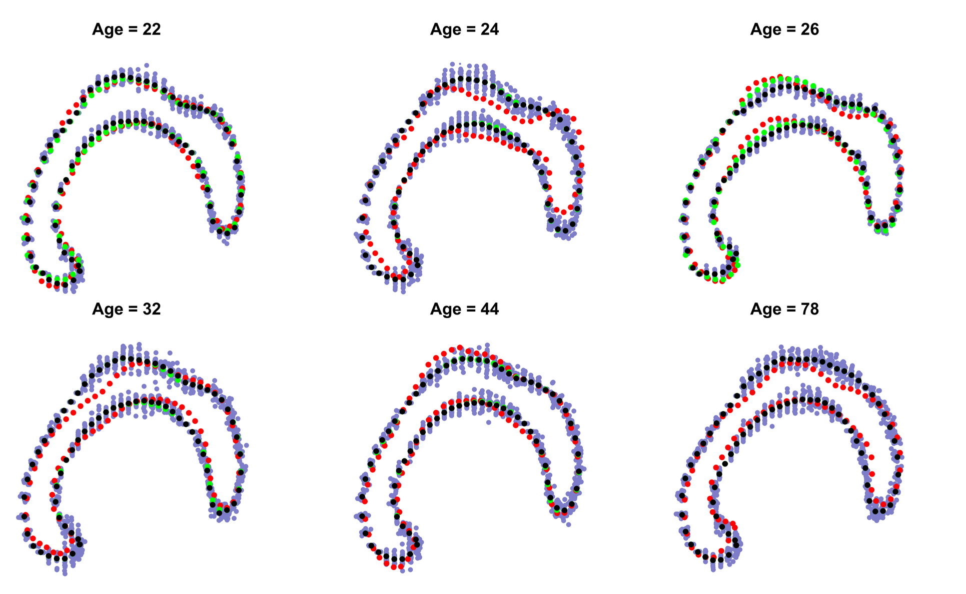

Figure 6: WGP regression on a population of *Corpus Callosum* shapes labeled by age. Red depicts data points from the test set. In black, the MAP estimates of the predictive distributions, in green values of the prior BPF at corresponding ages. Drawn in blue are 20 samples from the predictive distribution.

Acknowledgements

This research was supported by Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation. Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.