Gaussian Process Landmarking on Manifolds

Tingran Gao

Committee on Computational and Applied Mathematics
Department of Statistics
The University of Chicago

Applied Math Youth Forum
Beijing International Center for Mathematical Research
Beijing, China

December 22, 2019

Outline

Motivation

- Geometric Morphometrics
- Experimental Design

Gaussian Process Landmarking

- Sequential Experimental Design
- Witten Laplacian
- Reduced Basis Method

Other Applications

Joint work with **Shahar Z. Kovalsky**, **Doug M. Boyer**, **Ingrid Daubechies**

Outline

Motivation

- **Geometric Morphometrics**
- Experimental Design

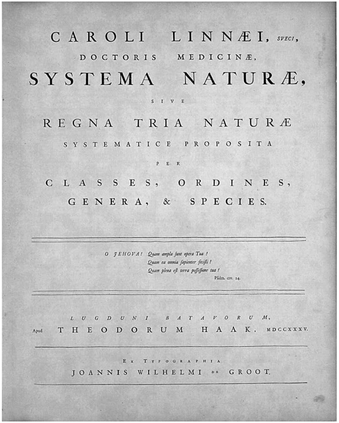
Gaussian Process Landmarking

- Sequential Experimental Design
- Witten Laplacian
- Reduced Basis Method

Other Applications

Joint work with **Shahar Z. Kovalsky**, **Doug M. Boyer**, **Ingrid Daubechies**

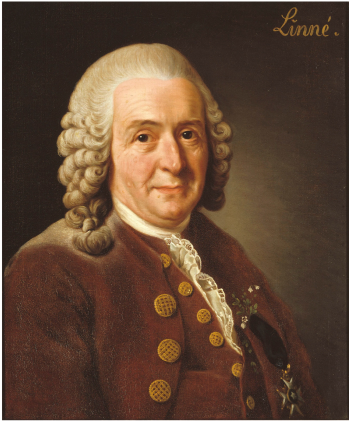
Morphology and Classification


CAROLI LINNÆI, SVECI,
DOCTORIS MEDICINÆ,
SYSTEMA NATURÆ,
SIVE
REGNA TRIA NATURÆ
SYSTEMATICE PROPOSITA
PER
CLASSES, ORDINES,
GENERA, & SPECIES.

O JEHOVA! Quam ampla sunt opera Tua!
Quam ea omnia sapienter fecisti!
Quam plena est terra possessione tua!
Psalm. civ. 24.

LUGDUNI BATAVORUM,
Apud THEODORUM HAAK, MDCCXXXV.
Ex Typographia
JOANNIS WILHELMI DE GROOT.

Systema Naturae, 1735



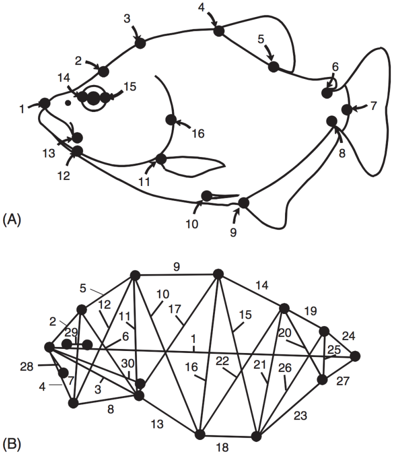
Carl Linnaeus (1707-1778)

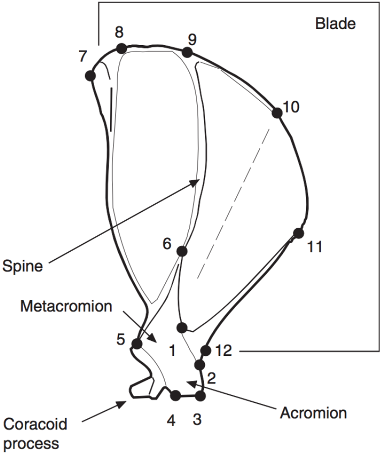
Iris — A Classical Example of Morphometrics

Iris setosa				*Iris versicolor*				*Iris virginica*			
Sepal length	Sepal width	Petal length	Petal width	Sepal length	Sepal width	Petal length	Petal width	Sepal length	Sepal width	Petal length	Petal width
5·1	3·5	1·4	0·2	7·0	3·2	4·7	1·4	6·3	3·3	6·0	2·5
4·9	3·0	1·4	0·2	6·4	3·2	4·5	1·5	5·8	2·7	5·1	1·9
4·7	3·2	1·3	0·2	6·9	3·1	4·9	1·5	7·1	3·0	5·9	2·1
4·6	3·1	1·5	0·2	5·5	2·3	4·0	1·3	6·3	2·9	5·6	1·8
5·0	3·6	1·4	0·2	6·5	2·8	4·6	1·5	6·5	3·0	5·8	2·2
5·4	3·9	1·7	0·4	5·7	2·8	4·5	1·3	7·6	3·0	6·6	2·1
4·6	3·4	1·4	0·3	6·3	3·3	4·7	1·6	4·9	2·5	4·5	1·7
5·0	3·4	1·5	0·2	4·9	2·4	3·3	1·0	7·3	2·9	6·3	1·8

R.A. Fisher "The Use of Multiple Measurements in Taxonomic Problems." *Annals of Eugenics* 7.2 (1936): 179-188.

Landmark-based Geometric Morphometrics

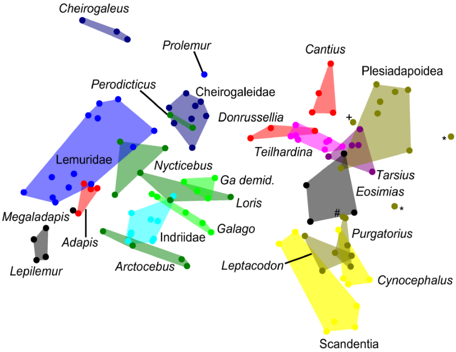




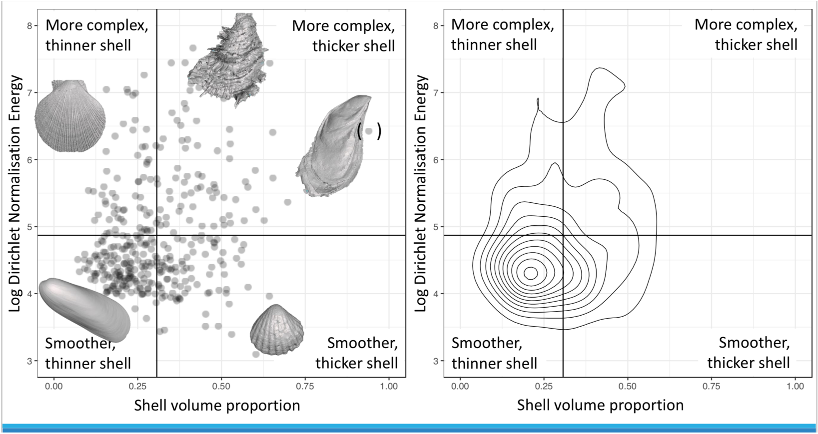
Geometric Morphometrics

$$\left(S_1, \{x_j\}_{j=1}^J\right), \left(S_2, \{y_j\}_{j=1}^J\right) \longrightarrow$$

$$d^2_{Procrustes}\left(S_1, S_2\right) = \min_{R \text{ rigid motion}} \frac{1}{J} \sum_{j=1}^{J} \left\| R\left(x_j\right) - y_j \right\|^2$$



Boyer et al. "Algorithms to Automatically Quantify the Geometric Similarity of Anatomical Surfaces." *Proceedings of the National Academy of Sciences* 108.45 (2011): 18221-18226.


More complex, thinner shell
More complex, thicker shell
Smoother, thinner shell
Smoother, thicker shell
Log Dirichlet Normalisation Energy
Shell volume proportion
0.00
0.25
0.50
0.75
1.00
3
4
5
6
7
8
More complex, thinner shell
More complex, thicker shell
Smoother, thinner shell
Smoother, thicker shell
Log Dirichlet Normalisation Energy
Shell volume proportion

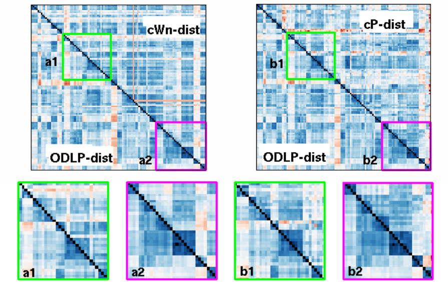
Shape Distances

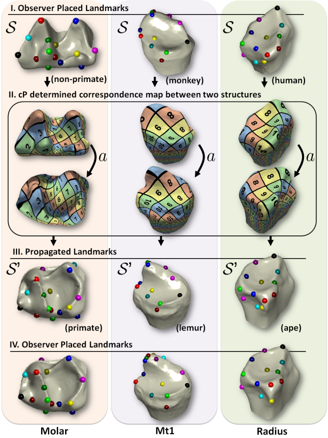
- John Clifford Gower, John C. Gower, and Garmt B. Dijksterhuis. "**Procrustes Problems.**" vol. 3 of Oxford Statistical Science Series, Oxford University Press Oxford, 2004.
- Rima Alaifari, Ingrid Daubechies, and Yaron Lipman. "**Continuous Procrustes distance between two surfaces.**" *Communications on Pure and Applied Mathematics 66*, no. 6 (2013): 934-964.
- Doug M. Boyer, Yaron Lipman, Elizabeth St Clair, Jesus Puente, Biren A. Patel, Thomas Funkhouser, Jukka Jernvall, and Ingrid Daubechies. "**Algorithms to automatically quantify the geometric similarity of anatomical surfaces.**" *Proceedings of the National Academy of Sciences* 108, no. 45 (2011): 18221-18226.
- Facundo Mémoli. "**Gromov–Wasserstein distances and the metric approach to object matching.**" *Foundations of Computational Mathematics* 11, no. 4 (2011): 417-487.
- Rongjie Lai, Hongkai Zhao. "**Multi-scale Non-Rigid Point Cloud Registration Using Robust Sliced-Wasserstein Distance via Laplace-Beltrami Eigenmap**", *SIAM Journal on Imaging Sciences* 10(2), pp. 449-483, 2017.
- Patrice Koehl, Joel Hass, "**Landmark-Free Geometric Methods in Biological Shape Analysis**", *Journal of The Royal Society Interface*, 12 no. 113 (2015): 20150795
- ......

Shape Distances

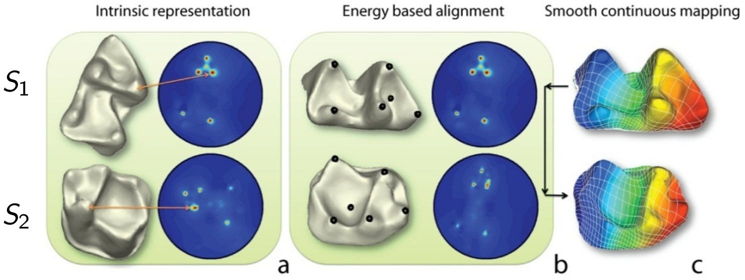
Boyer et al. 2011, 2012
Daubechies et al. 2011, 2013 — Conformal Wasserstein Distance
Al-Aifari et al. 2013 — Continuous Procrustes Distance

$$d_{cP}(S_1, S_2) = \inf_{\mathcal{C}\in\mathcal{A}(S_1,S_2)} \inf_{R\in\mathbb{E}(3)} \left(\int_{S_1} \| R(x) - \mathcal{C}(x) \|^2 \, d\mathrm{vol}_{S_1}(x) \right)^{\frac{1}{2}}$$





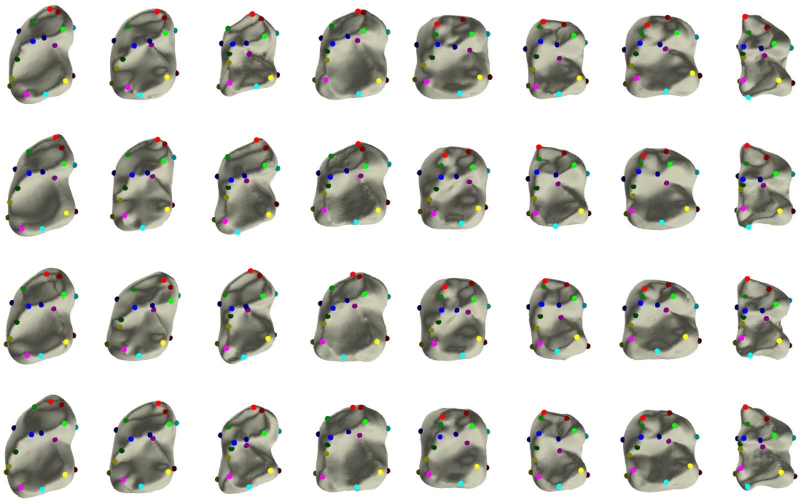
Landmark-Free Approaches: Bypass Feature Extraction



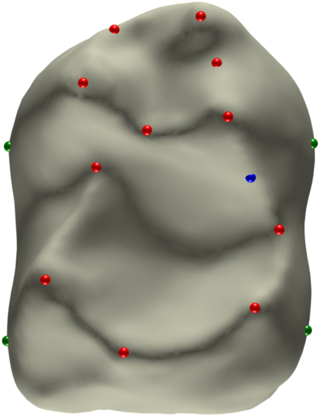
Correspondence-Based Shape Distances

$$D\left(S_1, S_2\right) = \inf_{f \in \mathscr{A}(S_1, S_2)} F\left(f; S_1, S_2\right)$$

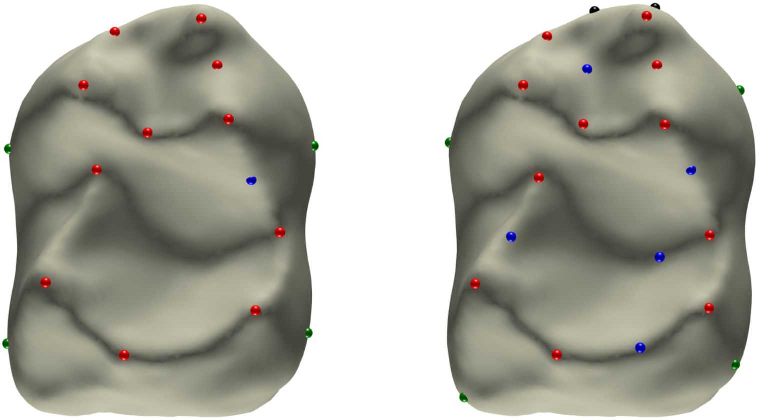
Revisiting Landmarks: For the Sake of Interpretability, or Turning the Clock Back?



Bookstein's Typology



Bookstein's Typology ...... *Cracked?*



Outline

Motivation

- Geometric Morphometrics
- **Experimental Design**

Gaussian Process Landmarking

- Sequential Experimental Design
- Witten Laplacian
- Reduced Basis Method

Other Applications

Joint work with **Shahar Z. Kovalsky**, **Doug M. Boyer**, **Ingrid Daubechies**

Experimental Design: A Classical Paradigm in Statistics

Experimental Design: A Classical Paradigm in Statistics

- **Estimation**: Given i.i.d. data $\{(X_i, Y_i), 1 \le i \le n\}$ sampled from a joint distribution $\mathcal{D}$, find good estimators by solving

$$\min_{\hat{f} \in \mathscr{F}} \mathbb{E}_{(X,Y) \sim \mathcal{D}} \left[\left(\hat{f}_n \left(X \mid \{(X_i, Y_i)\} \right) - Y \right)^2 \right]$$

Experimental Design: A Classical Paradigm in Statistics

- **Estimation**: Given i.i.d. data $\{(X_i, Y_i), 1 \le i \le n\}$ sampled from a joint distribution $\mathcal{D}$, find good estimators by solving

$$\min_{\hat{f} \in \mathscr{F}} \mathbb{E}_{(X,Y)\sim\mathcal{D}} \left[\left(\hat{f}_n \left(X \mid \{(X_i, Y_i)\} \right) - Y \right)^2 \right]$$

- **Experimental Design**: Given an estimation procedure $f \mapsto \hat{f}_n$ for a class of target functions $f \sim \mathscr{F}$, find samples $x_1, \cdots, x_n$ that minimize

$$\min_{x_1,\cdots,x_n} \mathbb{E}_{f\sim\mathscr{F}} \left[\left(\hat{f}_n \left(x \mid \{(x_i, f(x_i))\} \right) - f(x) \right)^2 \right]$$

Experimental Design: A Classical Paradigm in Statistics

- **Estimation**: Given i.i.d. data $\{(X_i, Y_i), 1 \leq i \leq n\}$ sampled from a joint distribution $\mathcal{D}$, find good estimators by solving

$$\min_{\hat{f} \in \mathscr{F}} \mathbb{E}_{(X,Y)\sim\mathcal{D}} \left[\left(\hat{f}_n \left(X \mid \{(X_i, Y_i)\} \right) - Y \right)^2 \right]$$

- **Experimental Design**: Given an estimation procedure $f \mapsto \hat{f}_n$ for a class of target functions $f \sim \mathscr{F}$, find samples $x_1, \cdots, x_n$ that minimize

$$\min_{x_1, \cdots, x_n} \mathbb{E}_{f\sim\mathscr{F}} \left[\left(\hat{f}_n \left(x \mid \{(x_i, f(x_i))\} \right) - f(x) \right)^2 \right]$$

- **Gaussian Process Experimental Design**: $\mathscr{F} = \mathrm{GP}(m, K)$ is a Gaussian process on domain $\mathcal{X}$

Gaussian Process Experimental Design: (Simple) Kriging

- Gaussian process $\mathrm{GP}(0, K)$ defined on a domain Ω by
  - kernel function $K : \Omega \times \Omega \to \mathbb{R}$
  - For $f \sim \mathrm{GP}(0, K)$, its values at any n points $x_1, \cdots, x_n \in \Omega$

$$(f(x_1), \cdots, f(x_n))^\top \in \mathbb{R}^n$$

follow a multivariate normal distribution

$$\mathcal{N}(0, \mathscr{K}_n)$$

where

$$\mathscr{K}_n = \begin{pmatrix} K(x_1, x_1) & \dots & K(x_1, x_n) \\ \vdots & & \vdots \\ K(x_n, x_1) & \dots & K(x_n, x_n) \end{pmatrix} \in \mathbb{R}^{n \times n}$$

Gaussian Process Experimental Design: (Simple) Kriging

- Conditioned on values at $x_1, \cdots, x_n$, i.e.

$$f(x_i) = y_i, \quad 1 \le i \le n,$$

the value $f(x)$ at any new point $x \in \Omega$ follows a normal distribution

$$\begin{aligned} f(x) \mid \{f(x_i) = y_i,\, 1 \le i \le n\} \\ \sim \mathscr{N}\left(\mathbf{r}_n(x)^\top \mathscr{K}_n^{-1}\mathbf{y}, K(x,x) - \mathbf{r}_n(x)^\top \mathscr{K}_n^{-1}\mathbf{r}_n(x)\right) \end{aligned}$$

where

$$\begin{aligned} \mathbf{r}_n(x) &= (K(x,x_1), \cdots, K(x,x_n))^\top \in \mathbb{R}^n \\ \mathbf{y} &= (y_1, \cdots, y_n)^\top \in \mathbb{R}^n. \end{aligned}$$

Gaussian Process Experimental Design: (Simple) Kriging

- Mean Squared Error (**MSE**) of the predictor

$$\hat{f}_n(x) = \mathbb{E}\left(f(x) \mid \{f(x_i) = y_i,\ 1 \le i \le n\}\right)$$
$$= \mathbf{r}_n(x)^\top \mathscr{K}_n^{-1}\mathbf{y}$$

is simply

$$\mathrm{MSE}\left(\hat{f}_n(x)\right) = \mathbb{E}\left(\hat{f}_n(x) - f(x)\right)^2$$
$$= K(x, x) - \mathbf{r}_n(x)^\top \mathscr{K}_n^{-1}\mathbf{r}_n(x)$$

Gaussian Process Experimental Design: (Simple) Kriging

- Mean Squared Error (**MSE**) of the predictor

$$\begin{aligned}\hat{f}_n(x) &= \mathbb{E}\left(f(x) \mid \{f(x_i) = y_i,\, 1 \le i \le n\}\right) \\ &= \mathbf{r}_n(x)^\top \mathscr{K}_n^{-1}\mathbf{y}\end{aligned}$$

  is simply

$$\begin{aligned}\mathrm{MSE}\left(\hat{f}_n(x)\right) &= \mathbb{E}\left(\hat{f}_n(x) - f(x)\right)^2 \\ &= K(x,x) - \mathbf{r}_n(x)^\top \mathscr{K}_n^{-1}\mathbf{r}_n(x)\end{aligned}$$

- **Kriging: How to pick $x_1, \cdots, x_n \in M$ so as to minimize the Integrated MSE (IMSE)**

$$\mathrm{IMSE}\left(\hat{f}_n\right) := \int_\Omega \mathrm{MSE}\left(\hat{f}_n(x)\right) \mathrm{d}x$$

Universal Convergence of (Simple) Kriging

Spectral Convergence (Wang-Tuo-Wu 2017). If K is a Gaussian kernel, $\{x_1, \cdots, x_n\} \subset \Omega$ a bounded open subset of a Euclidean space, then w.h.p.

$$\sup_{x \in \Omega} \left| \hat{f}_n(x) - f(x) \right| = O_P\left(h_n^{\frac{c}{h_n} - \frac{1}{2}} \log^{\frac{1}{2}} (1/h_n) \right)$$

where h_n is the *fill distance*

$$h_n := \sup_{x \in \Omega} \min_{1 \le i \le n} \|x - x_i\| .$$

Corollary. Let h_* denote the minimum fill distance on Ω for n points. Then

$$\inf_{\{x_1, \cdots, x_n\} \subset \Omega} \mathrm{IMSE}\left(\hat{f}_n\right) = O_P\left(h_*^{\frac{c}{h_*} - 1} \log (1/h_*) \right).$$

Wenjia Wang, Rui Tuo, and C. F. Wu. "Universal Convergence of Kriging." arXiv:1710.06959 (2017).

Outline

Motivation

- Geometric Morphometrics
- Experimental Design

Gaussian Process Landmarking

- Sequential Experimental Design
- Witten Laplacian
- Reduced Basis Method

Other Applications

Joint work with **Shahar Z. Kovalsky**, **Doug M. Boyer**, **Ingrid Daubechies**

Outline

Motivation

- Geometric Morphometrics
- Experimental Design

Gaussian Process Landmarking

- **Sequential Experimental Design**
- Witten Laplacian
- Reduced Basis Method

Other Applications

Joint work with **Shahar Z. Kovalsky**, **Doug M. Boyer**, **Ingrid Daubechies**

Sequential Experimental Design

- Select $x_1, \cdots, x_n$ sequentially
- **Methodology:** Iteratively pick the *most uncertain* location

Sequential Experimental Design

- Select $x_1, \cdots, x_n$ sequentially
- **Methodology:** Iteratively pick the *most uncertain* location
- With $x_1, \cdots, x_{i-1}$ given, pick

$$x_i := \arg\max_{x \in \Omega} \mathrm{MSE}\left(\hat{f}_{i-1}(x)\right)$$

Sequential Experimental Design

- Select $x_1, \cdots, x_n$ sequentially
- **Methodology:** Iteratively pick the *most uncertain* location
- With $x_1, \cdots, x_{i-1}$ given, pick

$$x_i := \arg\max_{x \in \Omega} \mathrm{MSE}\left(\hat{f}_{i-1}(x)\right)$$

- For Gaussian process experimental design, this amounts to iteratively picking the point maximizing

$$\mathrm{MSE}\left(\hat{f}_{i-1}(x)\right) = K(x, x) - \mathbf{r}_{i-1}(x)^\top \mathscr{K}_{i-1}^{-1} \mathbf{r}_{i-1}(x)$$

Sequential Experimental Design

- Select $x_1, \cdots, x_n$ sequentially
- **Methodology:** Iteratively pick the *most uncertain* location
- With $x_1, \cdots, x_{i-1}$ given, pick

$$x_i := \arg\max_{x \in \Omega} \mathrm{MSE}\left(\hat{f}_{i-1}(x)\right)$$

- For Gaussian process experimental design, this amounts to iteratively picking the point maximizing

$$\mathrm{MSE}\left(\hat{f}_{i-1}(x)\right) = K(x, x) - \mathbf{r}_{i-1}(x)^\top \mathscr{K}_{i-1}^{-1} \mathbf{r}_{i-1}(x)$$

- **Numerical Linear Algebra Perspective:** Cholesky Decomposition with Complete Pivoting

Gaussian Process Landmarking

Input: d-dimensional Manifold M isometrically embedded in $\mathbb{R}^D$, where $d < D$, and number of landmarks N

- Construct a kernel $K^c : M \times M \to \mathbb{R}$

$$K_\epsilon^c(x, y) = \int_M e^{-\frac{1}{2\epsilon}\|x-z\|_D^2} c(z)\, e^{-\frac{1}{2\epsilon}\|z-y\|_D^2} \mathrm{dvol}_M(z)$$

  where $c : M \to \mathbb{R}$ is the (Gauss/mean/L^2-) curvature of M
- For $i = 1$

$$x_1 = \arg\max_{x \in M} K_\epsilon^c(x, x)$$

Gaussian Process Landmarking

- For $i = 2, \ldots, N$

$$\begin{aligned} x_i &= \operatorname*{arg\,max}_{x \in M} \mathrm{MSE}\left(\hat{f}_{i-1}(x)\right) \\ &= \operatorname*{arg\,max}_{x \in M} \left[K_\epsilon^c(x,x) - \mathbf{r}_{i-1}^c(x)^\top \mathscr{K}_{i-1}^{-1} \mathbf{r}_{i-1}^c(x)\right] \end{aligned}$$

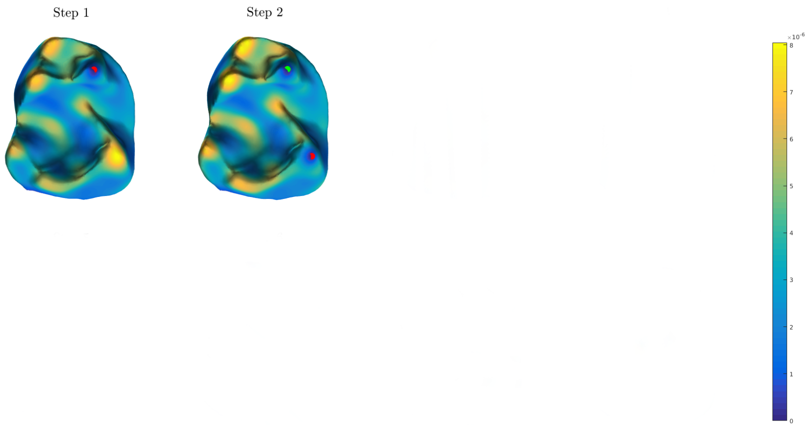
where

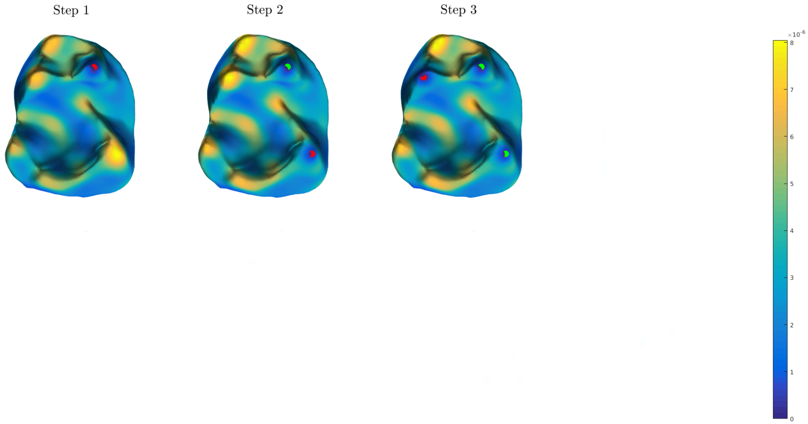
$$\mathbf{r}_{i-1}^c(x) = \left(K_\epsilon^c(x, x_1), \cdots, K_\epsilon^c(x, x_{i-1})\right)^\top \in \mathbb{R}^{i-1},$$

$$\mathscr{K}_{i-1} = \begin{pmatrix} K(x_1, x_1) & \ldots & K(x_1, x_{i-1}) \\ \vdots & & \vdots \\ K(x_{i-1}, x_1) & \ldots & K(x_{i-1}, x_{i-1}) \end{pmatrix} \in \mathbb{R}^{(i-1)\times(i-1)}$$

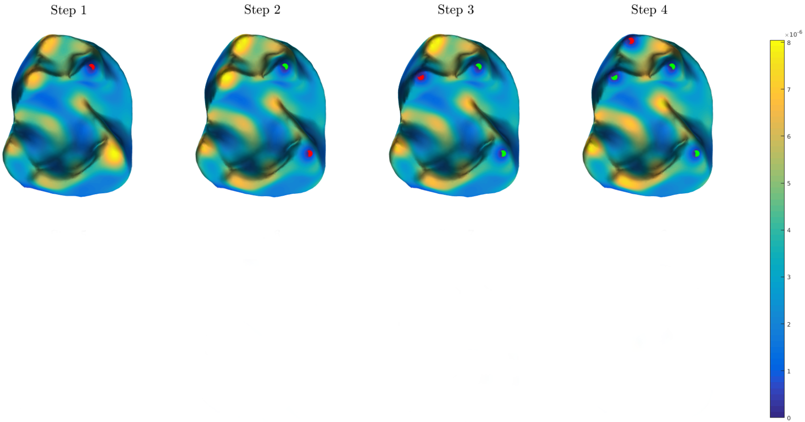
Output: N landmarks $x_1, \cdots, x_N \in M$

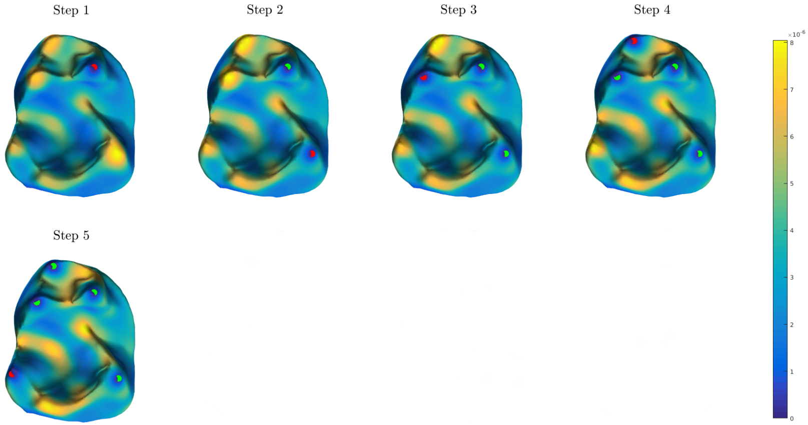
Step 1

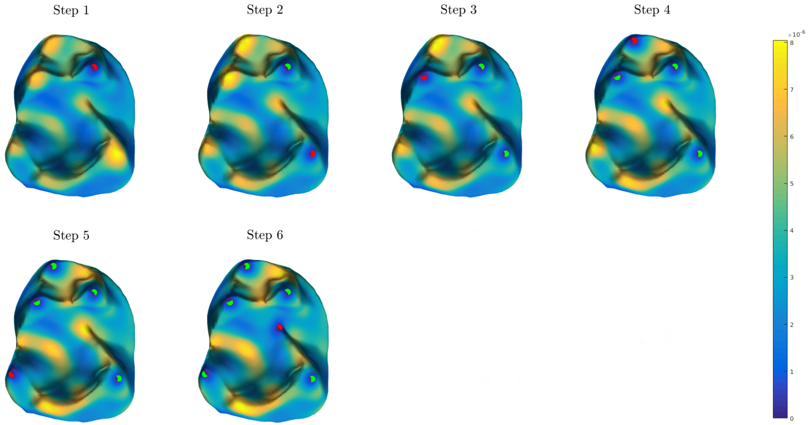

Step 1
Step 2
×10^-6
8
7
6
5
4
3
2
1
0

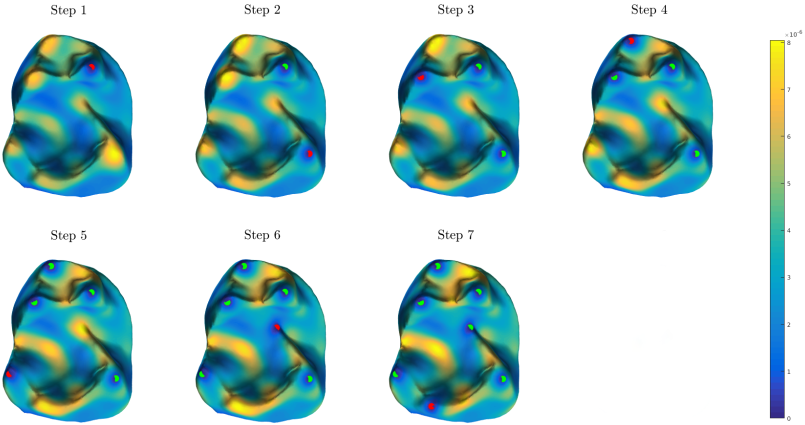


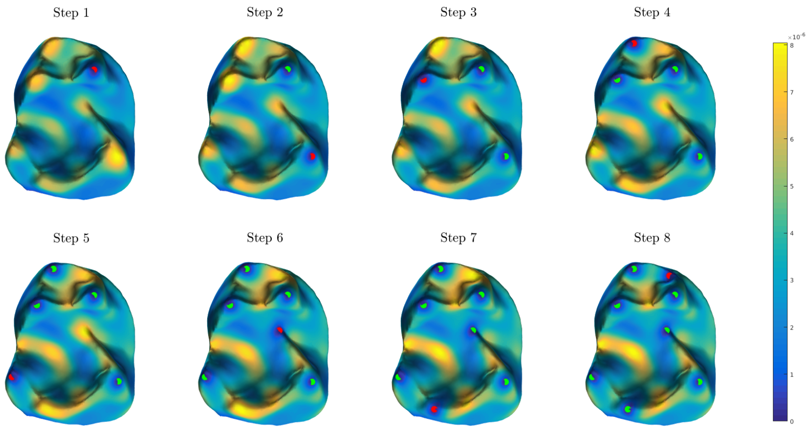
Step 1
Step 2
Step 3
×10^-6
8
7
6
5
4
3
2
1
0

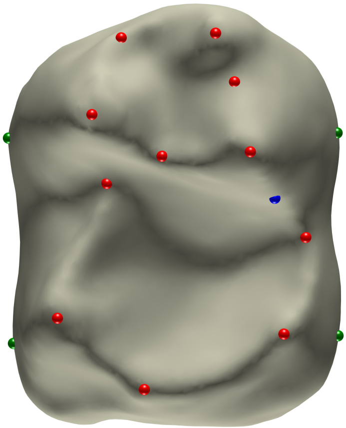

Step 1
Step 2
Step 3
Step 4
×10⁻⁶
8
7
6
5
4
3
2
1
0


Step 1
Step 2
Step 3
Step 4
Step 5
×10^-6
8
7
6
5
4
3
2
1
0

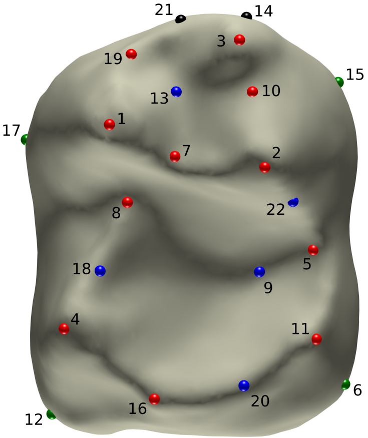

Step 1
Step 2
Step 3
Step 4
Step 5
Step 6
$\times 10^{-6}$
8
7
6
5
4
3
2
1
0


Step 1
Step 2
Step 3
Step 4
Step 5
Step 6
Step 7
×10^-6
8
7
6
5
4
3
2
1
0


Step 1
Step 2
Step 3
Step 4
Step 5
Step 6
Step 7
Step 8
×10^-6
8
7
6
5
4
3
2
1
0

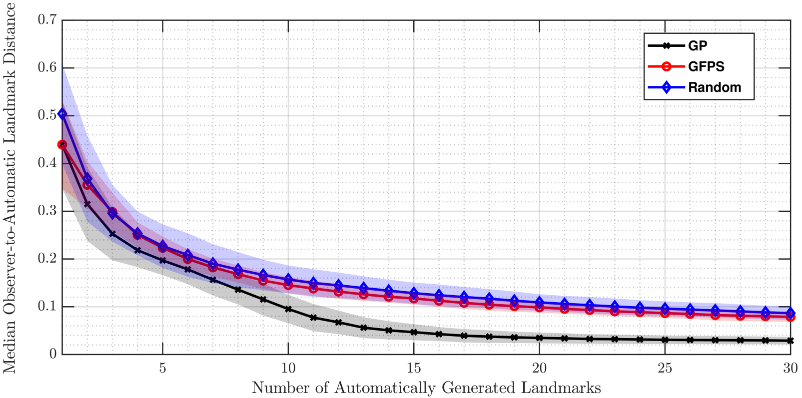


(a) Observer Landmarks

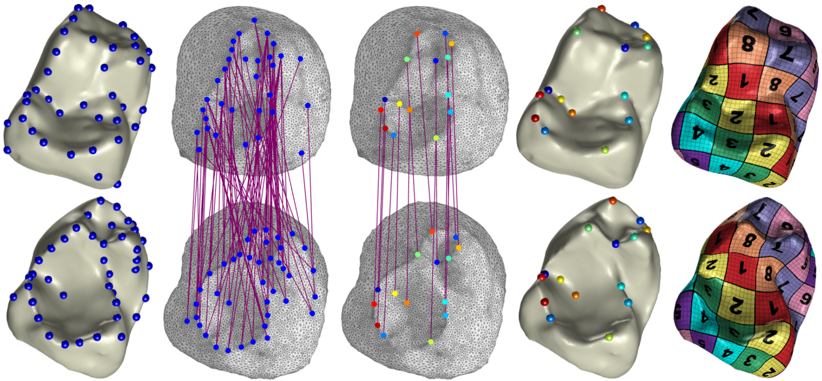


(b) First 22 Gaussian Process Landmarks

Efficient Adequate Coverage



Feature Matching by Bounded Distortion Filtering



(a) (b) (c) (d) (e)

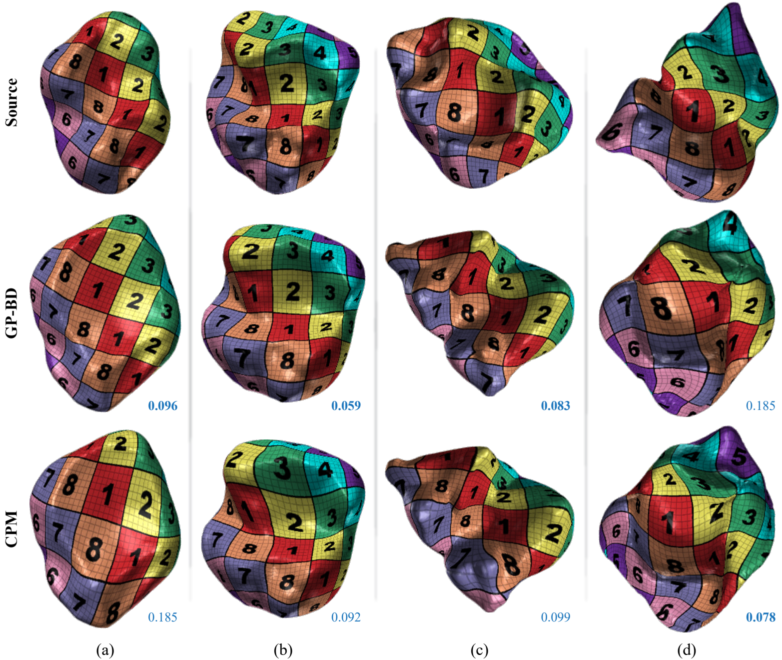
Feature Matching by Bounded Distortion Filtering

- GP landmarks $\left\{\zeta_1^{(1)}, \cdots, \zeta_L^{(1)}\right\}$ on S_1, $\left\{\zeta_1^{(2)}, \cdots, \zeta_L^{(2)}\right\}$ on S_2
- For each $\zeta_\ell^{(1)}$ propose T matches $\zeta_{\ell\to 1}^{(2)}, \cdots, \zeta_{\ell\to T}^{(2)}$ on S_2
- Solve the minimization problem (with *iterative reweighted least squares* (IRLS))

$$\min_{\Psi\in\mathscr{BD}(K)} \sum_{\ell=1}^{L}\sum_{t=1}^{T}\left\|\Psi\left(\zeta_\ell^{(1)}\right)-\zeta_{\ell\to t}^{(2)}\right\|_0$$

where $\mathscr{BD}(K)$ is the space of quasiconformal maps between S_1 and S_2 with conformal distortion bounded by $K \geq 0$

- Yaron Lipman, Stav Yagev, Roi Poranne, David W. Jacobs, and Ronen Basri. "Feature Matching with Bounded Distortion." *ACM Transactions on Graphics (TOG)*, 33, no. 3 (2014): 26.


Source
GP-BD
CPM
0.096
0.059
0.083
0.185
0.185
0.092
0.099
0.078
(a)
(b)
(c)
(d)

Outline

Motivation

- Geometric Morphometrics
- Experimental Design

Gaussian Process Landmarking

- Sequential Experimental Design
- **Witten Laplacian**
- Reduced Basis Method

Other Applications

Joint work with **Shahar Z. Kovalsky**, **Doug M. Boyer**, **Ingrid Daubechies**

Reweighted Kernel and the Witten Laplacian

$$K_\epsilon^c(x,y) = \int_M e^{-\frac{1}{2\epsilon}\|x-z\|_D^2} c(z)\, e^{-\frac{1}{2\epsilon}\|z-y\|_D^2} \mathrm{dvol}_M(z)$$

THM. (G. et al. 2019b). For $f \in C^2(M)$

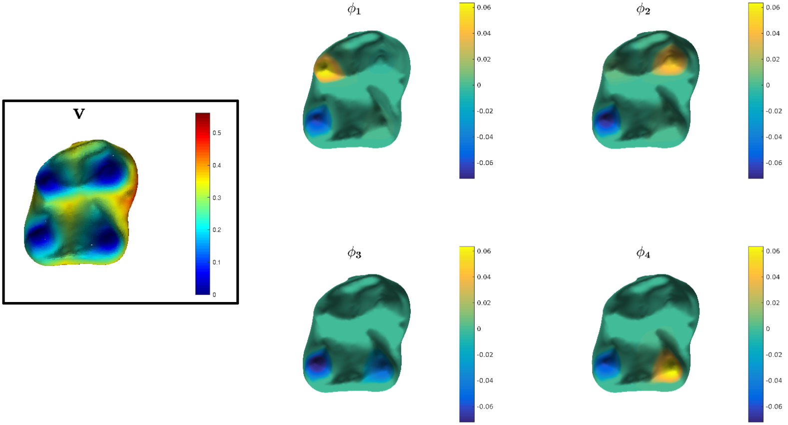
$$\lim_{\epsilon\to 0} \frac{1}{\epsilon}\left[\frac{\int_M K_\epsilon^c(x,y) f(y)\,\mathrm{dvol}_M(y)}{\int_M K_\epsilon^c(x,y)\,\mathrm{dvol}_M(y)} - f(x)\right] = \Delta f(x) + \nabla f(x)\cdot \nabla \log c(x).$$

I.e., the infinitesimal generator of the diffusion process defined by transition kernel $K_\epsilon^c(x,y)$ is conjugate to the Witten Laplacian

$$L_\epsilon = -\Delta - \frac{1}{\epsilon}\nabla \log c \cdot \nabla$$

• Tingran Gao, Shahar Z. Kovalsky, Doug M. Boyer, Ingrid Daubechies. "Gaussian Process Landmarking for Three-Dimensional Geometric Morphometrics." *SIAM Journal on Mathematics of Data Science*, 1(1), 237–267 (2019)

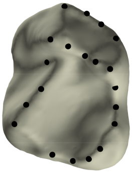
Localization of Eigenfunctions



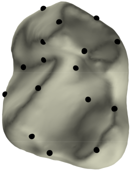
- Tingran Gao, Shahar Z. Kovalsky, Doug M. Boyer, Ingrid Daubechies. "Gaussian Process Landmarking for Three-Dimensional Geometric Morphometrics." *SIAM Journal on Mathematics of Data Science*, 1(1), 237–267 (2019)

The Importance of Reweighting

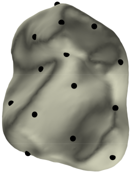
Gaussian Process Landmarking



Local Weight Maximum



Geodesic Farthest Point Sampling



- Tingran Gao, Shahar Z. Kovalsky, Doug M. Boyer, Ingrid Daubechies. “Gaussian Process Landmarking for Three-Dimensional Geometric Morphometrics.” *SIAM Journal on Mathematics of Data Science,* 1(1), 237–267 (2019)

Outline

Motivation

- Geometric Morphometrics
- Experimental Design

Gaussian Process Landmarking

- Sequential Experimental Design
- Witten Laplacian
- **Reduced Basis Method**

Other Applications

Joint work with **Shahar Z. Kovalsky**, **Doug M. Boyer**, **Ingrid Daubechies**

Sequential Experimental Design: Some Theory

- **Statistically:** If problem has submodularity (e.g. maximizing $\det(\mathscr{K}_n)$ in entropy-based sequential experimental design), can obtain near-optimality [Ko et al. 1995, Kause et al. 2008, Bouhtou et al. 2010]

$$\mathrm{OPT}_n \geq \mathrm{GPL}_n \geq (1-\epsilon)\,\mathrm{OPT}_n$$

- **Algorithmically:** This problem is NP-hard, life is short, find polynomial time approximations [Avron and Boutsidis 2013, Nikolov 2015, Wang et al. 2017, Allen-Zhu et al 2018]
- **Machine Learning:** Active Learning [Lewis and Gale 1994, Settles 2010]
- **Our Contribution:** Estimates for the decay rate of

$$\left\|\mathrm{MSE}\left(\hat{f}_n(\cdot)\right)\right\|_\infty = \sup_{x\in M} \mathrm{MSE}\left(\hat{f}_n(x)\right)$$

Faster Decay Than Any Inverse Polynomials

THM (G. et al. 2019a). Let M be a d-dimensional Riemannian manifold isometrically embedded in $\mathbb{R}^D$, with $d < D$. Let $x_1, \cdots, x_n$ be sequentially sampled on M using the Gaussian process landmarking algorithm. If K_ϵ^c is in $C^\ell(M)$, then for any $1 \leq k \leq \ell$ there exists $C_k > 0$ such that

$$\sup_{x \in M} \mathrm{MSE}\left(\hat{f}_n(x)\right) \leq C_k n^{-\frac{k}{d}}$$

for all sufficiently large $n \in \mathbb{N}$.

Faster Decay Than Any Inverse Polynomials

THM (G. et al. 2019a). Let M be a d-dimensional Riemannian manifold isometrically embedded in $\mathbb{R}^D$, with $d < D$. Let $x_1, \cdots, x_n$ be sequentially sampled on M using the Gaussian process landmarking algorithm. If K_ϵ^c is in $C^\ell(M)$, then for any $1 \leq k \leq \ell$ there exists $C_k > 0$ such that

$$\sup_{x \in M} \mathrm{MSE}\left(\hat{f}_n(x)\right) \leq C_k n^{-\frac{k}{d}}$$

for all sufficiently large $n \in \mathbb{N}$.

- Proof relies on the interpretation of the Gaussian process landmarking algorithm as applying the reduced basis method to the reproducing kernel Hilbert space associated with the Gaussian process $\mathrm{GP}(0, K_\epsilon^c)$.

• Tingran Gao, Shahar Z. Kovalsky, Ingrid Daubechies. "Gaussian Process Landmarking on Manifolds." *SIAM Journal on Mathematics of Data Science*, 1(1), 208–236 (2019)

Reproducing Kernel Hilbert Space for Gaussian Processes

- **Mercer's Theorem:** $K_\epsilon^c(x,y) = \sum_{k=0}^{\infty} e^{-\lambda_k} \phi_k(x)\,\phi_k(y)$
- **RKHS:** Closure under the RKHS norm of the set

$$\left\{ \sum_{i \in I} a_i K_\epsilon^c(\cdot, z_i) \mid a_i \in \mathbb{R}, z_i \in M, \mathrm{card}(I) < \infty \right\}$$

- **Feature Mapping:** $M \ni x \mapsto K_\epsilon^c(x, \cdot) \in \mathrm{RKHS}$

Key Observation: If $V_n := \mathrm{span}\{K_\epsilon^c(x_1, \cdot), \cdots, K_\epsilon^c(x_{n-1}, \cdot)\}$, then

$$\mathrm{MSE}\left(\hat{f}_{n-1}(x)\right) = \mathrm{dist}_{\mathrm{RKHS}}\left(K_\epsilon^c(x, \cdot), V_n\right)$$

Reduced Basis Method vs. Sequential Experimental Design

RBM	GP-SED
compact $K \subset \mathscr{H}$	compact $M^d \hookrightarrow \mathbb{R}^D$
$f_1 := \arg\max_{h \in K} \lVert h \rVert$	$x_1 := \arg\max_{x \in M} K_\epsilon^c(x, x)$
$f_2 := \arg\max_{h \in K} \mathrm{dist}_{\mathscr{H}}(h, V_1)$ $V_1 := \mathrm{span}\{f_0\}$	$x_2 := \arg\max_{x \in M} \mathrm{MSE}\left(\hat{f}_1(x)\right)$
$\vdots$	$\vdots$
$f_n := \arg\max_{h \in K} \mathrm{dist}_{\mathscr{H}}(h, V_{n-1})$ $V_{n-1} := \mathrm{span}\{f_0, \cdots, f_{n-1}\}$	$x_n := \arg\max_{x \in M} \mathrm{MSE}\left(\hat{f}_{n-1}(x)\right)$

Key Observation: If $V_n := \mathrm{span}\{K_\epsilon^c(x_1, \cdot), \cdots, K_\epsilon^c(x_{n-1}, \cdot)\}$, then

$$\mathrm{MSE}\left(\hat{f}_{n-1}(x)\right) = \mathrm{dist}_{\mathrm{RKHS}}\left(K_\epsilon^c(x, \cdot), V_n\right)$$

Greedy Algorithms in Reduced Basis Methods

$$d_n := \inf_{\substack{\dim Y = n \\ Y \subset \mathrm{RKHS}}} \sup_{x \in M} \mathrm{dist}_{\mathrm{RKHS}}\left(K_\epsilon^c(x,\cdot), Y\right)$$

$$\sigma_n := \sup_{x \in M} \mathrm{dist}_{\mathrm{RKHS}}\left(K_\epsilon^c(x,\cdot), V_n\right)$$

- $\sigma_n = \left\| \mathrm{MSE}\left(\hat{f}_n(\cdot)\right) \right\|_\infty$
- d_n is the Kolmogorov width, and $d_n = O\left(h_n^k\right)$ if M is a C^k manifold, where h_n is the *fill distance* with n points (Wendland 2004)
- Results from the reduced basis method (Binev 2011, DeVore et al. 2013) can be used to obtain $\sigma_n = O\left(d_{\lfloor n/2 \rfloor}^{1/2}\right)$

H. Wendland, **Scattered Data Approximation**, vol. 17, Cambridge University Press, 2004.

Greedy Algorithms in Reduced Basis Methods

Theorem (DeVore et al. 2013). For any $N \geq 0$, $n \geq 1$, and $1 \leq m < n$, there holds

$$\prod_{\ell=1}^{n} \sigma_{N+\ell}^2 \leq \left(\frac{n}{m}\right)^m \left(\frac{n}{n-m}\right)^{n-m} \sigma_{N+1}^{2m} d_m^{2n-2m}.$$

In particular, setting $N = 0$ and $m = \lfloor n/2 \rfloor$,

$$\sigma_n \leq \sqrt{2}\, \|K_\epsilon^c\|_{\infty, M\times M}^{\frac{1}{2}}\, d_{\lfloor n/2 \rfloor}^{\frac{1}{2}}$$

for all $n \in \mathbb{N}$, $n \geq 2$.

- P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, "Convergence Rates for Greedy Algorithms in Reduced Basis Methods." SIAM Journal on Mathematical Analysis, 43 (2011), pp. 1457–1472.
- R. DeVore, G. Petrova, and P. Wojtaszczyk, Greedy Algorithms for Reduced Bases in Banach Spaces, Constructive Approximation, 37 (2013), pp. 455–466.

Putting Everything Together......

- $\sigma_n = O\left(d_{\lfloor n/2 \rfloor}^{1/2}\right)$
- $d_n = O\left(h_n^{2k}\right)$
- $h_n = O\left(n^{-\frac{1}{d}}\right)$

Putting Everything Together......

- $\sigma_n = O\left(d_{\lfloor n/2 \rfloor}^{1/2}\right)$
- $d_n = O\left(h_n^{2k}\right)$
- $h_n = O\left(n^{-\frac{1}{d}}\right)$

Conclusion: The sequential, greedy algorithm guarantees

$$\left\| \mathrm{MSE}\left(\hat{f}_n(\cdot)\right) \right\|_\infty = \sigma_n \leq C_k n^{-\frac{k}{d}} \quad \text{for } C_k > 0, \text{ if } M \in C^k$$

Putting Everything Together......

- $\sigma_n = O\left(d_{\lfloor n/2 \rfloor}^{1/2}\right)$
- $d_n = O\left(h_n^{2k}\right)$
- $h_n = O\left(n^{-\frac{1}{d}}\right)$

Conclusion: The sequential, greedy algorithm guarantees

$$\left\| \mathrm{MSE}\left(\hat{f}_n(\cdot)\right) \right\|_\infty = \sigma_n \leq C_k n^{-\frac{k}{d}} \quad \text{for } C_k > 0, \text{ if } M \in C^k$$

- In particular, if $M \in C^\infty$, Gaussian process landmarking guarantees that MSE decays faster than any inverse polynomial in n
- Open Question: Exponential convergence?

Outline

Motivation

- Geometric Morphometrics
- Experimental Design

Gaussian Process Landmarking

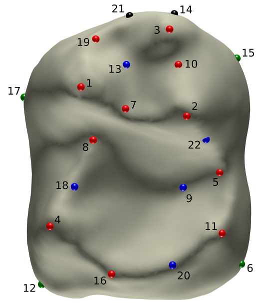
- Sequential Experimental Design
- Witten Laplacian
- Reduced Basis Method

Other Applications

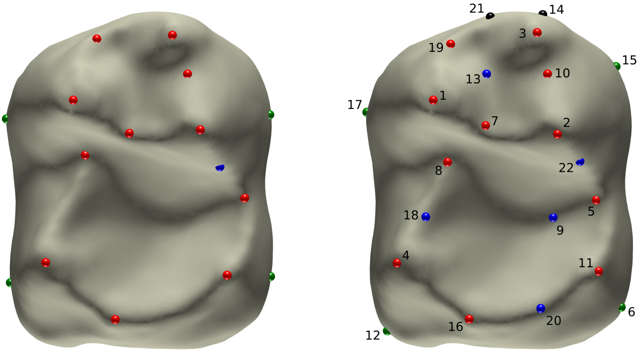
Joint work with **Shahar Z. Kovalsky**, **Doug M. Boyer**, **Ingrid Daubechies**

Post-Stories

- Ravier, Robert J. “Eyes on the Prize: Improved Registration via Forward Propagation.” arXiv preprint arXiv:1812.10592 (2018).
- Shan, Shan, Shahar Z. Kovalsky, Julie M. Winchester, Doug M. Boyer, and Ingrid Daubechies. “ariaDNE: A robustly implemented algorithm for Dirichlet energy of the normal.” *Methods in Ecology and Evolution*, 10, no. 4 (2019): 541–552.



Thank You!



(a) Observer Landmarks **(b) First 22 Gaussian Process Landmarks**

- Tingran Gao, Shahar Z. Kovalsky, Ingrid Daubechies. "Gaussian Process Landmarking on Manifolds." *SIAM Journal on Mathematics of Data Science,* 1(1), 208–236 (2019)
- Tingran Gao, Shahar Z. Kovalsky, Doug M. Boyer, Ingrid Daubechies. "Gaussian Process Landmarking for Three-Dimensional Geometric Morphometrics." *SIAM Journal on Mathematics of Data Science,* 1(1), 237–267 (2019)