Manifold Learning on Fibre Bundles

Tingran Gao

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

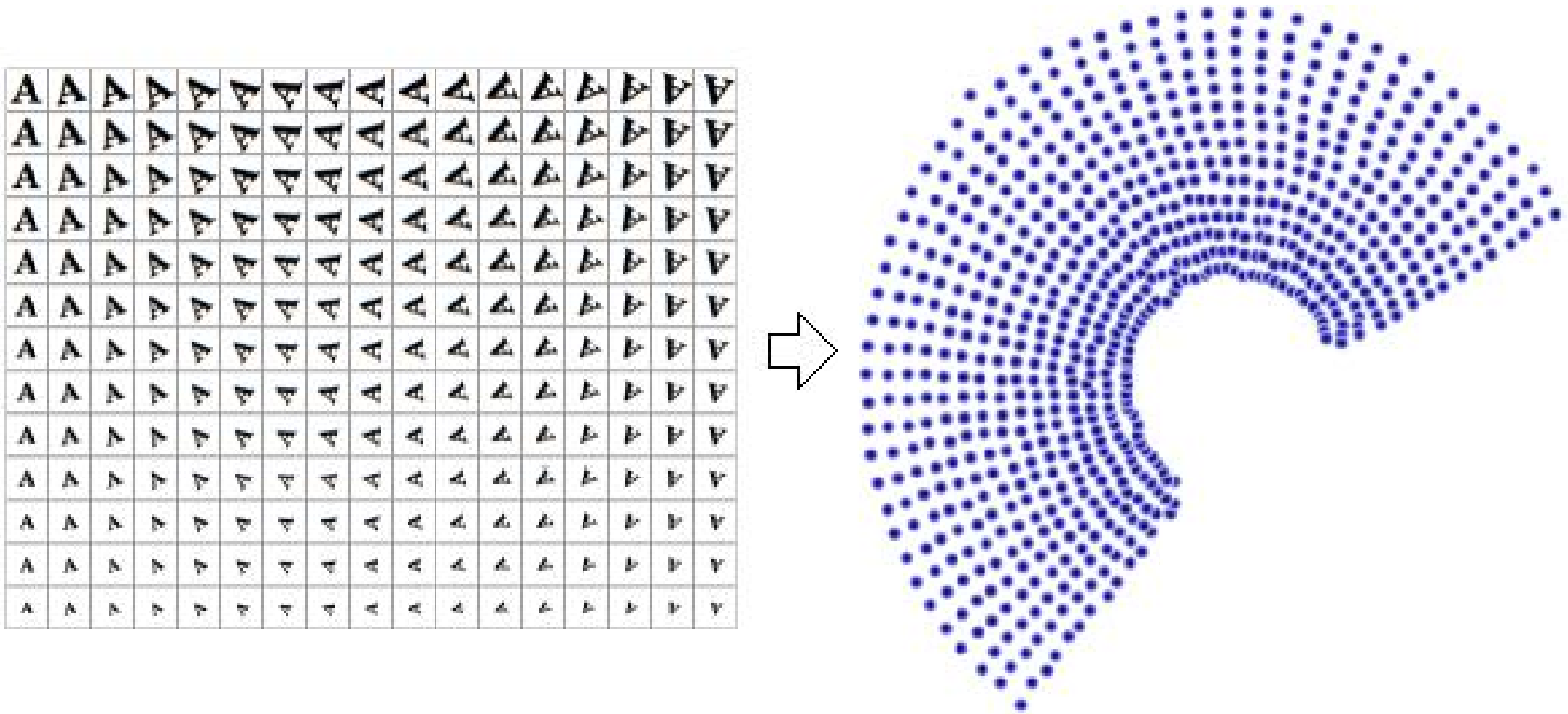
Computational and Applied Mathematics Colloquium
The University of Chicago, Chicago IL

Thursday May 16, 2019

Outline

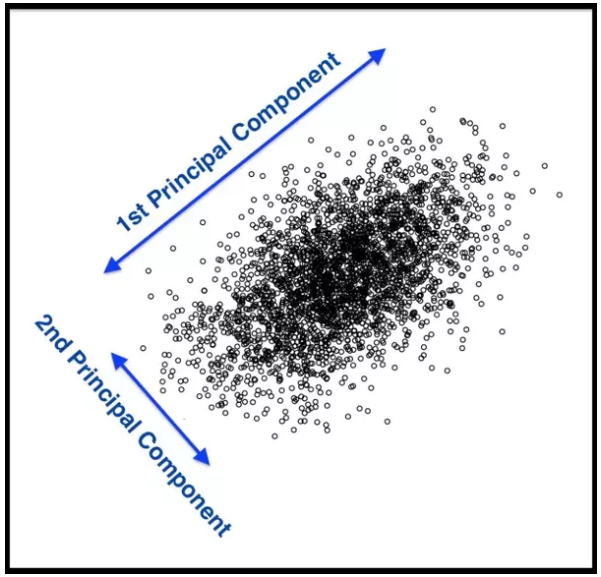
Manifold Learning

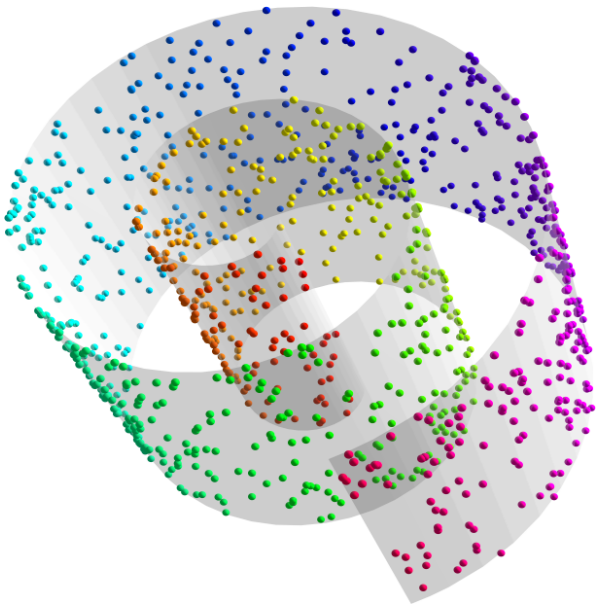
Uncovering the Manifold Structure in the Data



Manifold Learning

- **Motivation**
  - *Curse of Dimensionality*: For D-dimensional data, need $\Omega\left(\epsilon^{-1}\exp\left(D\log\frac{1}{\epsilon}\right)\right)$ samples to ensure estimation error $\leq \epsilon$
  - *Manifold Assumption*: Data lie approximately on a d-dimensional submanifold of $\mathbb{R}^D$, with $d \ll D$; expecting sampling complexity $\Omega\left(\epsilon^{-1}\exp\left(d\log\frac{1}{\epsilon}\right)\right)$ instead





Manifold Learning and Discrete Geometry – Various Interpretations

- **Nonparametric Statistics:** Regression and density estimation under manifold assumptions (L. Wasserman, P. Bickel et al.)
- **Gaussian Processes:** Gaussian process latent variable models (N. Lawrence et al.)
- **Coarse Geometry:** Geometric Whitney Problem – Metric space approximation under the Gromov–Hausdorff distance (C. Fefferman et al.)
- **Finite Elements:** Discrete exterior calculus (P. Schröder, D. Arnold et al.)

Manifold Learning and Discrete Geometry – Various Interpretations

- **Nonparametric Statistics:** Regression and density estimation under manifold assumptions (L. Wasserman, P. Bickel et al.)
- **Gaussian Processes:** Gaussian process latent variable models (N. Lawrence et al.)
- **Coarse Geometry:** Geometric Whitney Problem – Metric space approximation under the Gromov–Hausdorff distance (C. Fefferman et al.)
- **Finite Elements:** Discrete exterior calculus (P. Schröder, D. Arnold et al.)
- **Spectral Geometry**: Laplacian Eigenmaps (M. Belkin & P. Niyogi, S. Lafon & R. Coifman et al.)

Spectral Geometry

CAN ONE HEAR THE SHAPE OF A DRUM?

MARK KAC, The Rockefeller University, New York

To George Eugene Uhlenbeck on the occasion of his sixty-fifth birthday

"La Physique ne nous donne pas seulement l'occasion de résoudre des problèmes . . . , elle nous fait presentir la solution." H. POINCARÉ.

Spectral Geometry

$$\Delta_M u_n = -\lambda_n u_n, \quad n = 0, 1, 2, \ldots$$
$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \nearrow \infty$$

- Does knowing all λ_n's determine M up to isometry? (**Spoiler Alert: No**)

Spectral Geometry

$$\Delta_M u_n = -\lambda_n u_n, \quad n = 0, 1, 2, \ldots$$
$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \nearrow \infty$$

- Does knowing all λ_n's determine M up to isometry? (**Spoiler Alert: No**)
- Isospectral but non-isomstric:
  - Flat tori (Milnor 1964)
  - Riemann surfaces with constant negative curvature (Fignéras 1980)
  - Lens spaces with constant curvature (Ikeda 1983)
  - Riemannian covering spaces (Sunada 1985)

Spectral Geometry

$$\Delta_M u_n = -\lambda_n u_n, \quad n = 0, 1, 2, \ldots$$
$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \nearrow \infty$$

- Does knowing all λ_n's determine M up to isometry? (**Spoiler Alert: No**)
- Isospectral but non-isomstric:
  - Flat tori (Milnor 1964)
  - Riemann surfaces with constant negative curvature (Fignéras 1980)
  - Lens spaces with constant curvature (Ikeda 1983)
  - Riemannian covering spaces (Sunada 1985)
- Cospectral graphs

Spectral Geometry

$$\Delta_M u_n = -\lambda_n u_n, \quad n = 0, 1, 2, \ldots$$
$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \nearrow \infty$$

- Does knowing all λ_n's **and all $\{u_n\}$'s** determine M up to isometry?

Spectral Geometry

$$\Delta_M u_n = -\lambda_n u_n, \quad n = 0, 1, 2, \ldots$$
$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \nearrow \infty$$

- Does knowing all λ_n's **and all $\{u_n\}$'s** determine M up to isometry?
  Yes – via the *heat kernel*: $\forall x, y \in M, t \in [0, \infty)$

$$k_M(x, y; t) = \sum_{n=0}^{\infty} e^{-\lambda_n t} u_n(x) u_n(y)$$

- Heat kernel completely determines the metric up to isometry

Spectral Embedding

(Bérard-Besson-Gallot 1994) Any closed Riemannian manifold in

$$\mathcal{M}_{n,k,D} = \left\{ (M,g) \,\middle|\, \dim(M) = n, \mathrm{Ric}(g) \geq (n-1)\,kg, \right.$$
$$\left. \mathrm{diam}(M) \leq D \right\}$$

can be embedded into the infinite dimensional Hilbert space ℓ^2 using the *heat kernel map*

$$M \ni x \longmapsto \Phi_t(x) := \left(e^{-\lambda_0 t/2} u_0(x), e^{-\lambda_1 t/2} u_1(x), \ldots \right) \in \ell^2$$

- $\langle \Phi_t(x), \Phi_t(y) \rangle_{\ell^2} = k_M(x,y;t)$ RKHS
- Estimates for $k_M \leftrightarrow$ (Gromov) Precompactness of $\mathcal{M}_{n,k,D}$

Spectral Embedding: Beyond BBG'94

- **(Jones-Maggioni-Schul 2010)** **Local**, **finitely** many eigenfunctions used to make **bi-Lipschitz** coordinate charts
- **(Bates 2014)** Make (Jones-Maggioni-Schul 2010) **global**
- **(Wang-Zhu 2015)** Heat kernel maps can be made **isometric** using Nasher-Moser
- **(Portegies 2016)** **Global**, **finite**, and **almost isomstric** embedding using harmonic radius arguments
- **(Wu 2017)** BGG'94-type embedding, but with the heat kernels of the **connection Laplacian** (rough Laplacian)
- **(Lin-Wu 2018)** Embedding in (Wu 2017) can be made **finite**

Data Analysis: Discrete/Combinatorial/Probabilistic

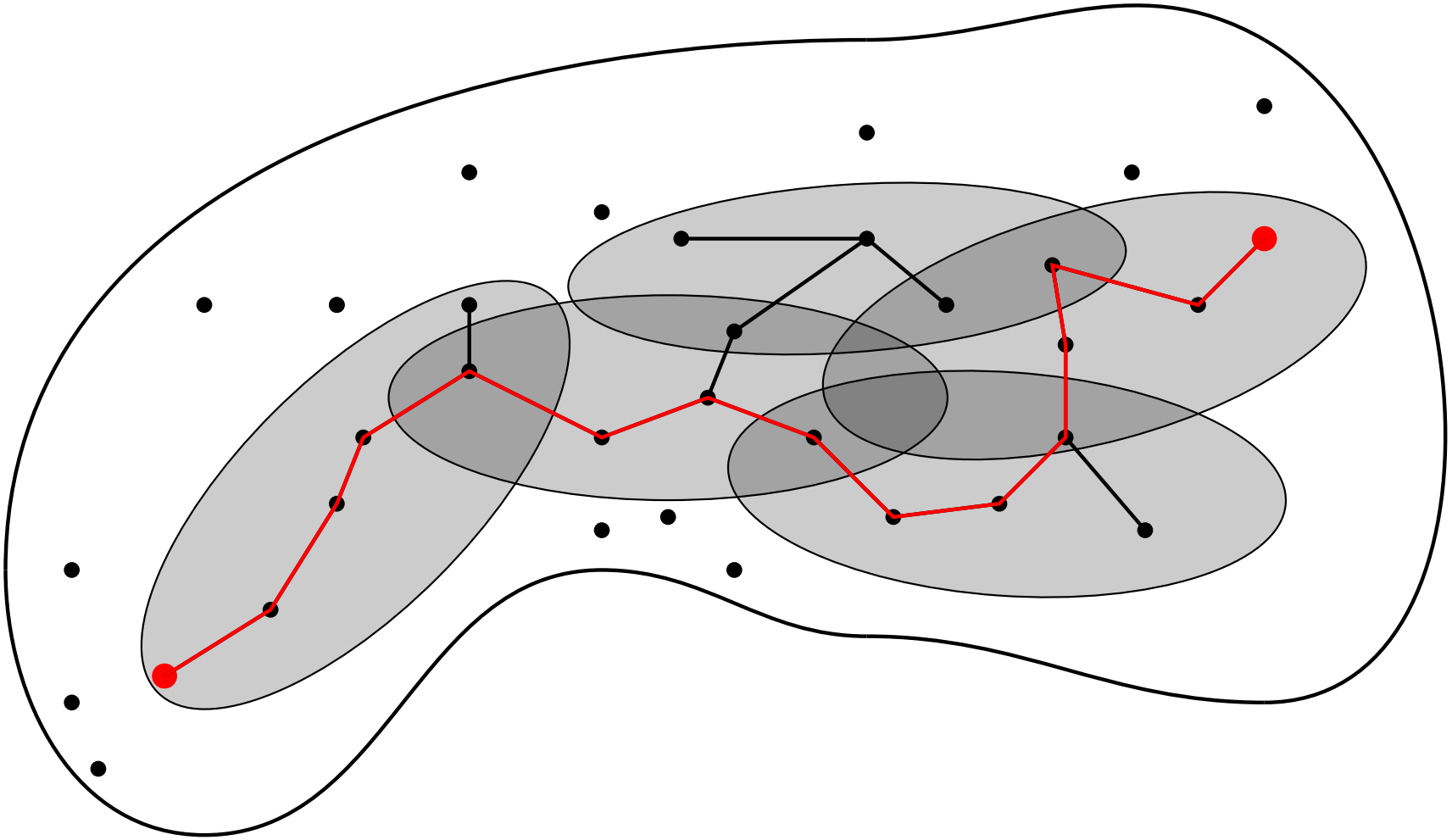
Underpinning methodology **(Lim 2015)**:

Graphs are discrete Riemannian manifolds

Riemannian Manifold	**Graph**
tangent vectors	edges
Laplacian	graph Laplacian
heat kernel	heat kernel
diffusion process	random walk

L.-K. Lim, "Hodge Laplacians on Graphs," arxiv:1507.05319
M. Belkin, P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Computation* 15 (6), 1373-1396
R. Coifman, S. Lafon, "Diffusion Maps," *Applied and Computational Harmonic Analysis* 21 (2006), no. 1, 5-30

Random Walks "Knit Together" Local Geometry



Small distances are much more reliable!

Diffusion Maps

- Data $\mathscr{X} := \{x_1, \ldots, x_n\} \subset \mathbb{R}^D$, point cloud
- Build *weighted* nearest-neighbor graph $G = (V, E, w)$ for $\mathscr{X}$
- Define weighted adjacency matrix W for G, and diagonal matrix $D \in \mathbb{R}^{n \times n}$ with

$$D_{ii} = \sum_{k=1}^{n} w_{ij}$$

- Build graph random walk Laplacian $L = I_n - D^{-1}W$, and perform eigen-decomposition

$$Lu_i = \lambda_i u_i, \quad i = 1, \ldots, n$$

- For any $1 \leq d \leq n$, embed $\mathscr{X}$ into $\mathbb{R}^d$ by

$$x_k \mapsto \left(\lambda_1^{1/2} u_1(k), \ldots, \lambda_d^{1/2} u_d(k)\right)$$

Graph Random Walk Laplacian

- $D^{-1}W$ defines a random walk on the graph G
- Typical choice for weights: For some *bandwidth* parameter $\epsilon > 0$,
$$W_{ij} = \exp\left(-\|x_i - x_j\|^2 / \epsilon\right)$$
- The choice of ϵ reflects our belief for the "locality" of the data
- If we eigen-decompose W instead of $D^{-1}W$, the algorithm is known as *Laplacian eigenmaps*

Continuous Limit

- Explicitly, $D^{-1}W$ is a Markov operator: for any $v \in \mathbb{R}^n$,

$$D^{-1}Wv = \frac{\sum_{j=1}^{n} w_{ij} v_j}{\sum_{j=1}^{n} w_{ij}}$$

- As $n \to \infty$, $D^{-1}W$ converges weakly to the integral operator

$$P_\epsilon f(x) = \frac{\int_M \exp\left(-\|x-y\|^2/\epsilon\right) f(y)\, \mathrm{dvol}(y)}{\int_M \exp\left(-\|x-y\|^2/\epsilon\right) \mathrm{dvol}(y)}$$

for all $f \in L^1(M)$.

Diffusion Maps: Asymptotic Theory

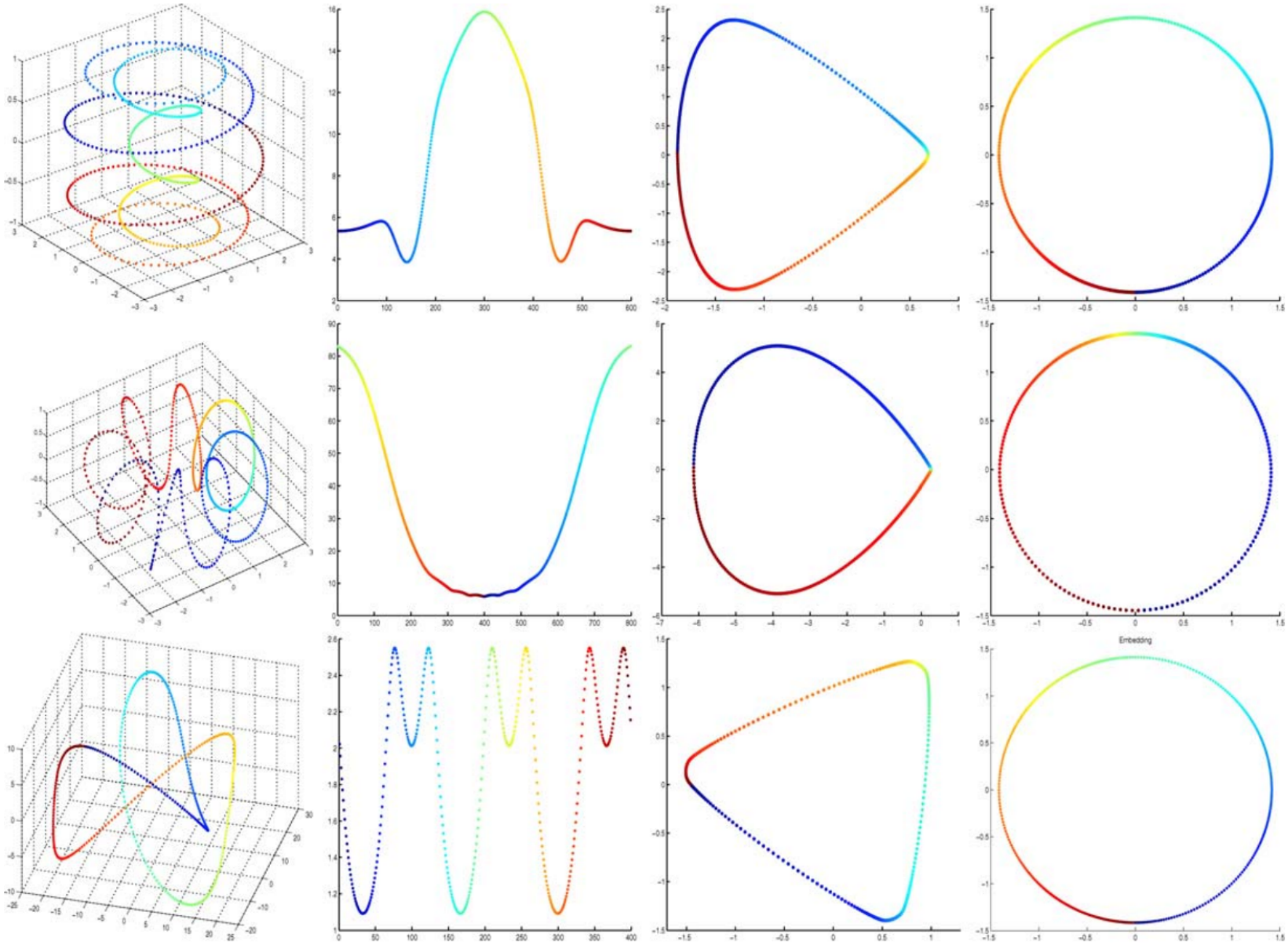
Theorem (Belkin-Niyogi, 2005). Let data points $x_1, \cdots, x_n$ be sampled from a **uniform** distribution on M. Under mild technical assumptions, there exist a sequence of real numbers $t_n \to 0$ and a constant C such that for any $f \in C^\infty(M)$

$$\lim_{n\to\infty} C \frac{(4\pi t_n)^{-\frac{k+2}{2}}}{n} \frac{P_{t_n} - I}{t_n} f(x) = \Delta_M f(x), \quad \forall x \in M.$$

Theorem (Coifman-Lafon, 2006). As $\epsilon \to 0$, for any $f \in C^\infty(M)$ and $x \in M$, if $\{x_i\}_{i=1}^n \sim p(x)\mathrm{dvol}_M(x)$, then w.h.p.

$$P_\epsilon^{(\alpha)} f(x)$$
$$= f(x) + \epsilon \frac{m_2}{2m_0} \left[\frac{\Delta_M \left[f p^{1-\alpha} \right](x)}{p^{1-\alpha}(x)} - f(x) \frac{\Delta_M p^{1-\alpha}(x)}{p^{1-\alpha}(x)} \right] + O(\epsilon^2).$$

Decoupling Geometry from Probability



The Convergence Rate: Diffusion Maps

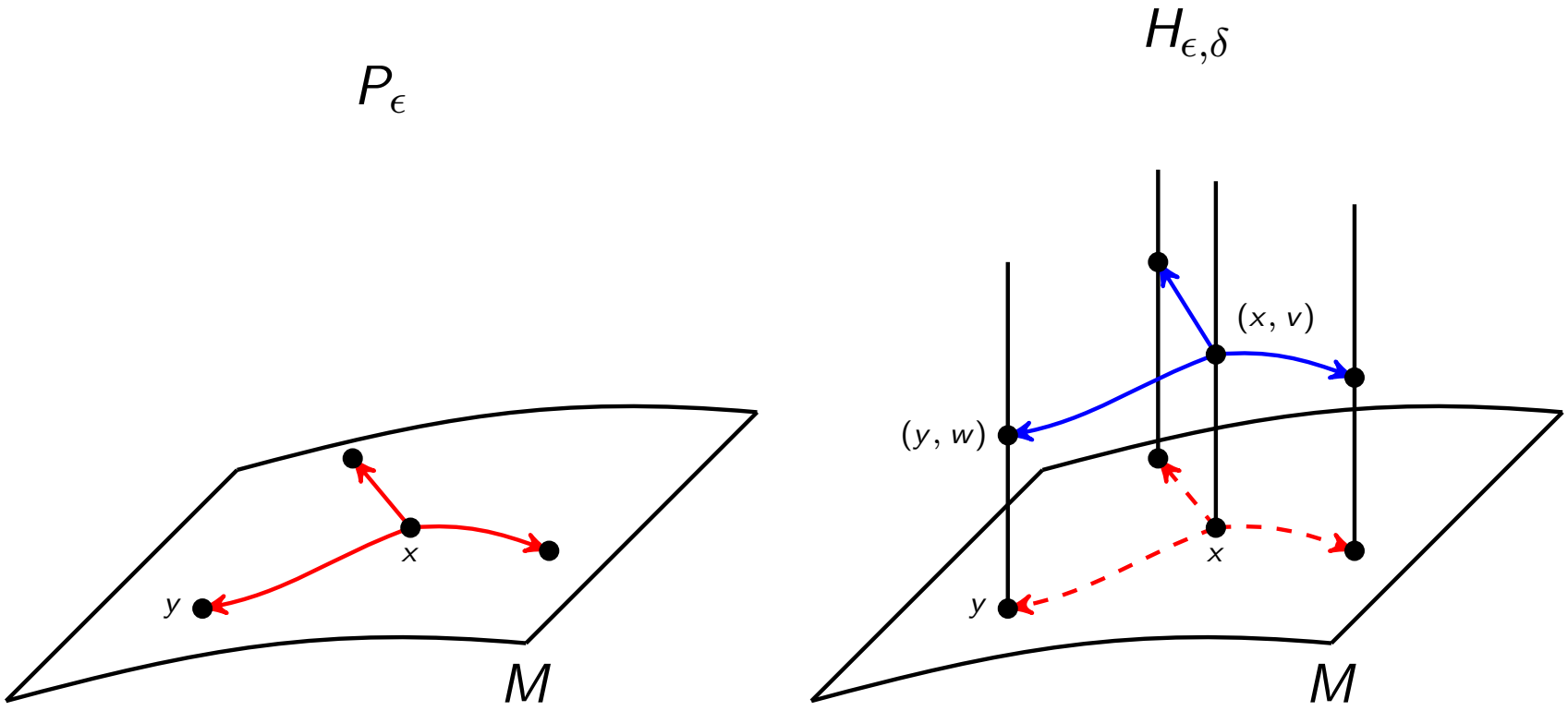
Theorem (Singer, 2006). Suppose N points are i.i.d. uniformly sampled from a d-dimensional Riemannian manifold M. The graph diffusion operator $P_{\epsilon,\alpha}$ converges to its smooth limit at rate

$$O\left(N^{-\frac{1}{2}}\epsilon^{\frac{1}{2}-\frac{d}{4}}\right).$$

Corollary. Under the same assumption, non-uniform sampling has convergence rate

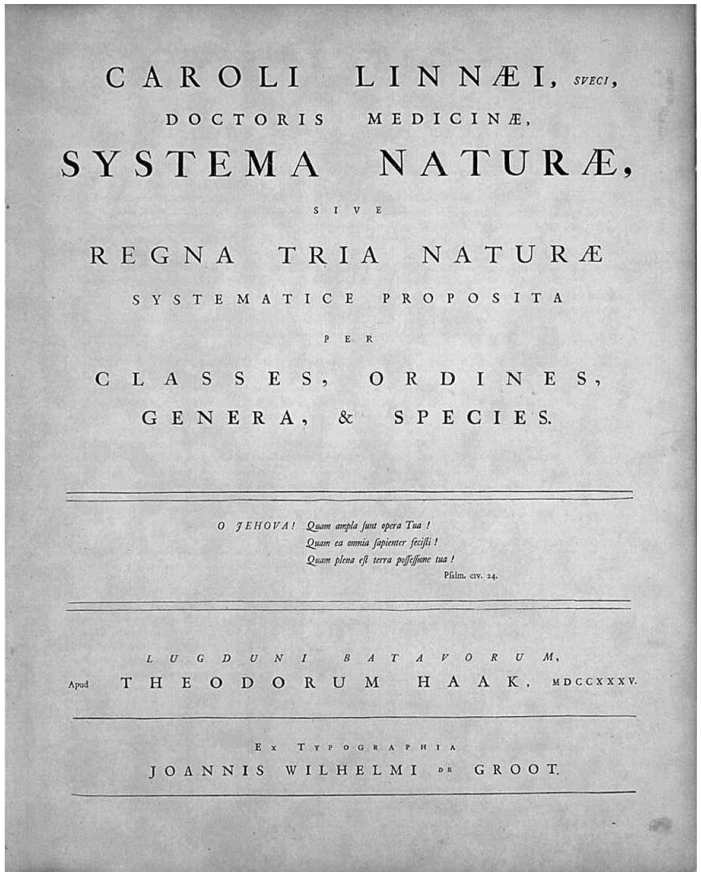
$$O\left(N^{-\frac{1}{2}}\epsilon^{-\frac{d}{4}}\right).$$

This Talk: *Horizontal* Diffusion Maps on Fibre Bundles



Outline

A Motivating Case Study: Morphology and Classification



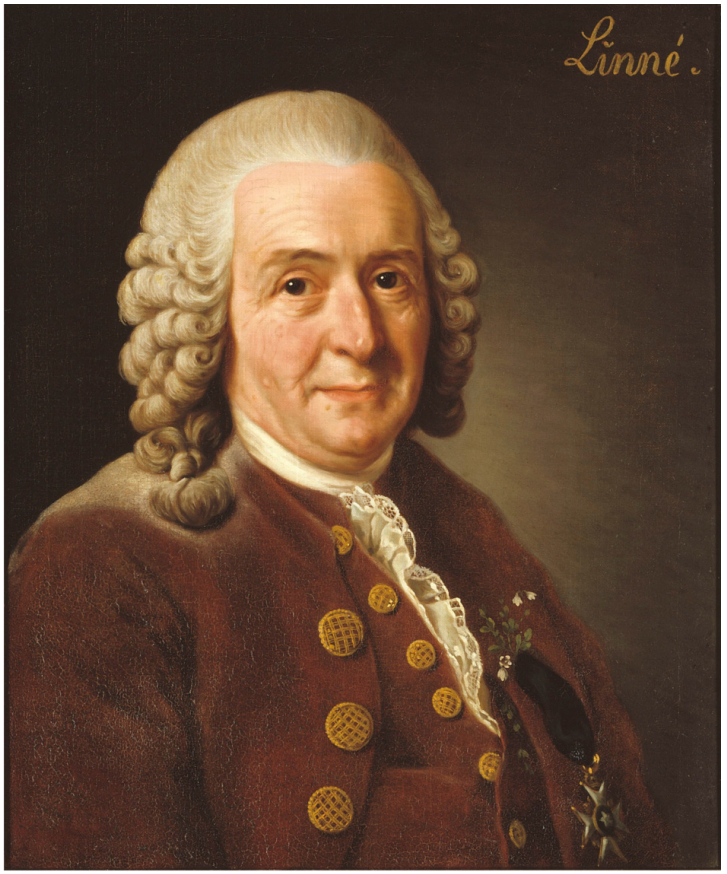
CAROLI LINNÆI, SVECI,
DOCTORIS MEDICINÆ,
SYSTEMA NATURÆ,
SIVE
REGNA TRIA NATURÆ
SYSTEMATICE PROPOSITA
PER
CLASSES, ORDINES,
GENERA, & SPECIES.

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



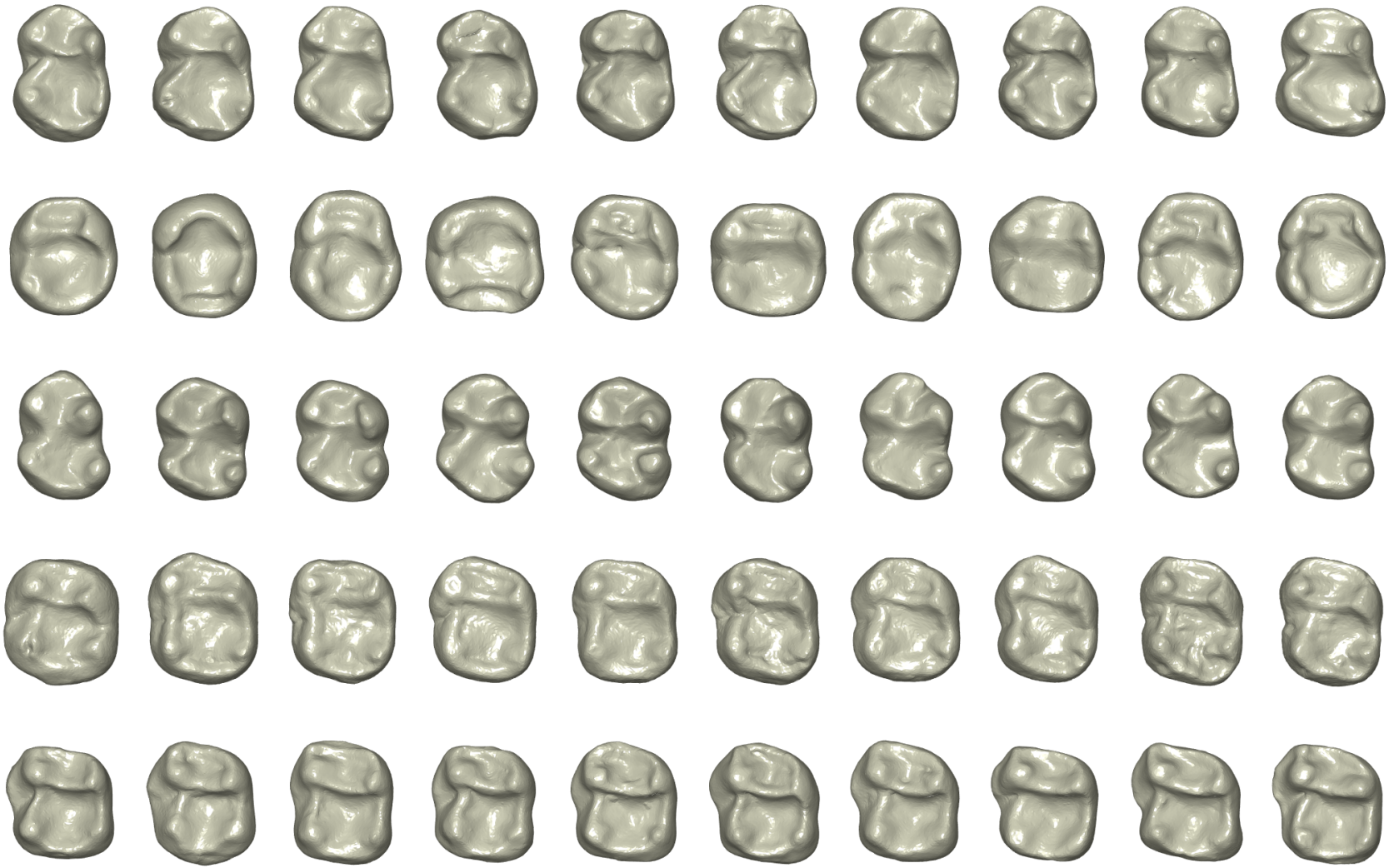
Carl Linnaeus (1707-1778)

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.

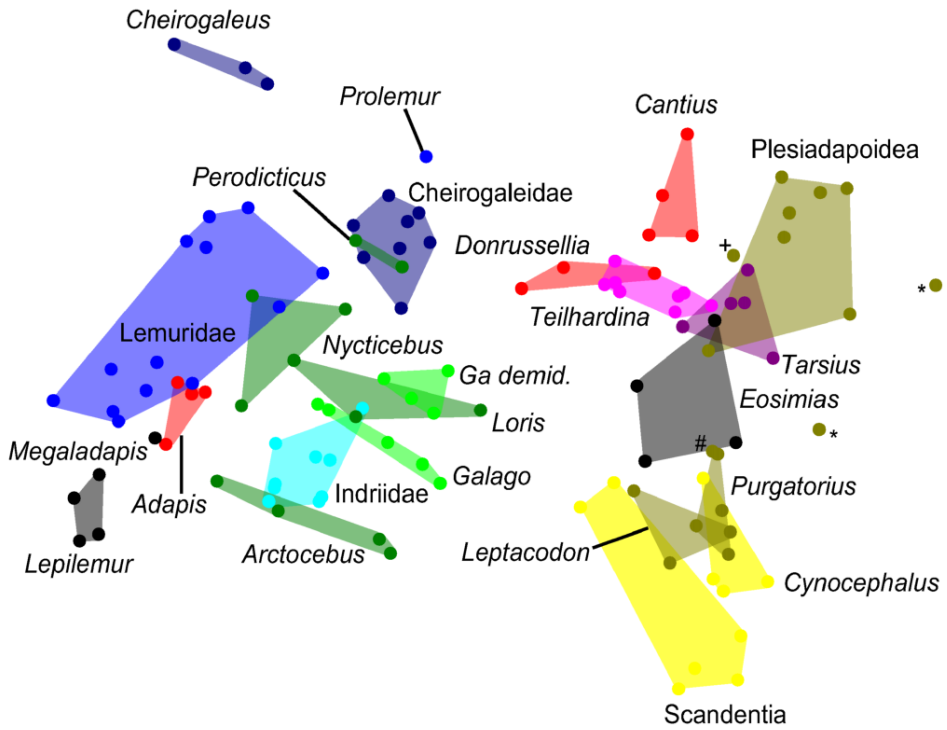
Geometric Morphometrics



Landmark-Based Geometric Morphometrics

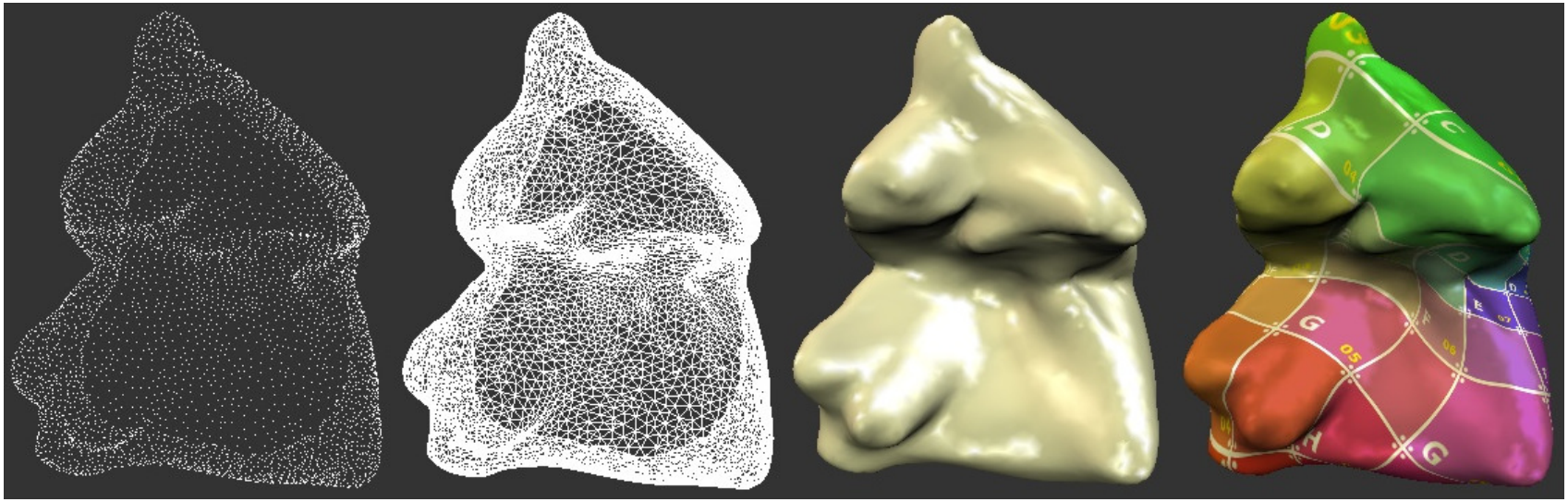
$$\left(S_1, \{x_j\}_{j=1}^{J}\right), \left(S_2, \{y_j\}_{j=1}^{J}\right) \longrightarrow$$

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



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.

Data Acquisition: microCT (High-Resolution X-ray CT)



Surface reconstructed from μCT-scanned voxel data

Data Acquisition: *Morphosource.org*

Sharing the Bones

Duke researchers bring digital tools to the Stone Age findings of the Rising Star cave expedition

Writer: Louise Flynn
December 11, 2015

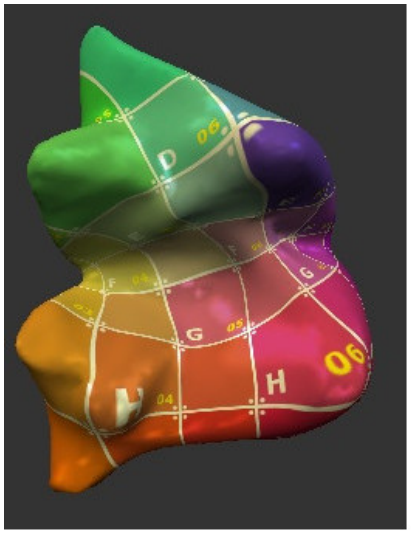
Streaked with ancient limestone dust, the scientist hovering near Steve Churchill's worktable waited to see his reaction to the fossil she'd surgically removed from its eons-old resting place eighty meters below ground. Crammed in the tent's open window, a camera crew from *National Geographic* gathered more footage. Researchers came in and out,

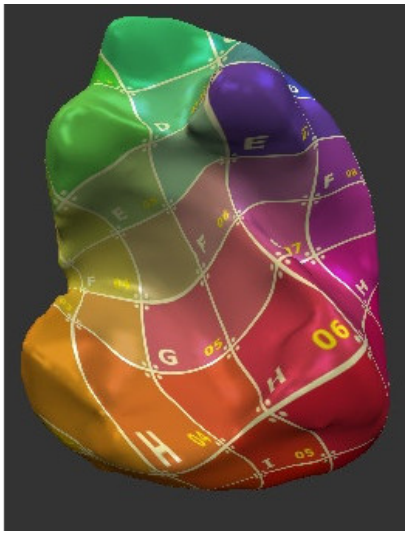
"Big Data" for Biologists: Impossible to landmark them all!!

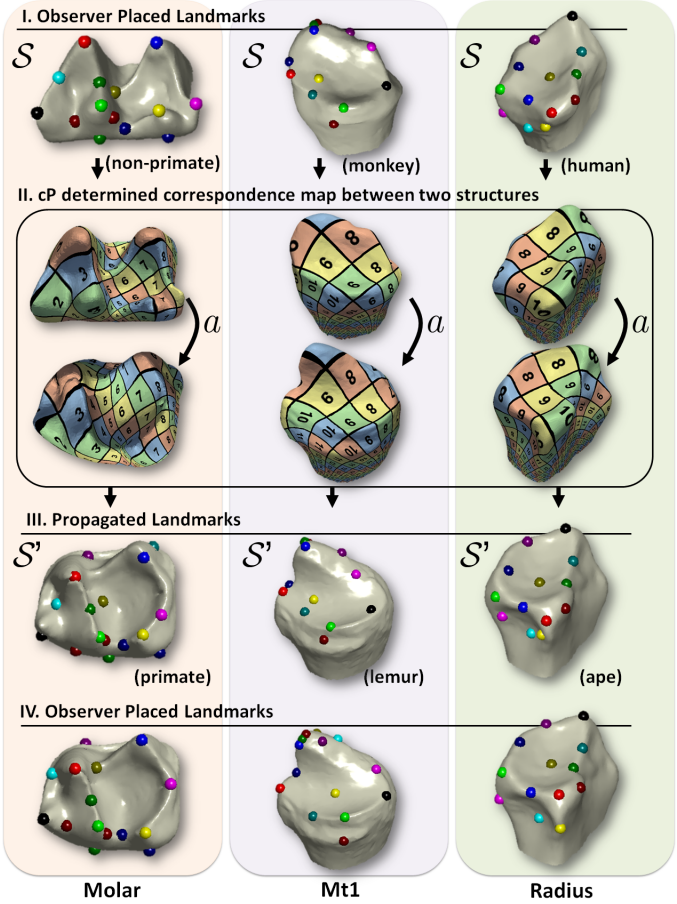
A Zoo of (Landmark-Free) Shape Distances...

$d_{\mathrm{cWn}}(S_1, S_2)$: Conformal Wasserstein Distance (CWD)
$d_{\mathrm{cP}}(S_1, S_2)$: Continuous Procrustes Distance (CPD)
$d_{\mathrm{cKP}}(S_1, S_2)$: Continuous Kantorovich-Procrustes Distance (CKPD)

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

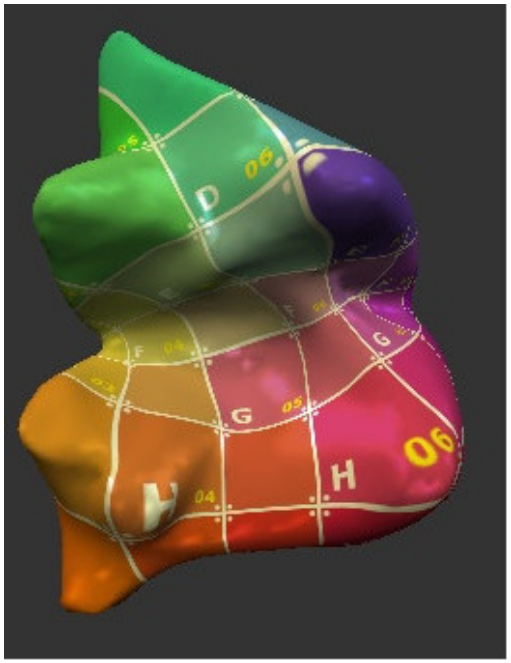




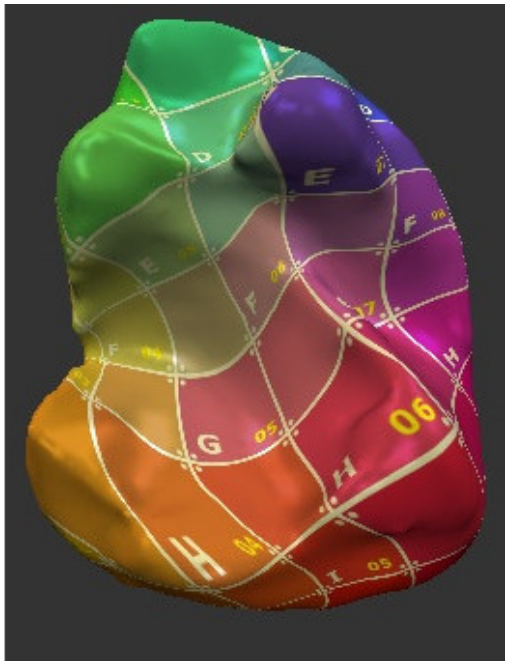


$$d_{\mathrm{cP}}\left(S_i, S_j\right) = \inf_{\mathcal{C}\in\mathcal{A}\left(S_i,S_j\right)} \inf_{R\in\mathbb{E}(3)} \left(\int_{S_i} \left\|R\left(x\right) - \mathcal{C}\left(x\right)\right\|^2 d\mathrm{vol}_{S_i}\left(x\right)\right)^{\frac{1}{2}}$$

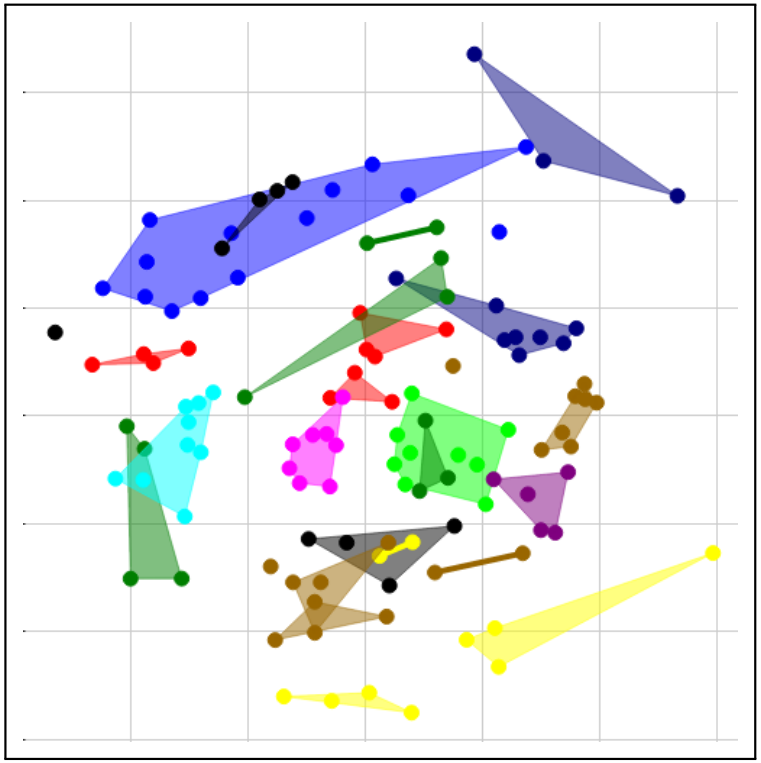
S_i



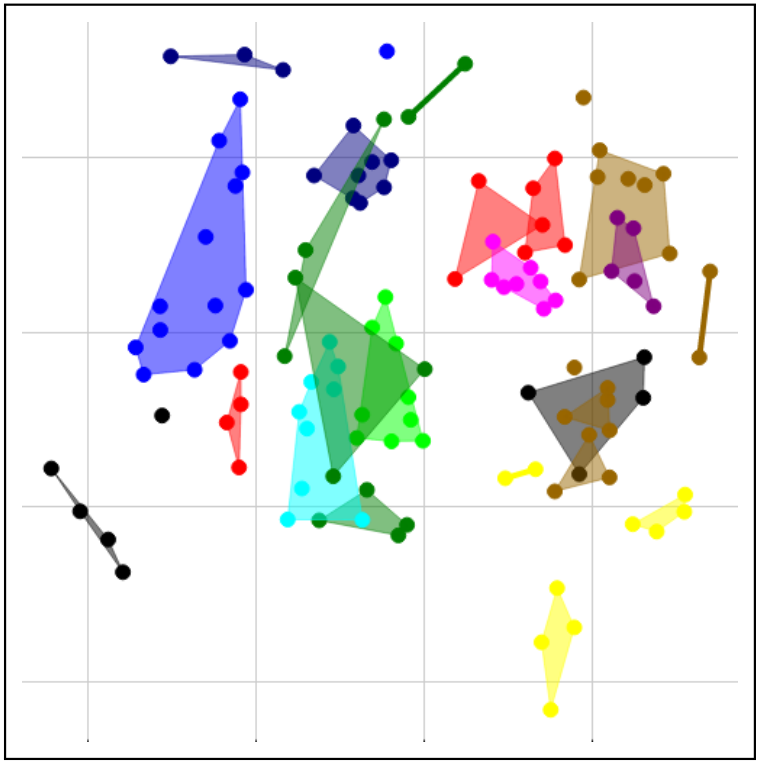
$\xrightarrow[f_{ij}]{d_{ij}}$



S_j

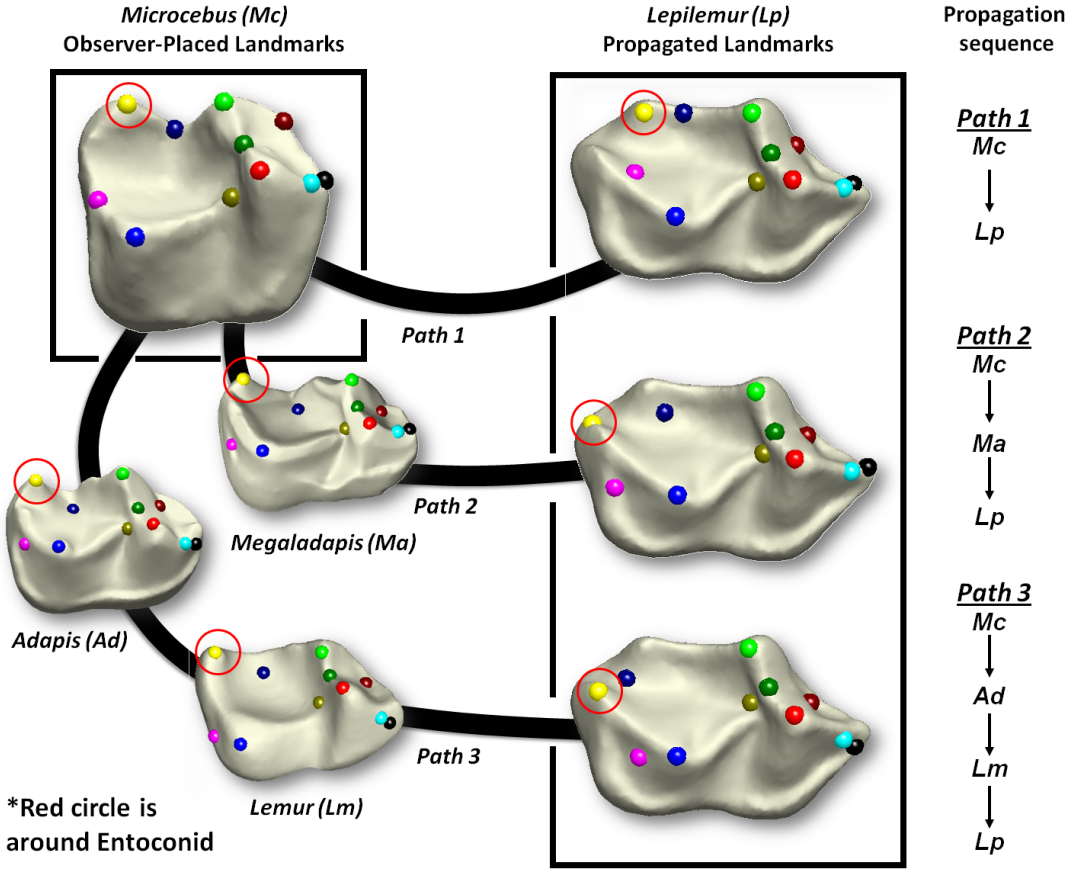


landmark-free



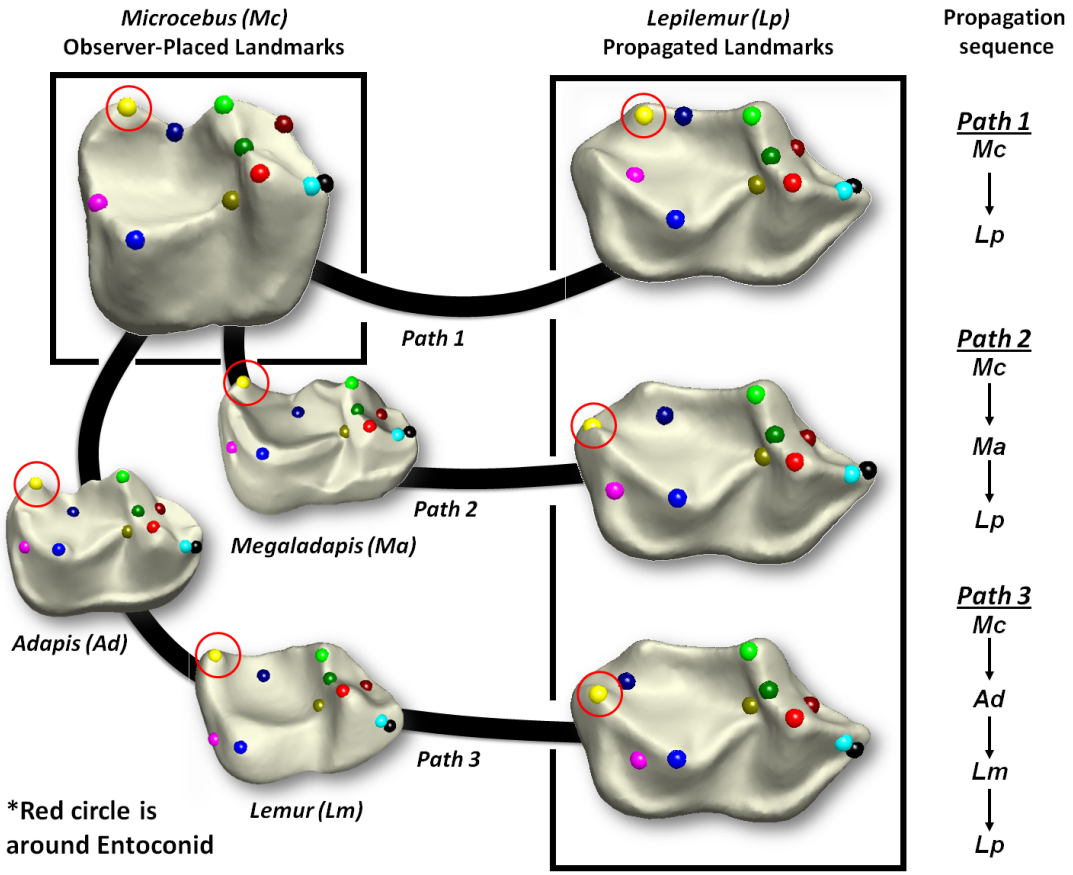
landmark

Interpretability Issue



Even mistakes made by CPD were similar to biologists' mistakes!

Trust Small Distances — Let Diffusion Maps Do it for You!

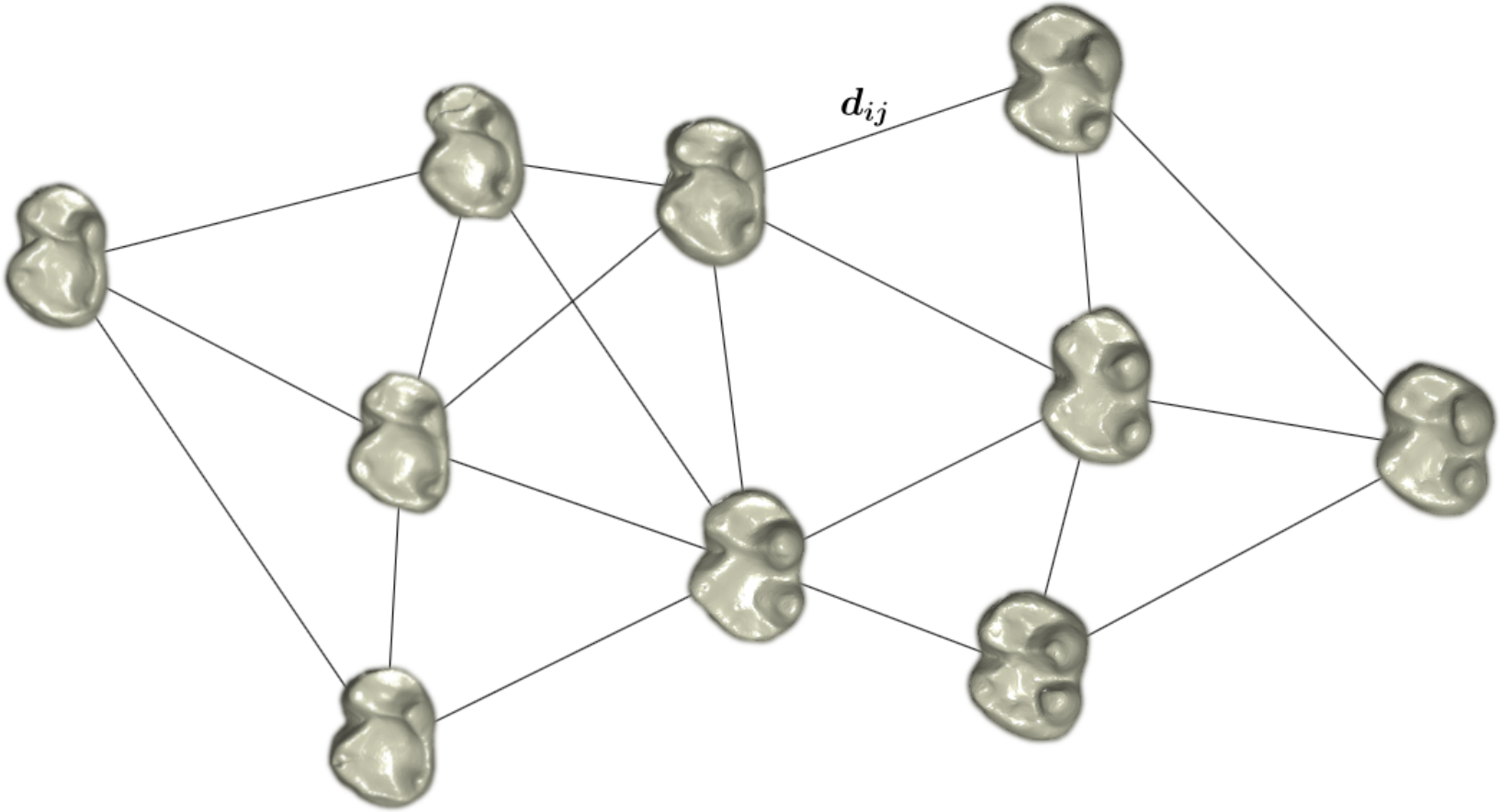


"Correct" like a biologist, but *automatically*?

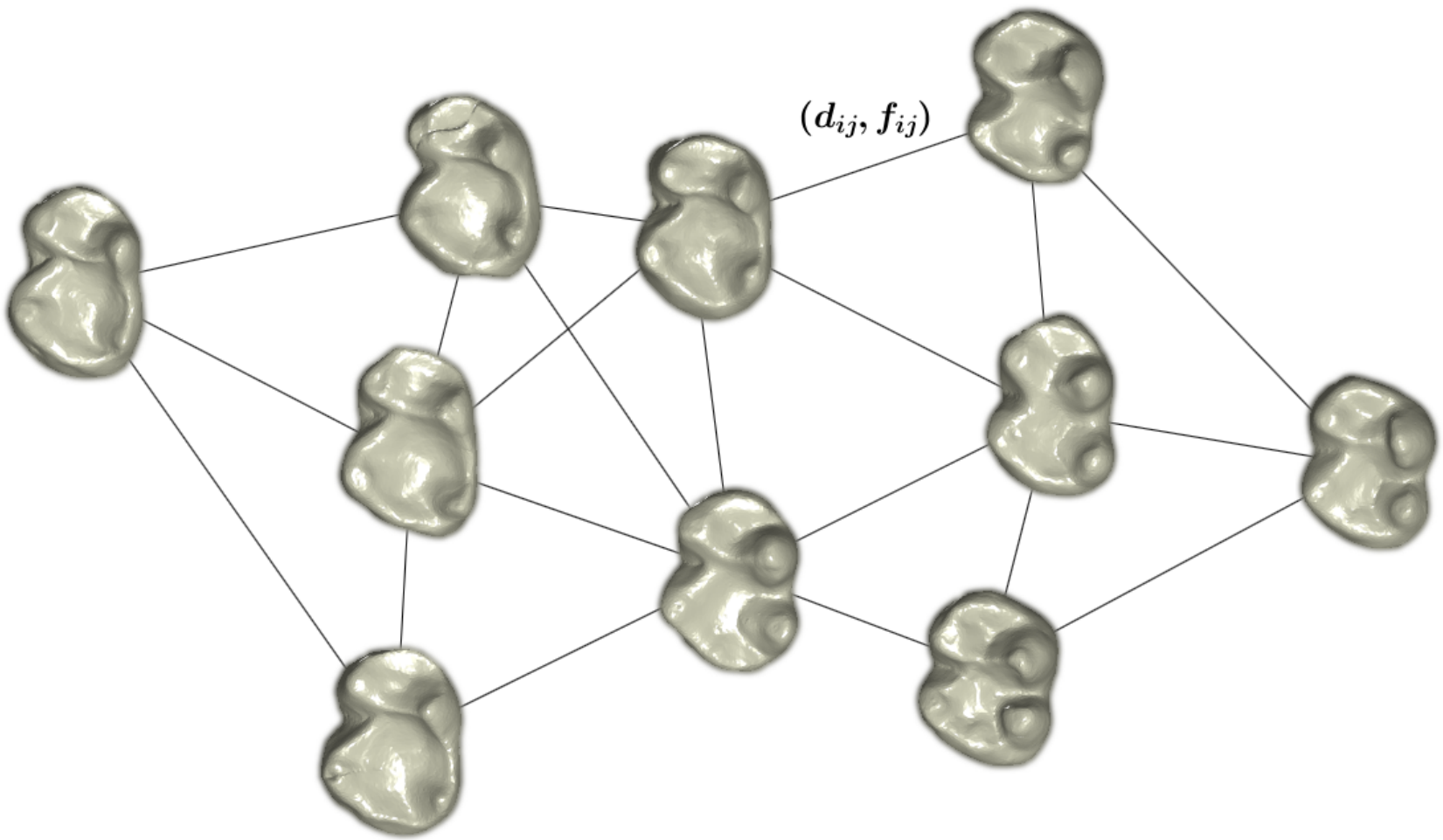
small distances between $S_1, S_2 \longrightarrow$ OK maps
larger distances $\longrightarrow$ not OK

Gao et al. (2018) "Development and Assessment of Fully Automated and Globally Transitive Geometric Morphometric Methods, with Application to a Biological Comparative Dataset with High Interspecific Variation," *The Anatomical Record* 301 (4), 636-658 (2017)

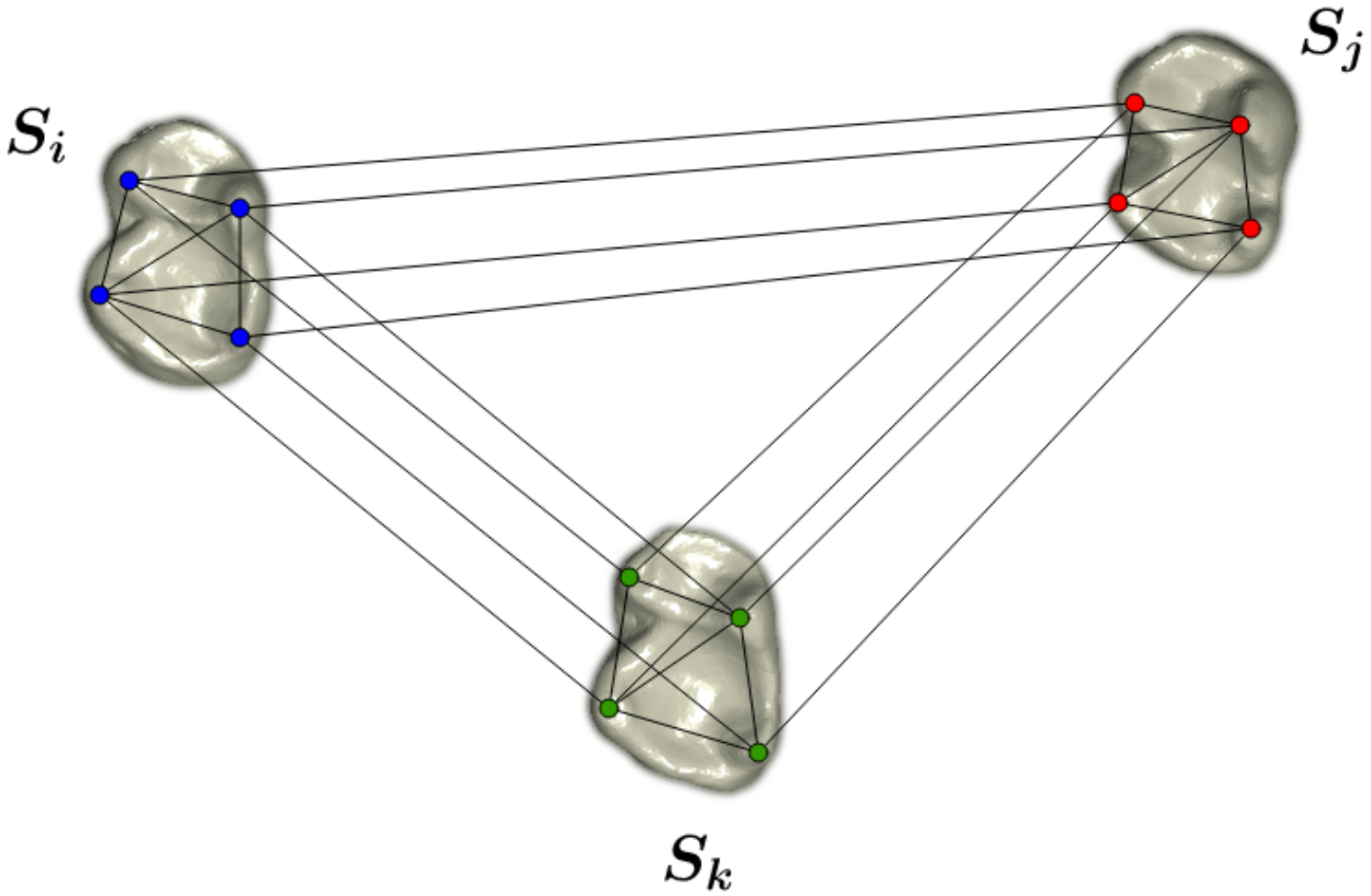
Trust Only *Small* Distances: *Geodesics* in Shape Space



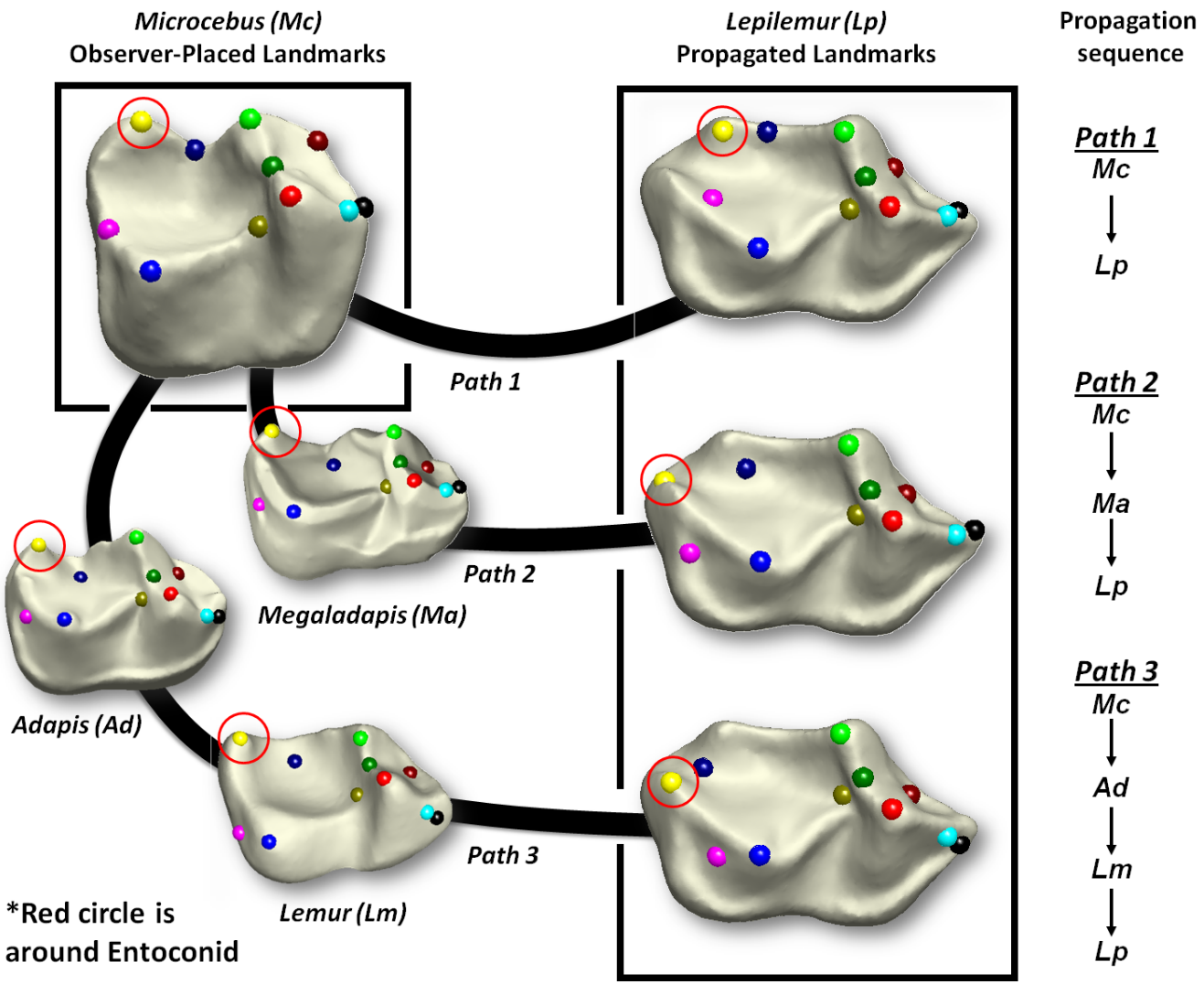
Geometric Model — *Parallel Transport* on *Fibre Bundles*



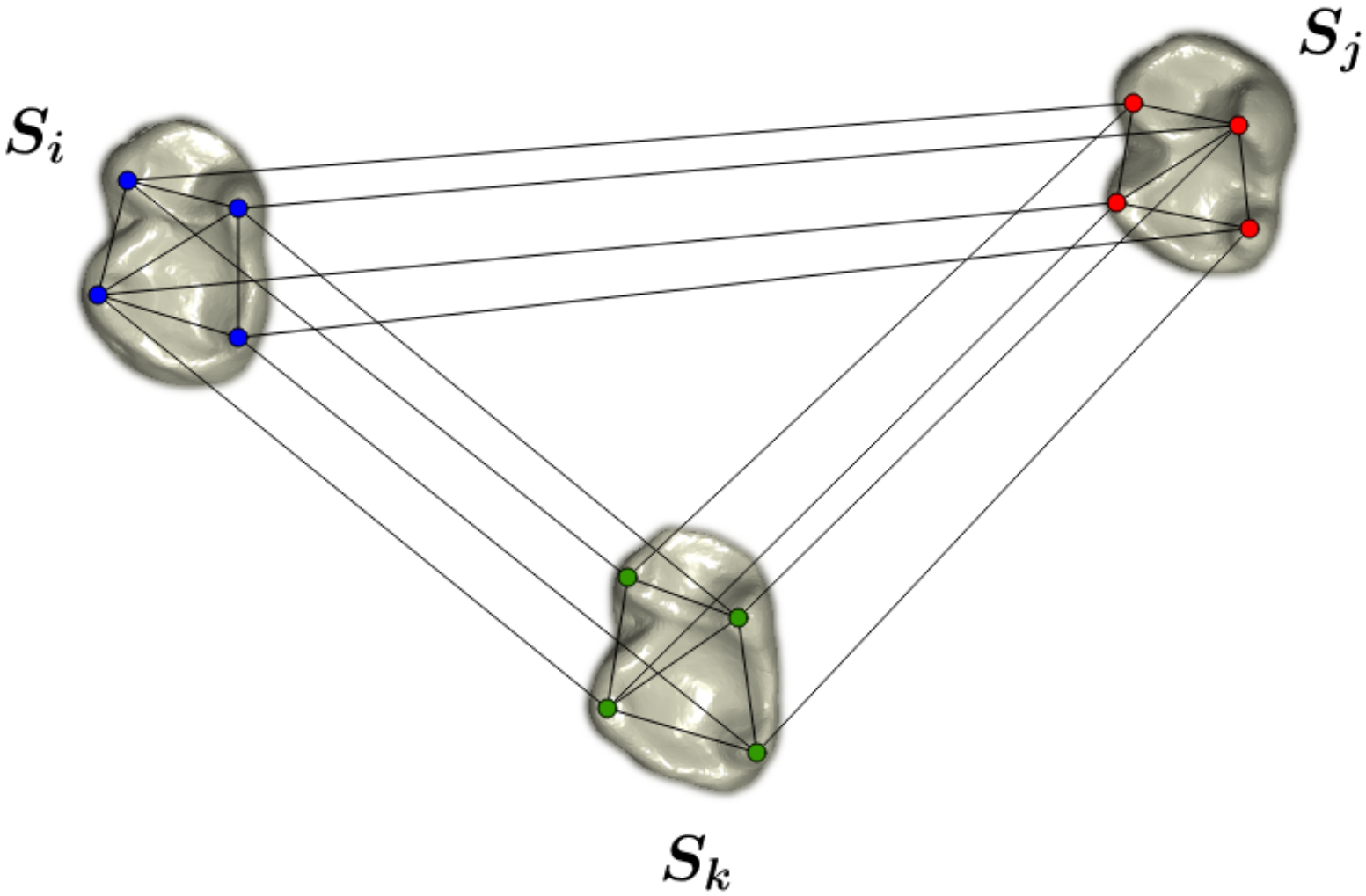
Geometric Model — *Parallel Transport* on *Fibre Bundles*



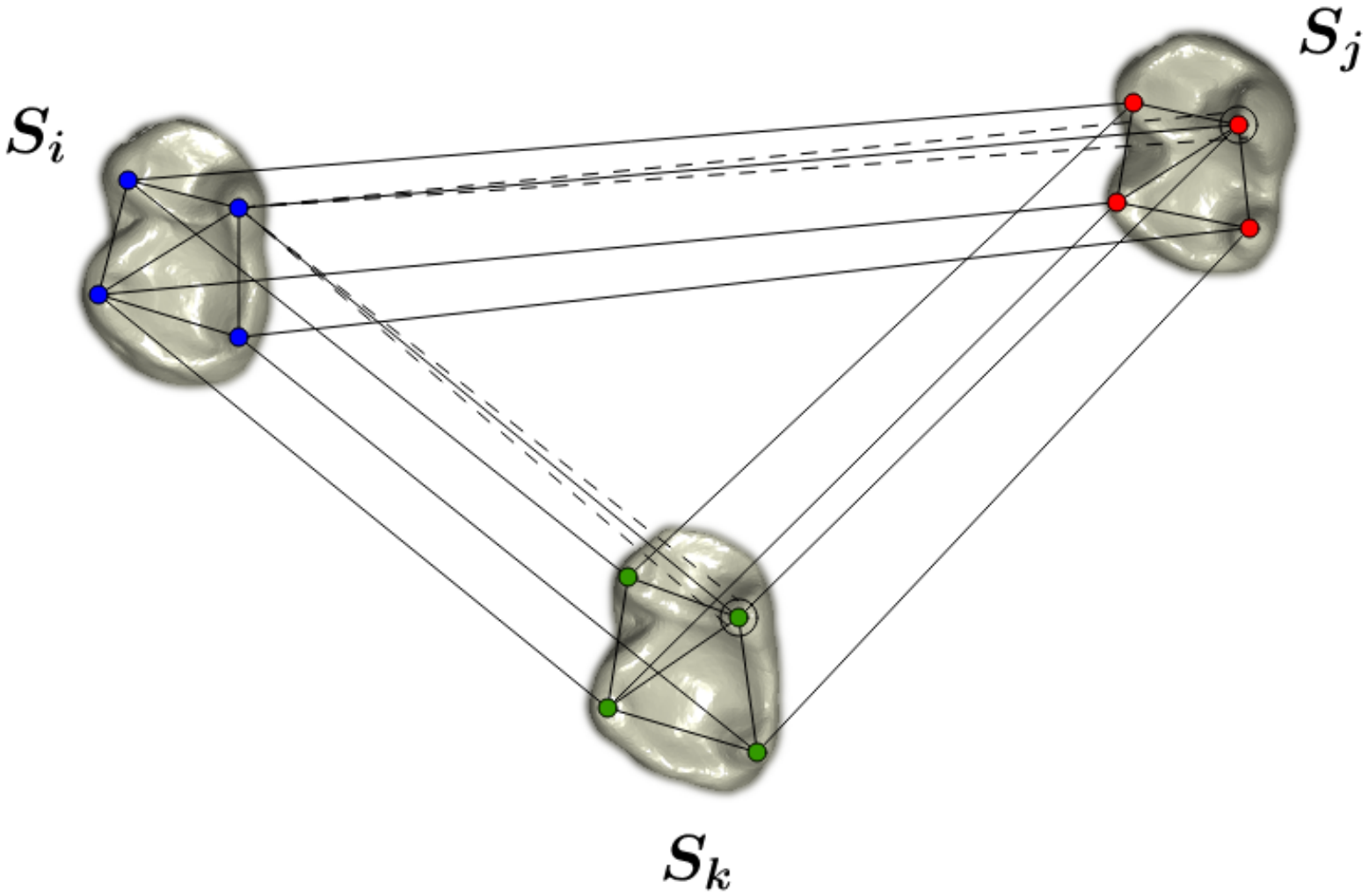
Shape Space is NOT a Trivial Fibre Bundle



Ideal, Noise-Free Case: Trivial Bundle/No Holonomy



Cruel, Real Life: Non-trivial Bundle/Holonomy

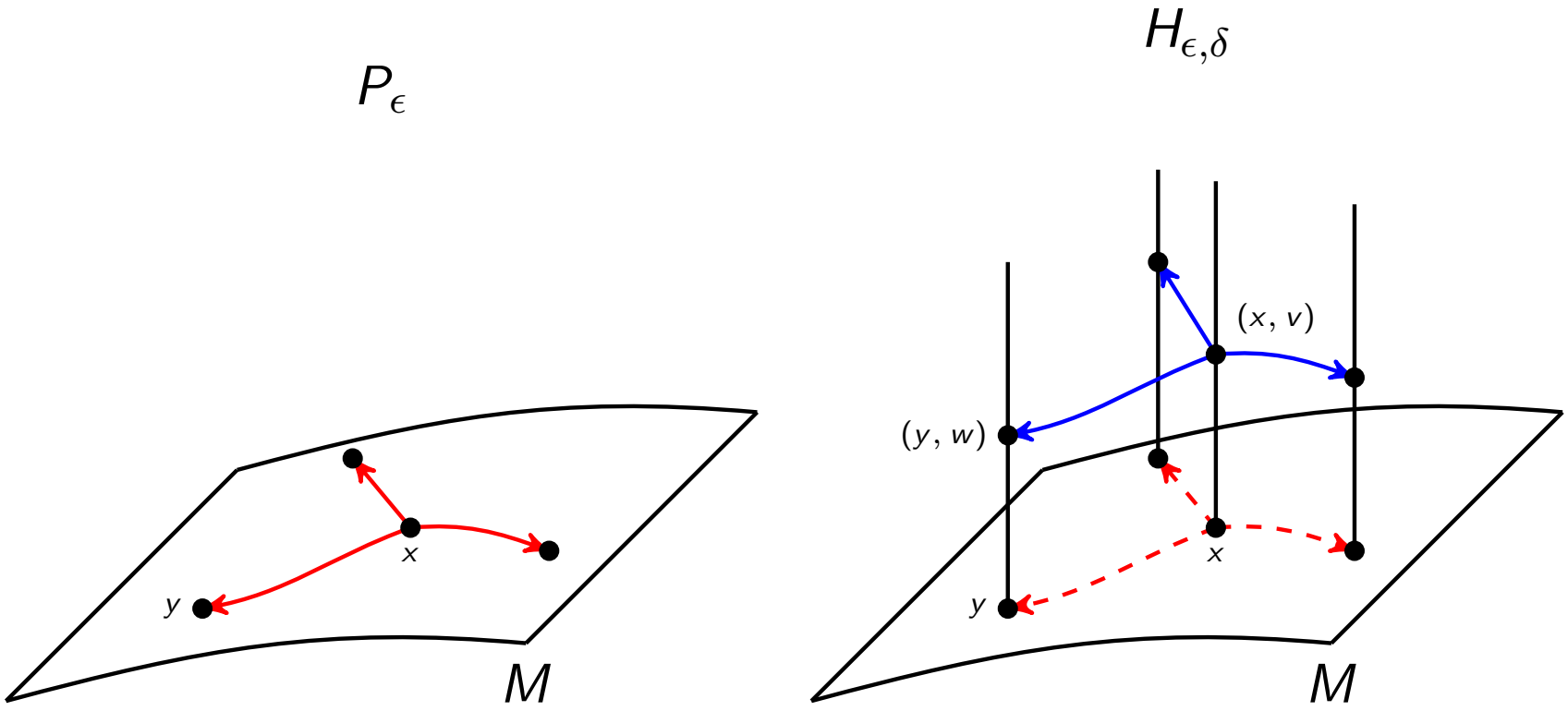


Two Possible Sources of the Holonomy

- Random Noise/Discretization Error: **Variance (G.,2015; G., 2016)**
- The fibre bundle is truly non-trivial: **Bias (G. et al., 2016; Bajaj et al., 2018)**

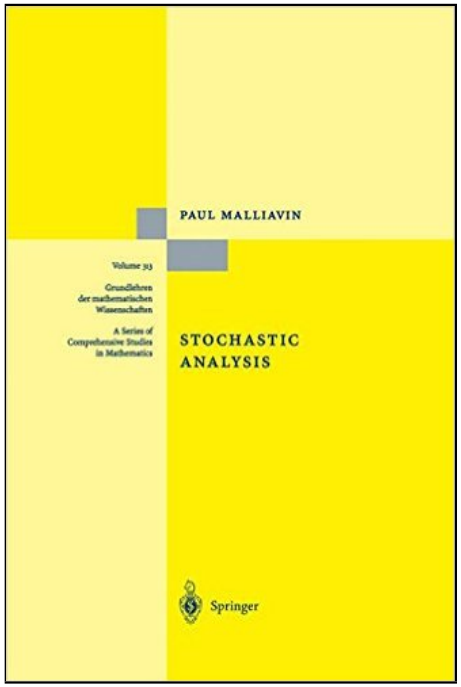
• **Tingran Gao**, "Hypoelliptic Diffusion Maps and Their Applications in Automated Geometric Morphometrics," PhD Thesis, Duke University (2015)
• **Tingran Gao**. "The Diffusion Geometry of Fibre Bundles: Horizontal Diffusion Maps." arXiv:1602.02330 (2016)
• **Tingran Gao**, Jacek Brodzki, Sayan Mukherjee. "The Geometry of Synchronization Problems and Learning Group Actions." *Discrete & Computational Geometry*, to appear (2019)
• Chandrajit Bajaj, **Tingran Gao**, Zihang He, Qixing Huang, and Zhenxiao Liang, "SMAC: Simultaneous Mapping and Clustering Using Spectral Decompositions," *Proceedings of the 35th International Conference on Machine Learning, PMLR* 80:324-333 (2018)

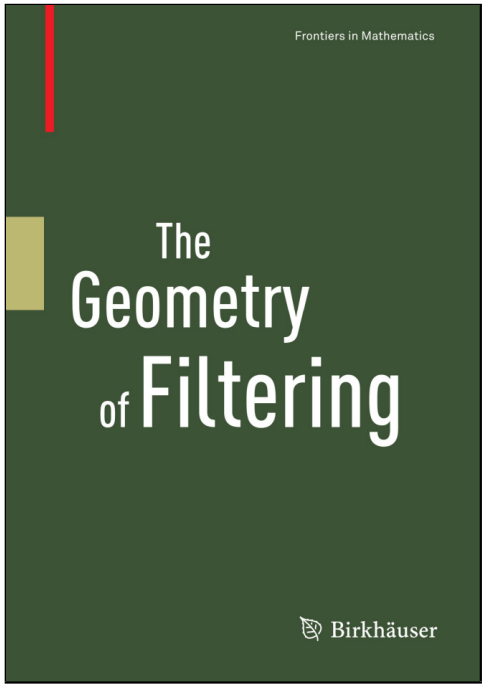
Horizontal Random Walk on a *Fibre Bundle*

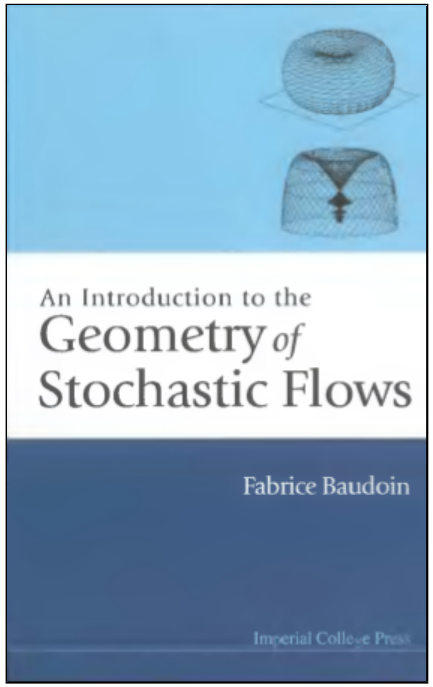


Horizontal Diffusion Process in Stochastic Geometry

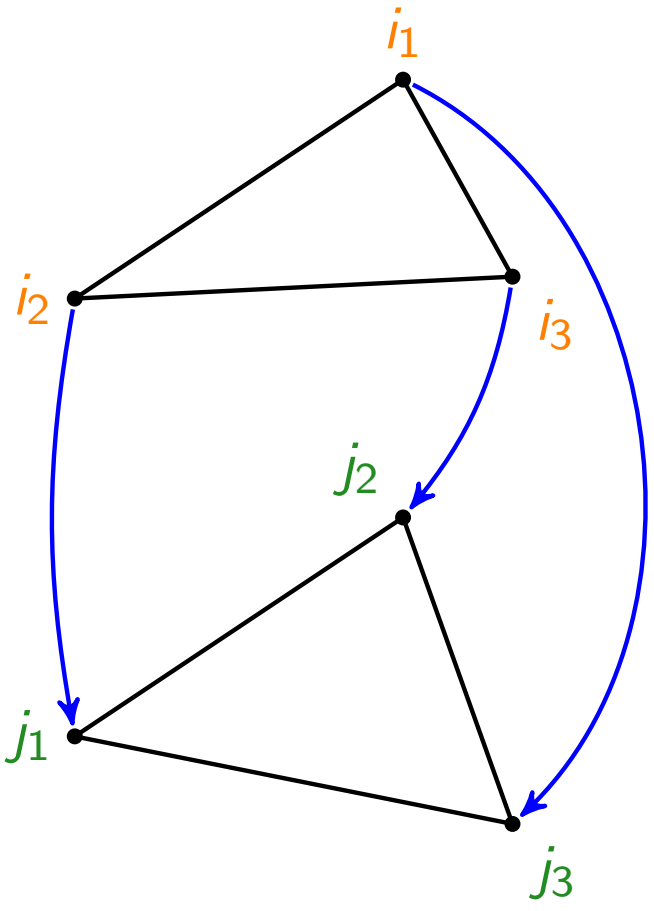
- K.D. Elworthy, W.S. Kendall. "Factorization of Harmonic Maps and Brownian Motions." University of Warwick, 1985.
- M. Liao, "Factorization of Diffusions on Fibre Bundles." *Transactions of the American Mathematical Society.* 311.2 (1989): 813-827.
- M. Arnaudon, A. Thalmaier. "Horizontal Martingales in Vector Bundles." *Séminaire de Probabilits de Strasbourg.* 36 (2002): 419-456.
- K.D. Elworthy, Y. Le Jan, and X. Li. "The Geometry of Filtering." Springer Basel, 2010. 33-59.
- F. Baudoin. "An Introduction to the Geometry of Stochastic Flows." London: Imperial College Press, 2004.



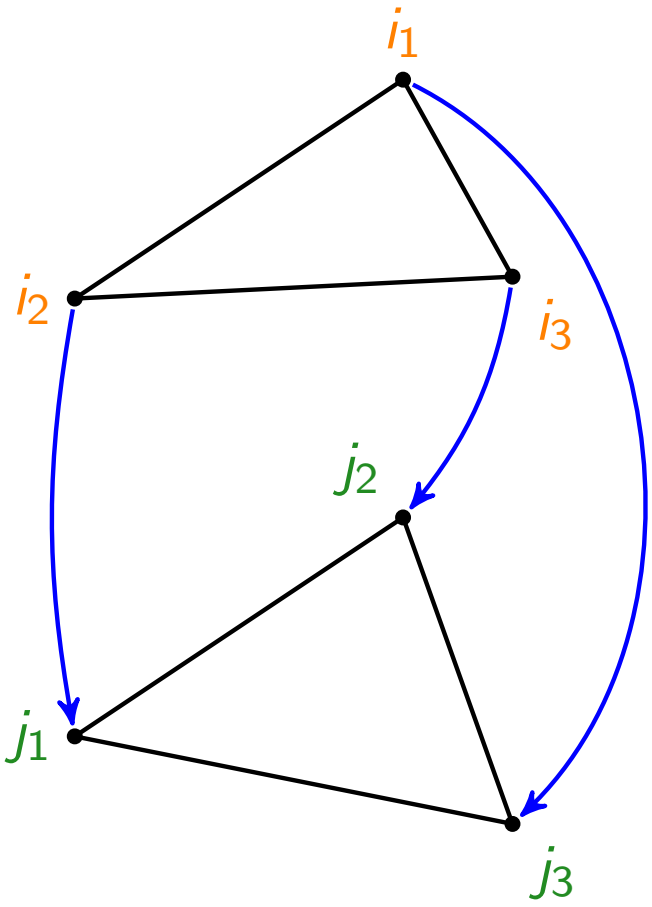




Horizontal Diffusion Maps – Matrix Form

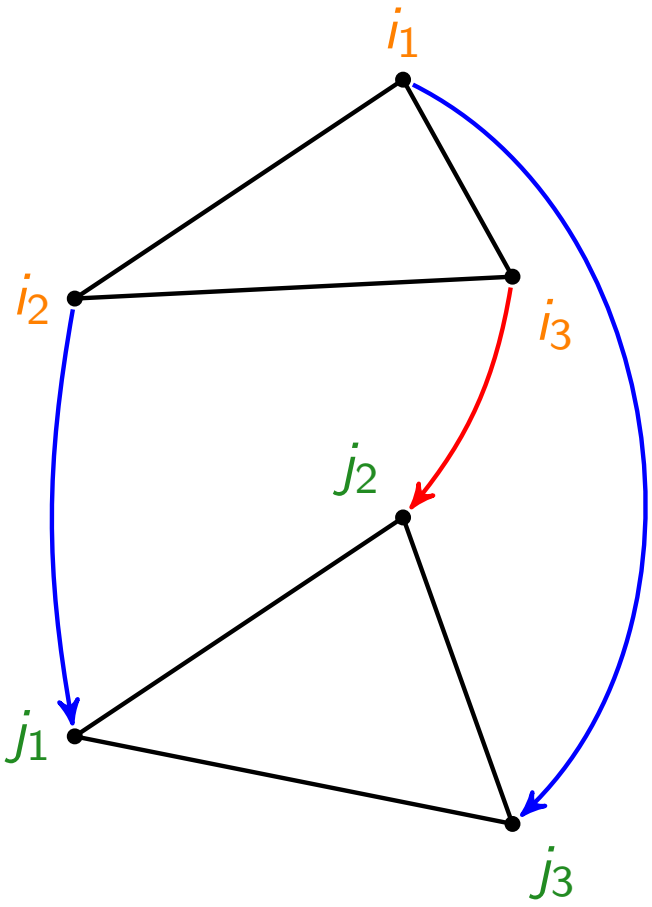


Horizontal Diffusion Maps – Matrix Form



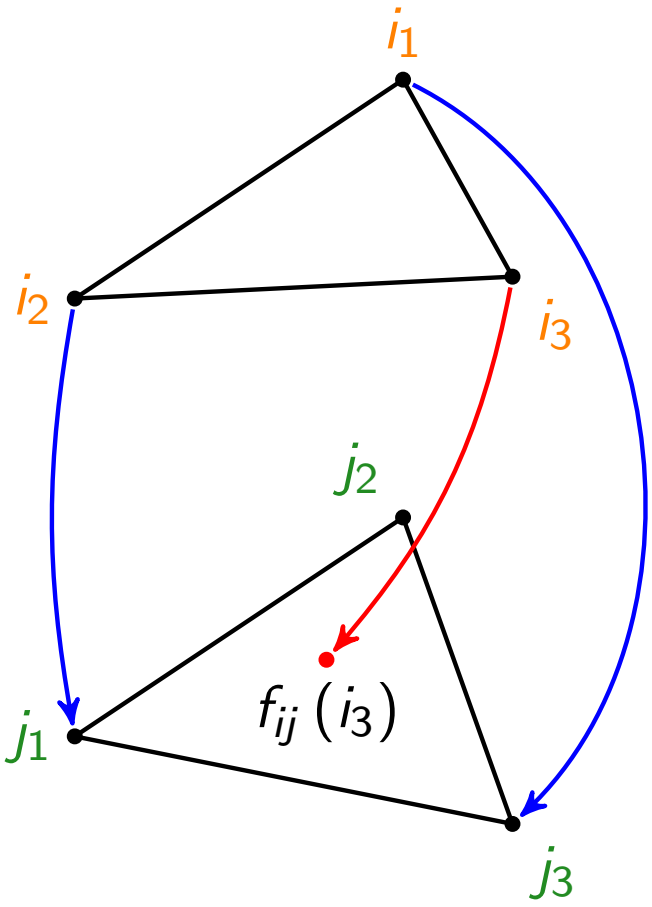
$$
\begin{array}{cc}
 & \xrightarrow{\quad S_j \quad} \\
 & \begin{array}{ccccc} \cdots & j_1 & j_2 & j_3 & \cdots \end{array} \\
S_i \downarrow \begin{array}{c} \vdots \\ i_1 \\ i_2 \\ i_3 \\ \vdots \end{array} & \begin{pmatrix} & \vdots & \vdots & \vdots & \\ \cdots & 0 & 0 & 1 & \cdots \\ \cdots & 1 & 0 & 0 & \cdots \\ \cdots & 0 & 1 & 0 & \cdots \\ & \vdots & \vdots & \vdots & \end{pmatrix}
\end{array}
$$

Horizontal Diffusion Maps – Matrix Form



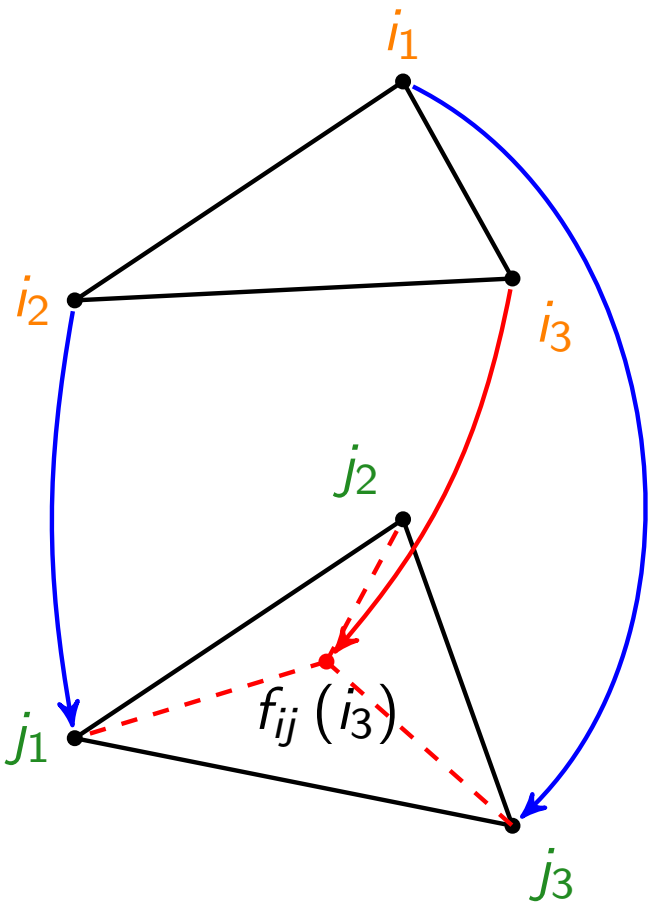
$$
\begin{array}{cc}
 & \xrightarrow{\quad S_j \quad} \\
 & \begin{array}{ccccc} \cdots & j_1 & j_2 & j_3 & \cdots \end{array} \\
S_i \downarrow \begin{array}{c} \vdots \\ i_1 \\ i_2 \\ i_3 \\ \vdots \end{array} & \begin{pmatrix} & \vdots & \vdots & \vdots & \\ \cdots & 0 & 0 & 1 & \cdots \\ \cdots & 1 & 0 & 0 & \cdots \\ \cdots & 0 & \boxed{1} & 0 & \cdots \\ & \vdots & \vdots & \vdots & \end{pmatrix}
\end{array}
$$

Horizontal Diffusion Maps – Matrix Form



$$
\begin{array}{cc}
 & \xrightarrow{\quad s_j \quad} \\
 & \begin{array}{ccccc} \cdots & j_1 & j_2 & j_3 & \cdots \end{array} \\
s_i \downarrow \begin{array}{c} \vdots \\ i_1 \\ i_2 \\ i_3 \\ \vdots \end{array} &
\begin{pmatrix}
 & \vdots & \vdots & \vdots & \\
\cdots & 0 & 0 & 1 & \cdots \\
\cdots & 1 & 0 & 0 & \cdots \\
\cdots & 0 & \boxed{?} & 0 & \cdots \\
 & \vdots & \vdots & \vdots &
\end{pmatrix}
\end{array}
$$

Horizontal Diffusion Maps – Matrix Form



$$
\begin{array}{cc}
 & \xrightarrow{\quad s_j \quad} \\
 & \begin{array}{ccccc} \cdots & j_1 & j_2 & j_3 & \cdots \end{array} \\
s_i \downarrow \begin{array}{c} \vdots \\ i_1 \\ i_2 \\ i_3 \\ \vdots \end{array} &
\begin{pmatrix}
 & \vdots & \vdots & \vdots & \\
\cdots & 0 & 0 & 1 & \cdots \\
\cdots & 1 & 0 & 0 & \cdots \\
\cdots & \boxed{0.91 \quad 0.95 \quad 0.88} & & & \cdots \\
 & \vdots & \vdots & \vdots &
\end{pmatrix}
\end{array}
$$

$$
\rho_{ij}^{\delta}(r,s) = \exp\left(-\frac{\|f_{ij}(i_r) - j_s\|^2}{\delta}\right)
$$

Horizontal Diffusion Maps – Matrix Form

Assume there are n shapes, $S_1, \ldots, S_n$, and each S_i has ℓ_i vertices.

- Compute $f_{ij} : S_i \to S_j$ for each pair of $1 \leq i \neq j \leq n$
- $f_{ij} \longleftrightarrow \rho_{ij}^{\delta} \in \mathbb{R}^{\ell_j \times \ell_i}$
- Let $\mathcal{W}$ be an n-by-n **block** matrix, with (i, j)th block

$$\exp\left(-\frac{d_{\mathrm{cP}}^2(S_i, S_j)}{\epsilon}\right) \cdot \rho_{ij}^{\delta} \in \mathbb{R}^{\ell_j \times \ell_i}$$

  ϵ: horizontal bandwidth parameter
  δ: vertical bandwidth parameter

- Let $\mathcal{D}$ be a diagonal matrix, with kth diagonal entry equaling the kth row sum of $\mathcal{W}$
- **Horizontal Random Walk Laplacian**: $I - \mathcal{D}^{-1}\mathcal{W}$

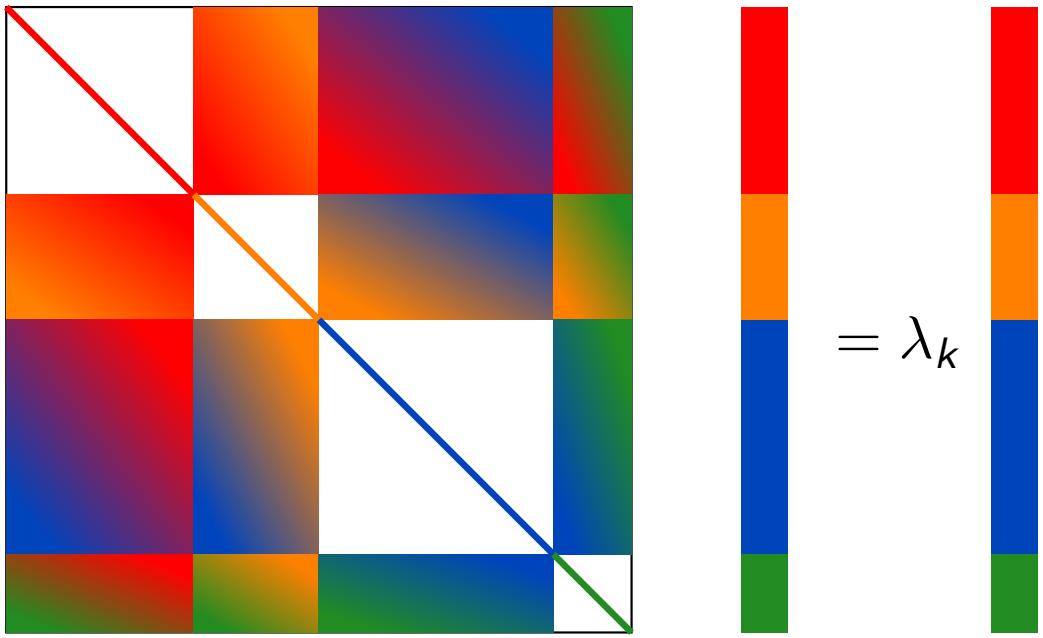
Remark. $\mathcal{W}$ can be viewed as a flattening of a 4-tensor

Horizontal Diffusion Maps – Matrix Form

For all $1 \leq k \leq \kappa$, solve eigenproblems:

$$\mathcal{D}^{-1}\mathcal{W} \qquad u_k = \lambda_k \; u_k,$$

$\mathcal{D}^{-1}$



$= \lambda_k$

HDM: Continuous Limit

M – base manifold; F – template fibre; E – total manifold; F_x – fibre over $x \in M$; $P_{yx} : F_x \to F_y$ – parallel transport

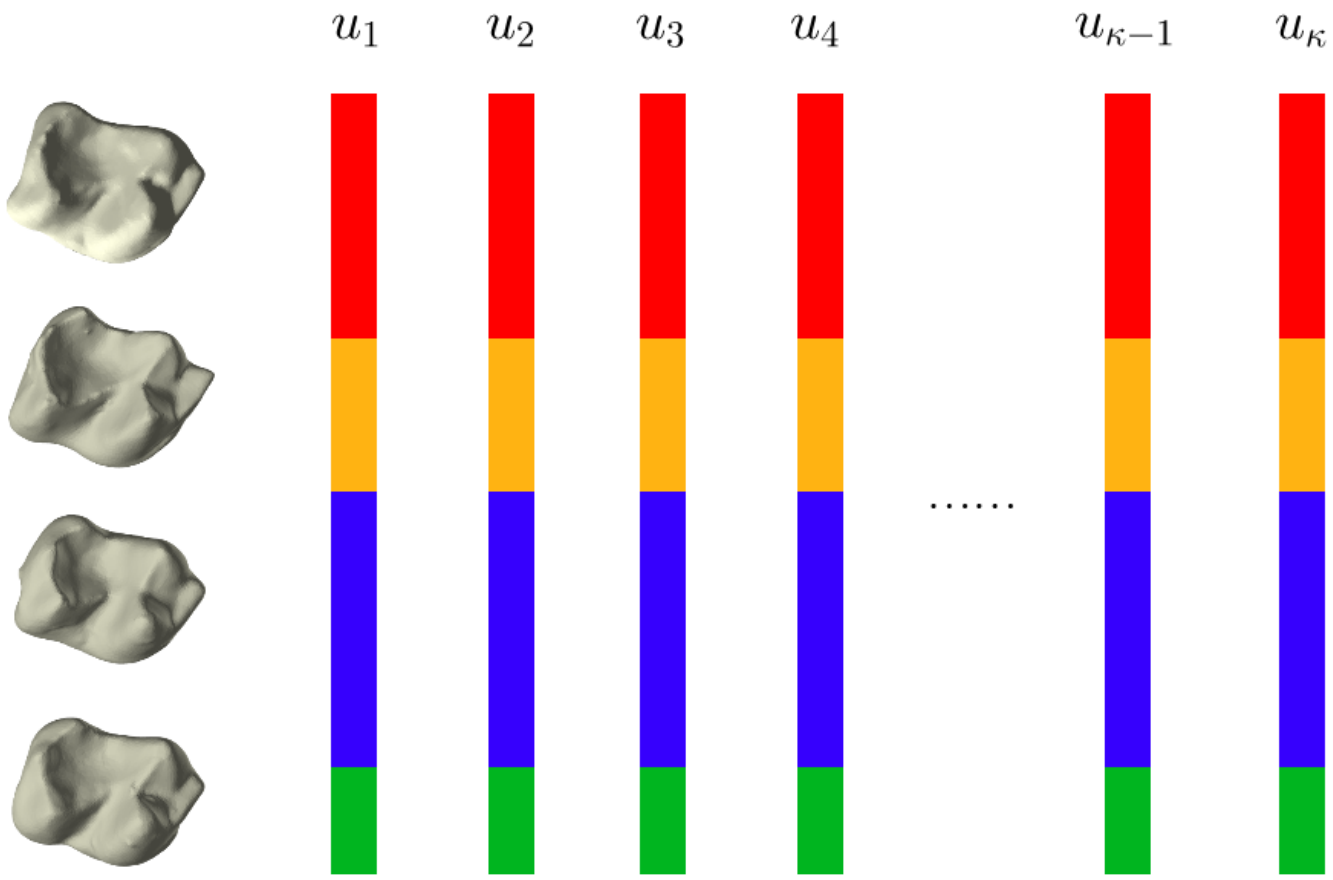
- As $n, \ell_i \to \infty$ for all $1 \leq \ell \leq n$, $\mathcal{D}^{-1}\mathcal{W}$ converges weakly to the integral operator $H_{\epsilon,\delta} : L^1(E) \to L^1(E)$

$$H_{\epsilon,\delta} f(x,v) = \frac{\int_M \int_{F_y} K_{\epsilon,\delta}(x,v;y,w) f(y,w) p(y,w) \, d\mathrm{vol}_{F_y}(w) \, d\mathrm{vol}_M(y)}{\int_M \int_{F_y} K_{\epsilon,\delta}(x,v;y,w) p(y,w) \, d\mathrm{vol}_{F_y}(w) \, d\mathrm{vol}_M(y)}$$

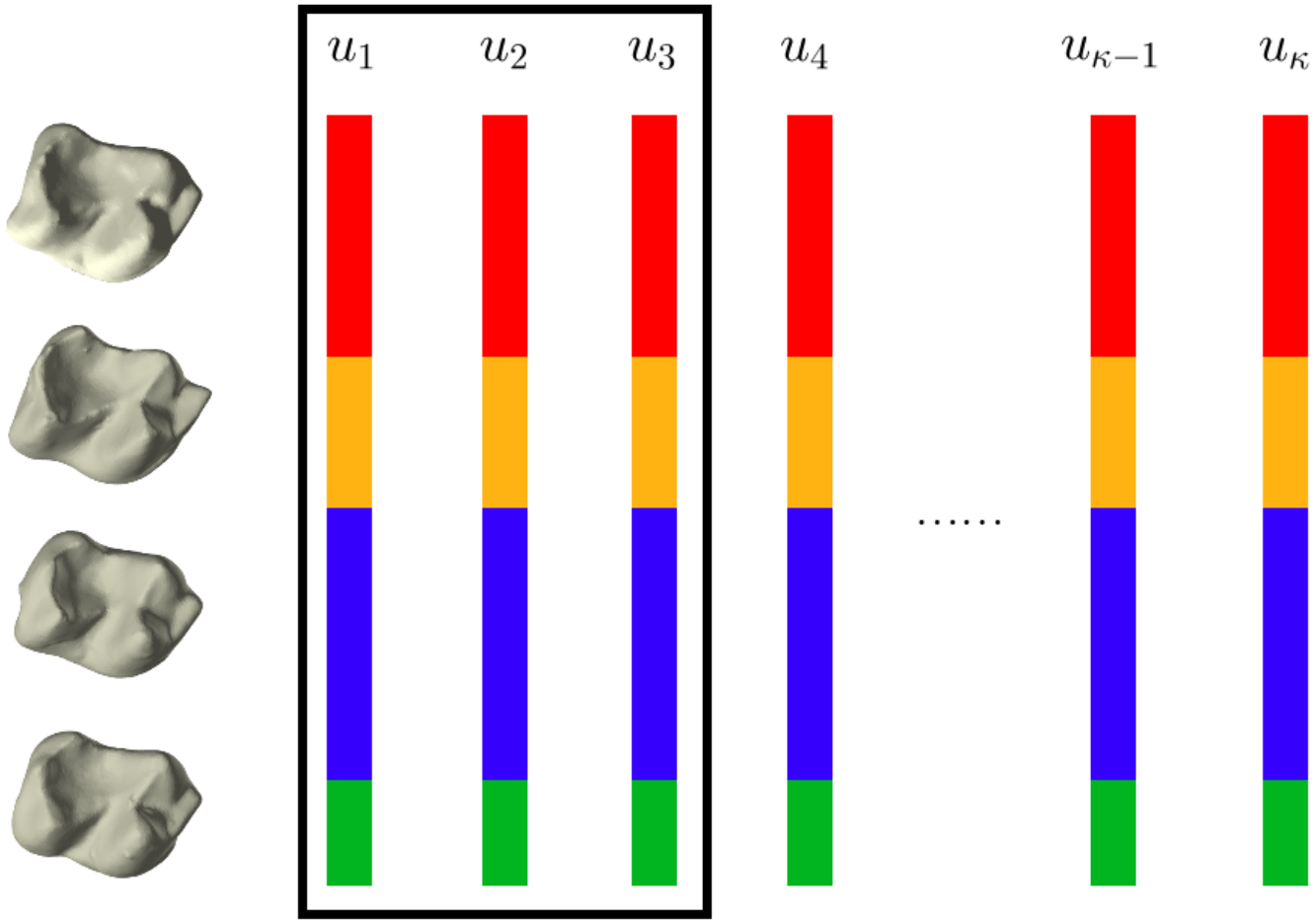
for all $f \in L^1(E)$ and $v \in F_x$, $w \in F_y$, where

$$K_{\epsilon,\delta}(x,v;y,w) = \exp\left(-\frac{d_M^2(x,y)}{\epsilon} - \frac{d_{F_y}^2(P_{yx}v,w)}{\delta}\right).$$

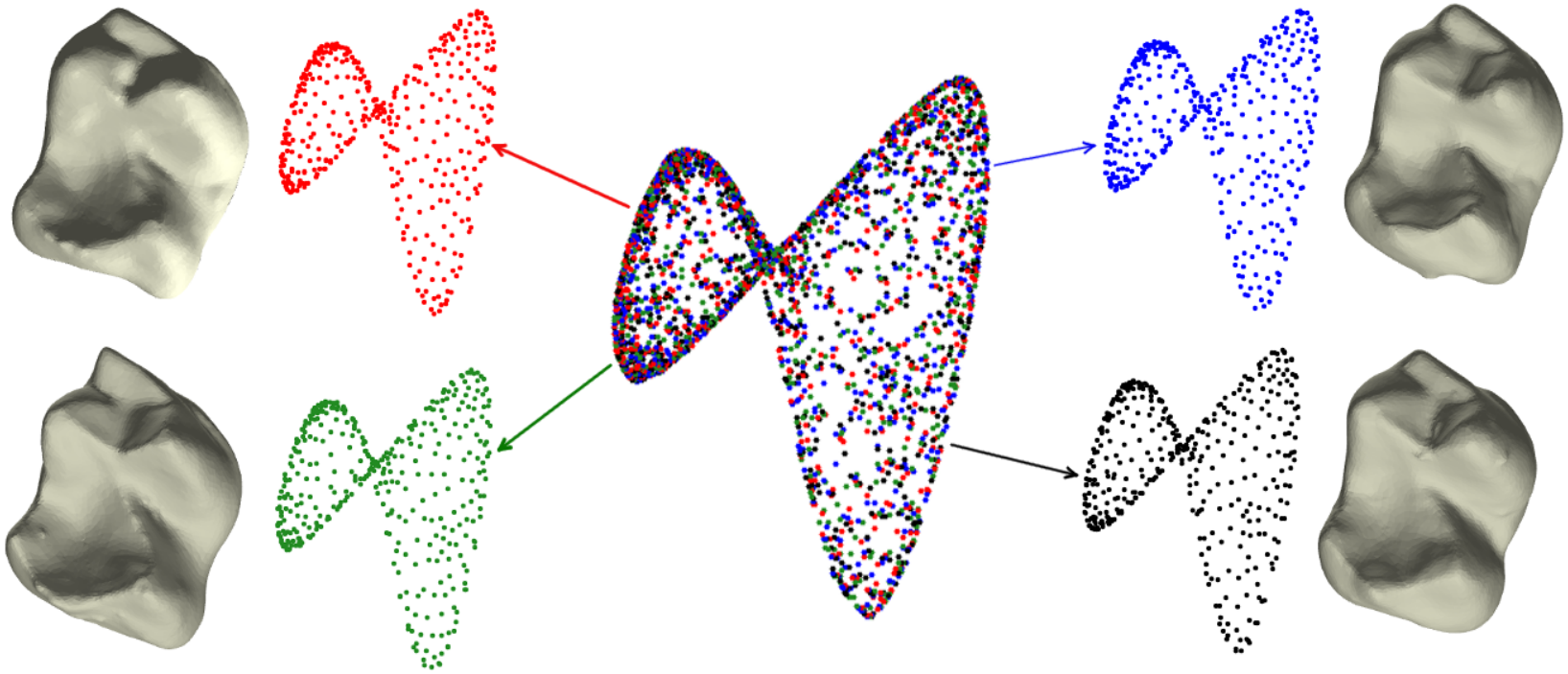
Horizontal Diffusion Maps: Embedding the Entire Bundle



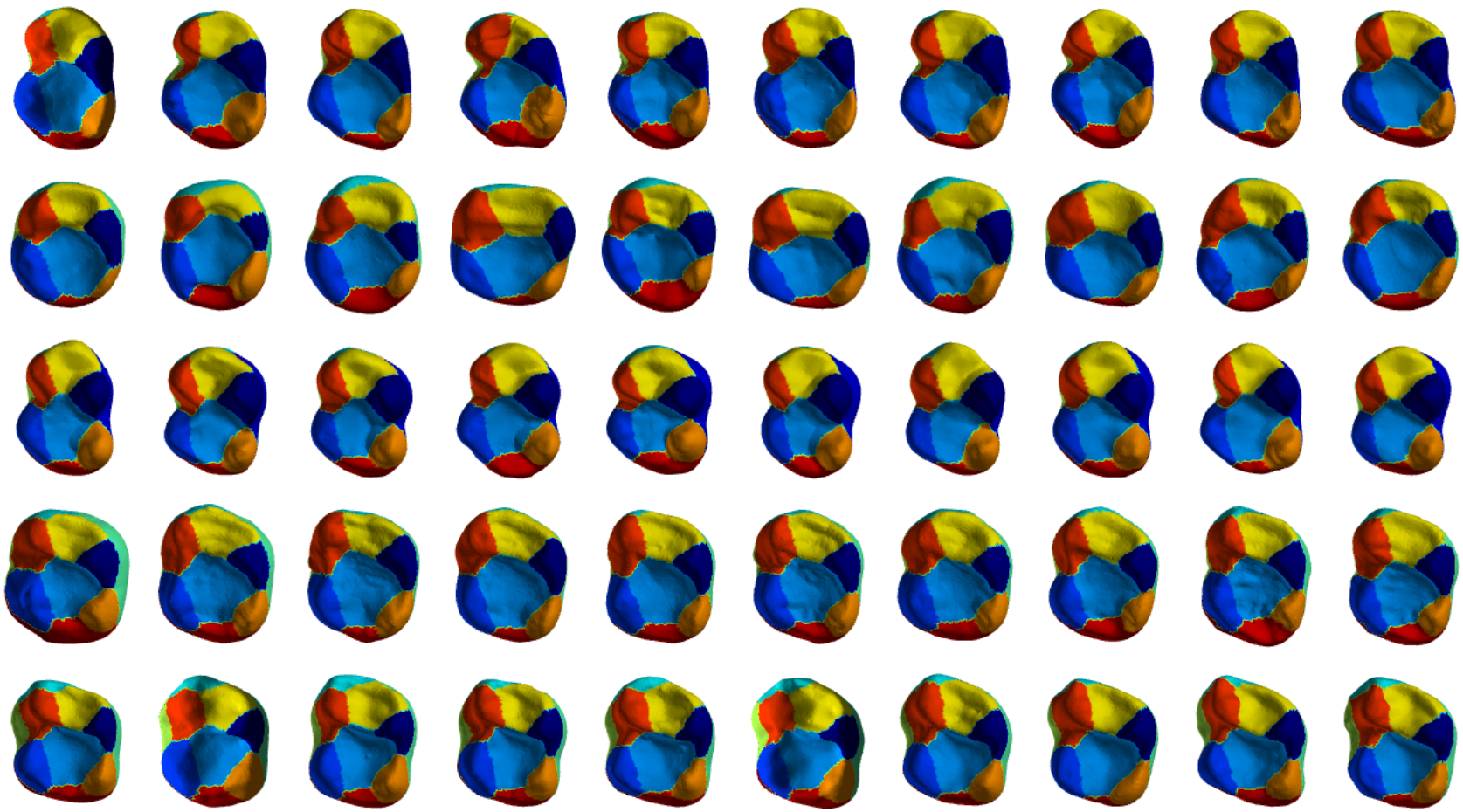
Horizontal Diffusion Maps: Embedding the Entire Bundle



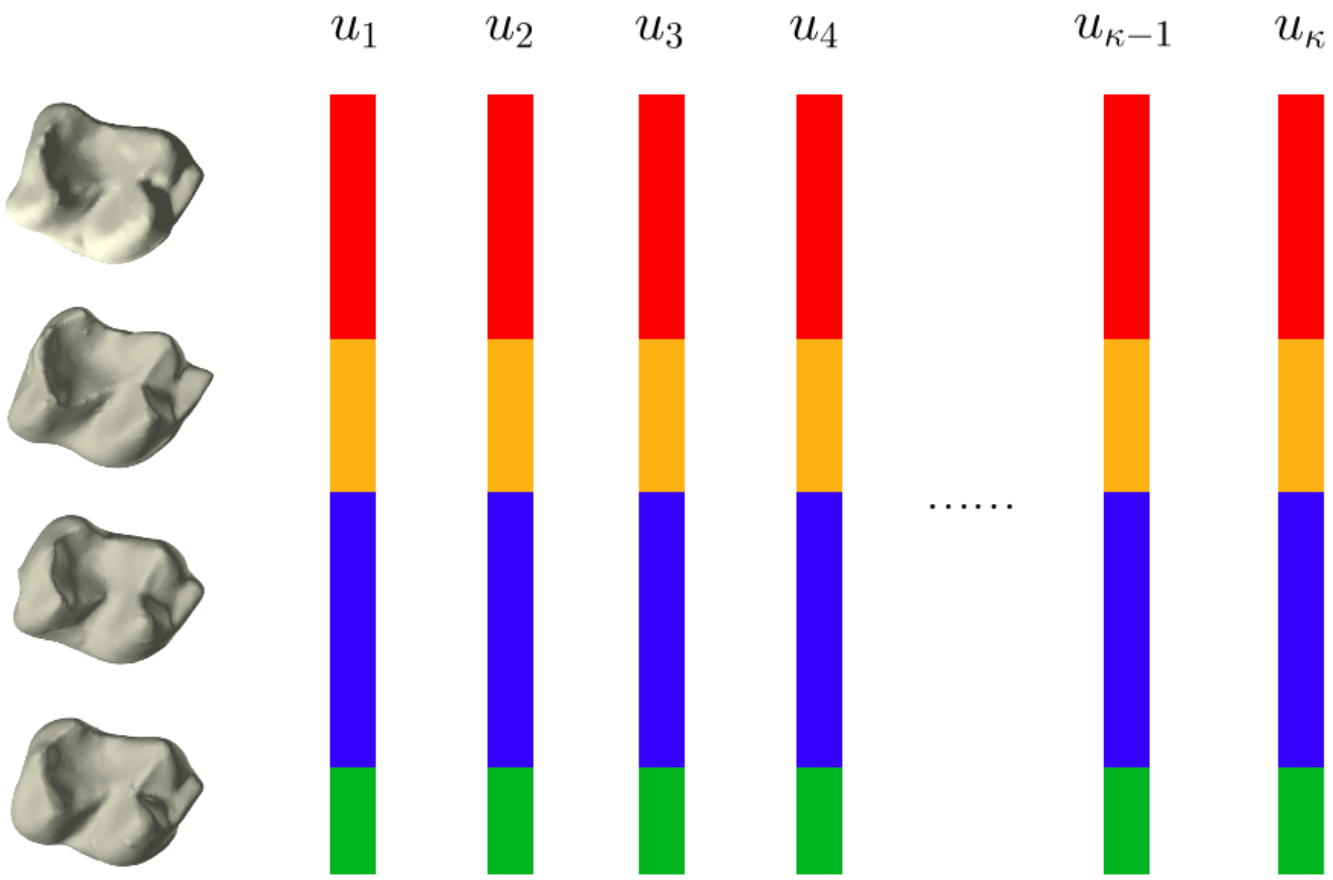
Horizontal Diffusion Maps



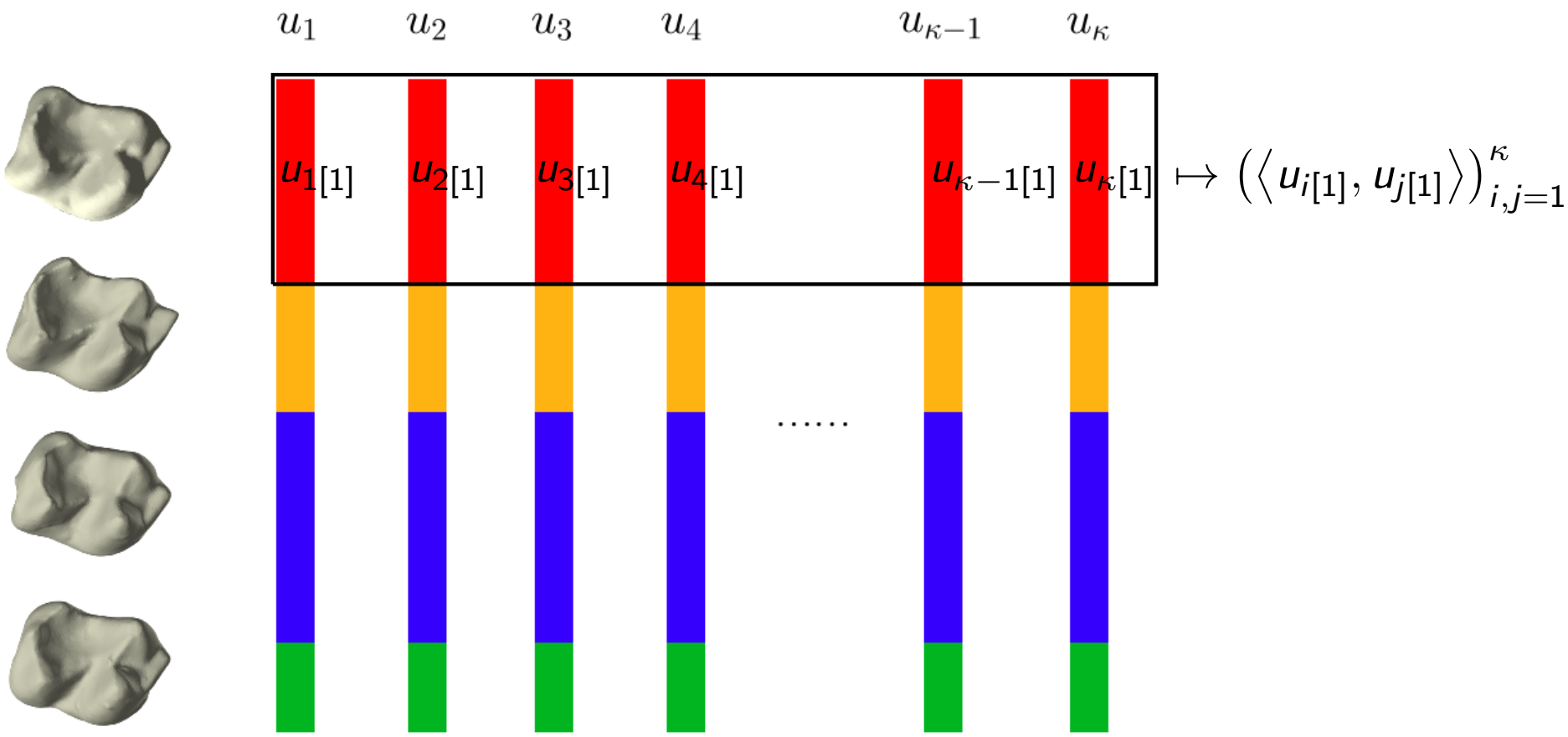
Automatic Landmarking — Interpretability



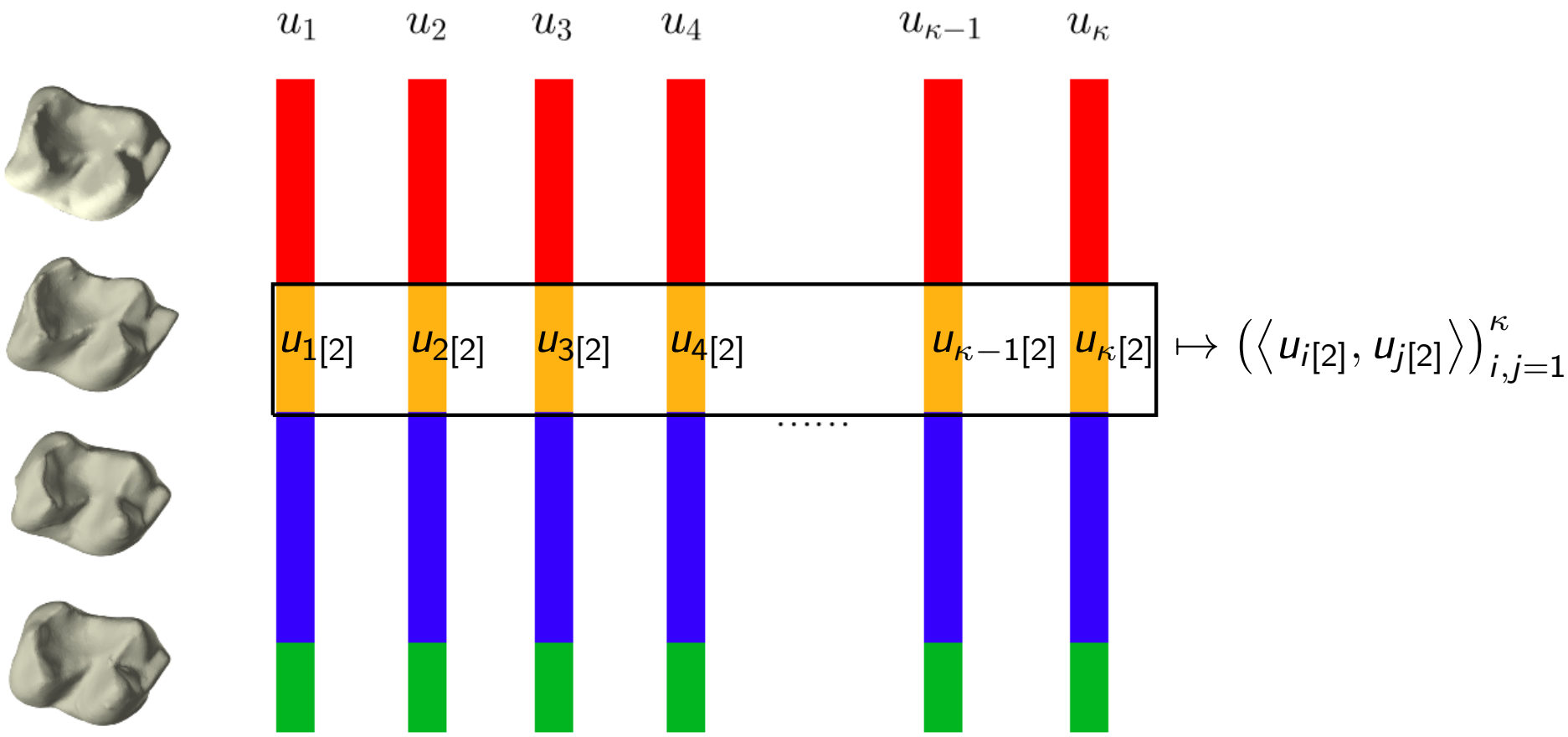
Horizontal Diffusion Maps: Embedding the Base Manifold



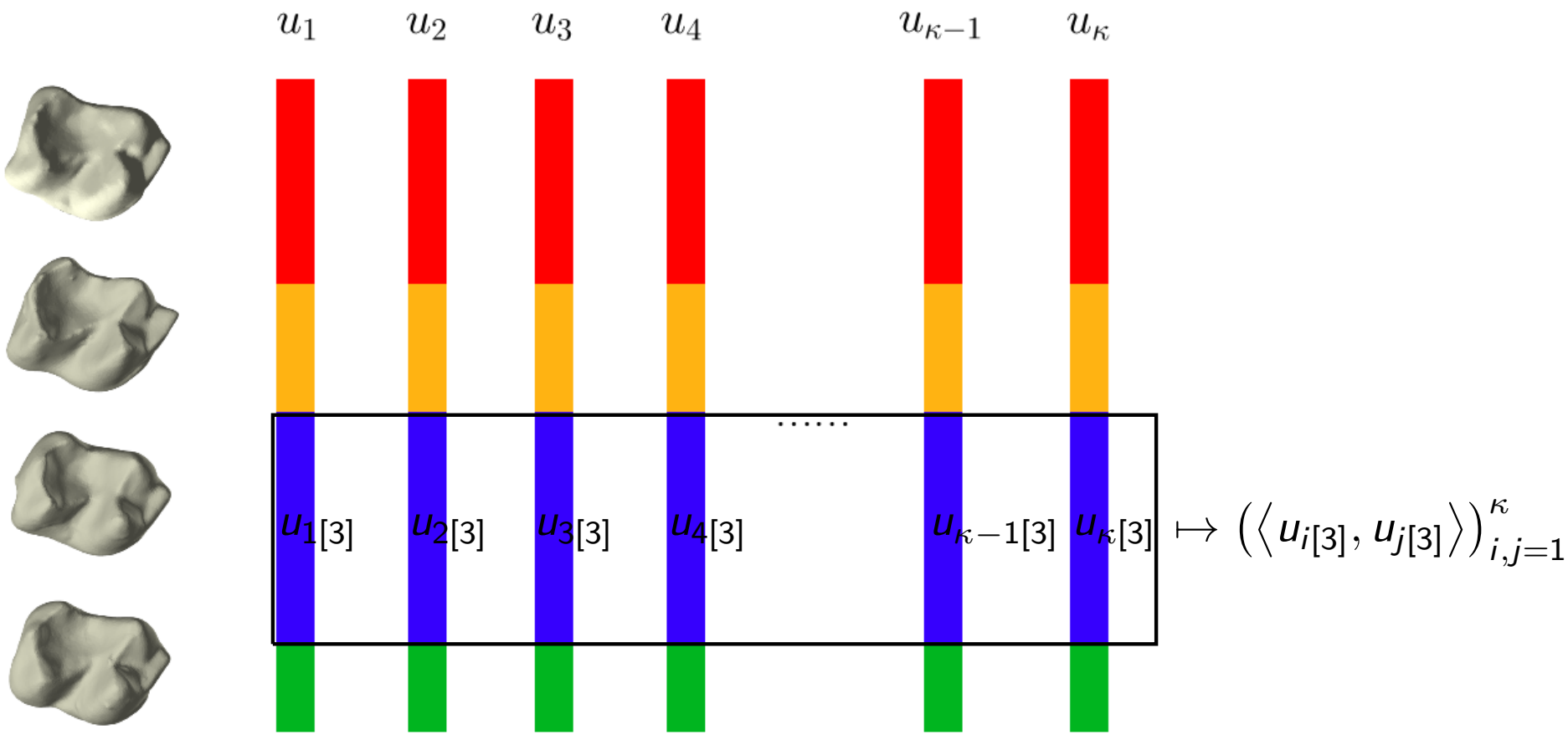
Horizontal Diffusion Maps: Embedding the Base Manifold



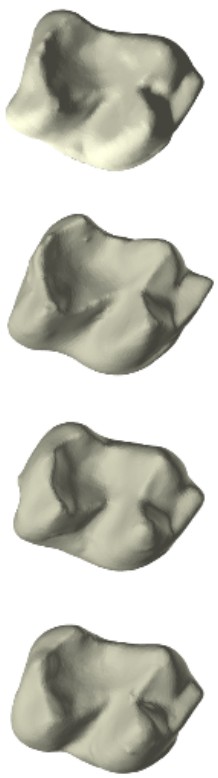
Horizontal Diffusion Maps: Embedding the Base Manifold

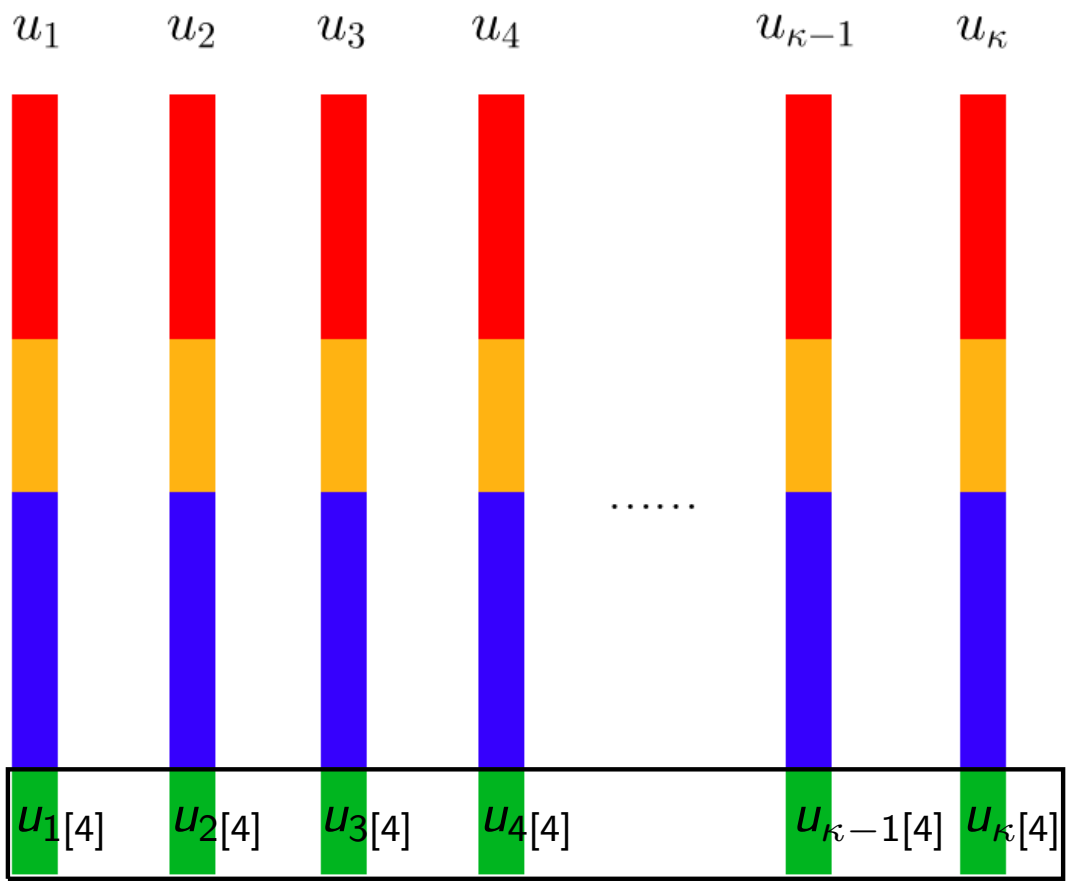


Horizontal Diffusion Maps: Embedding the Base Manifold



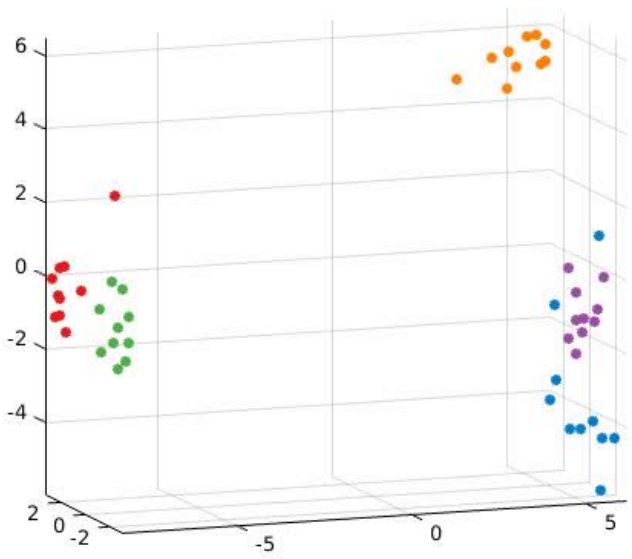
Horizontal Diffusion Maps: Embedding the Base Manifold



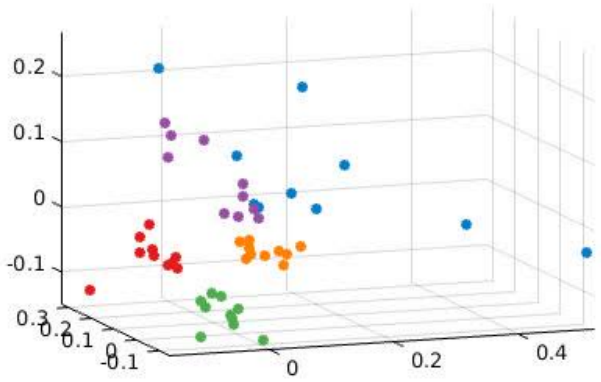


$$\mapsto \left(\left\langle u_{i[4]}, u_{j[4]} \right\rangle\right)_{i,j=1}^{\kappa}$$

Species Clustering

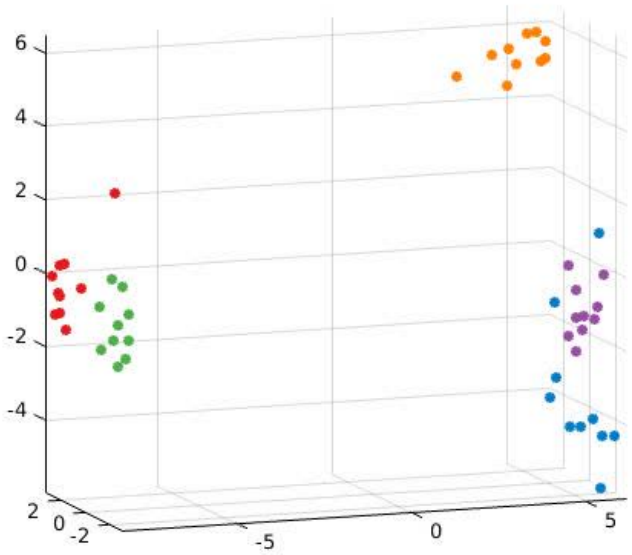


Horizontal Base Diffusion Distance (with Maps)

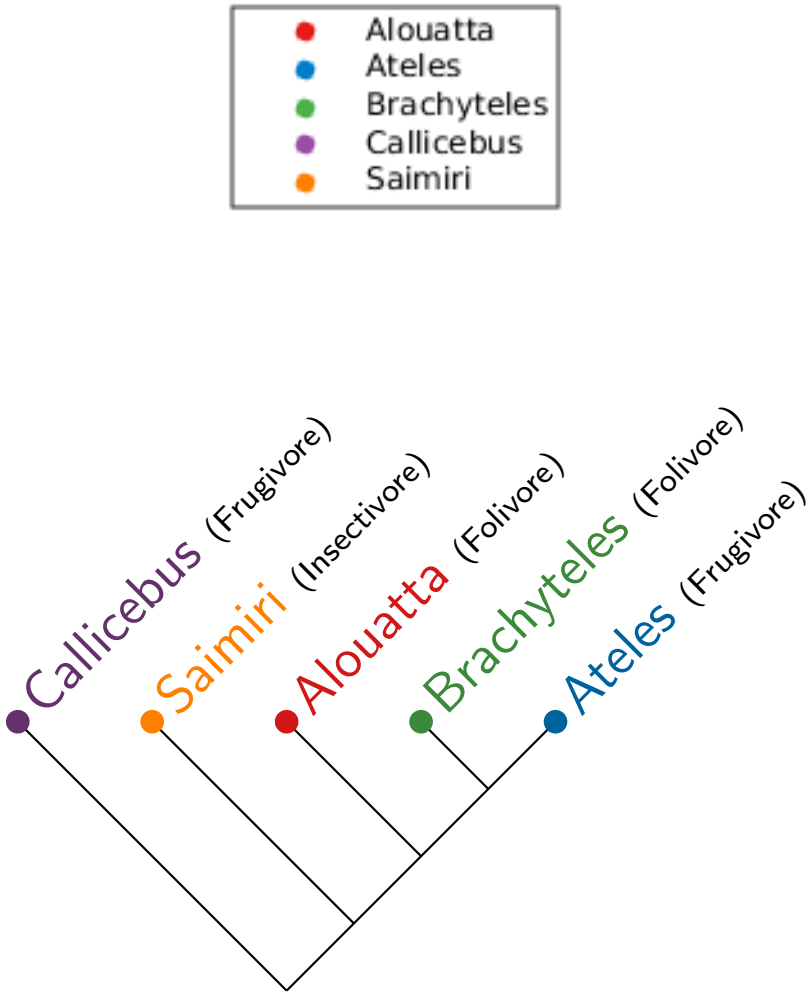


Diffusion Distance (without Maps)

Species Clustering



Horizontal Base Diffusion Distance (with Maps)



HDM: Continuous Limit

$$H_{\epsilon,\delta} f(x,v) = \frac{\int_M \int_{F_y} K_{\epsilon,\delta}(x,v;y,w)\, f(y,w)\, p(y,w)\, d\mathrm{vol}_{F_y}(w)\, d\mathrm{vol}_M(y)}{\int_M \int_{F_y} K_{\epsilon,\delta}(x,v;y,w)\, p(y,w)\, d\mathrm{vol}_{F_y}(w)\, d\mathrm{vol}_M(y)}$$

for all $f \in L^1(E)$ and $v \in F_x$, $w \in F_y$, where

$$K_{\epsilon,\delta}(x,v;y,w) = \exp\left(-\frac{d_M^2(x,y)}{\epsilon} - \frac{d_{F_y}^2(P_{yx}v,w)}{\delta}\right)$$

Asymptotic Theory: Generator of the Diffusion Process

Theorem (G., 2016). If $\delta = O(\epsilon)$ as $\epsilon \to 0$, then for any $f \in C^\infty(E)$ and $(x, v) \in E$, as $\epsilon \to 0$,

$$
\begin{aligned}
&H_{\epsilon,\delta}^{(\alpha)} f(x, v) \\
&= f(x, v) + \epsilon \frac{m_{21}}{2m_0} \left[\frac{\Delta_H \left(f p^{1-\alpha} \right)(x, v)}{p^{1-\alpha}(x, v)} - f(x, v) \frac{\Delta_H p^{1-\alpha}(x, v)}{p^{1-\alpha}(x, v)} \right] \\
&\quad + \delta \frac{m_{22}}{2m_0} \left[\frac{\Delta_E^V \left(f p^{1-\alpha} \right)(x, v)}{p^{1-\alpha}(x, v)} - f(x, v) \frac{\Delta_E^V p^{1-\alpha}(x, v)}{p^{1-\alpha}(x, v)} \right] \\
&\quad + O\left(\epsilon^2 + \epsilon\delta + \delta^2 \right).
\end{aligned}
$$

Tingran Gao, "The Diffusion Geometry of Fibre Bundles: Horizontal Diffusion Maps," arXiv:1602.02330

Asymptotic Theory: Generator of the Diffusion Process

Theorem (G., 2016). If $\delta = O(\epsilon)$ as $\epsilon \to 0$, then for any $f \in C^{\infty}(E)$ and $(x, v) \in E$, as $\epsilon \to 0$,

$$
\begin{aligned}
& H_{\epsilon,\delta}^{(\alpha)} f(x, v) \\
& = f(x, v) + \epsilon \frac{m_{21}}{2m_0} \left[\frac{\Delta_H \left(f p^{1-\alpha} \right)(x, v)}{p^{1-\alpha}(x, v)} - f(x, v) \frac{\Delta_H p^{1-\alpha}(x, v)}{p^{1-\alpha}(x, v)} \right] \\
& \quad + \delta \frac{m_{22}}{2m_0} \left[\frac{\Delta_E^V \left(f p^{1-\alpha} \right)(x, v)}{p^{1-\alpha}(x, v)} - f(x, v) \frac{\Delta_E^V p^{1-\alpha}(x, v)}{p^{1-\alpha}(x, v)} \right] \\
& \quad + O\left(\epsilon^2 + \epsilon\delta + \delta^2 \right).
\end{aligned}
$$

- Δ_E^V is the vertical Laplacian on E

Asymptotic Theory: Generator of the Diffusion Process

Theorem (G., 2016). If $\delta = O(\epsilon)$ as $\epsilon \to 0$, then for any $f \in C^\infty(E)$ and $(x, v) \in E$, as $\epsilon \to 0$,

$$
\begin{aligned}
&H_{\epsilon,\delta}^{(\alpha)} f(x, v) \\
&= f(x, v) + \epsilon \frac{m_{21}}{2m_0} \left[\frac{\Delta_H \left(f p^{1-\alpha} \right)(x, v)}{p^{1-\alpha}(x, v)} - f(x, v) \frac{\Delta_H p^{1-\alpha}(x, v)}{p^{1-\alpha}(x, v)} \right] \\
&\quad + \delta \frac{m_{22}}{2m_0} \left[\frac{\Delta_E^V \left(f p^{1-\alpha} \right)(x, v)}{p^{1-\alpha}(x, v)} - f(x, v) \frac{\Delta_E^V p^{1-\alpha}(x, v)}{p^{1-\alpha}(x, v)} \right] \\
&\quad + O\left(\epsilon^2 + \epsilon\delta + \delta^2 \right).
\end{aligned}
$$

- Δ_E^V is the vertical Laplacian on E
- Δ_H is the Bochner horizontal Laplacian on E

Asymptotic Theory: Generator of the Diffusion Process

Theorem (G., 2016). If $\delta = O(\epsilon)$ as $\epsilon \to 0$, then for any $f \in C^\infty(E)$ and $(x, v) \in E$, as $\epsilon \to 0$,

$$
\begin{aligned}
&H_{\epsilon,\delta}^{(\alpha)} f(x, v) \\
&= f(x, v) + \epsilon \frac{m_{21}}{2m_0} \left[\frac{\Delta_H \left(f p^{1-\alpha} \right)(x, v)}{p^{1-\alpha}(x, v)} - f(x, v) \frac{\Delta_H p^{1-\alpha}(x, v)}{p^{1-\alpha}(x, v)} \right] \\
&\quad + \delta \frac{m_{22}}{2m_0} \left[\frac{\Delta_E^V \left(f p^{1-\alpha} \right)(x, v)}{p^{1-\alpha}(x, v)} - f(x, v) \frac{\Delta_E^V p^{1-\alpha}(x, v)}{p^{1-\alpha}(x, v)} \right] \\
&\quad + O\left(\epsilon^2 + \epsilon\delta + \delta^2 \right).
\end{aligned}
$$

- Δ_E^V is the vertical Laplacian on E
- Δ_H is the Bochner horizontal Laplacian on E
- In general $\Delta_H + \Delta_E^V \neq \Delta_E$, true if and only if π is *harmonic*

Asymptotic Theory: Generator of the Diffusion Process

Theorem (G., 2016). If $\delta = O(\epsilon)$ as $\epsilon \to 0$, then for any $f \in C^\infty(E)$ and $(x, v) \in E$, as $\epsilon \to 0$,

$$
\begin{aligned}
&H_{\epsilon,\delta}^{(\alpha)} f(x, v) \\
&= f(x, v) + \epsilon \frac{m_{21}}{2m_0} \left[\frac{\Delta_H \left(f p^{1-\alpha} \right)(x, v)}{p^{1-\alpha}(x, v)} - f(x, v) \frac{\Delta_H p^{1-\alpha}(x, v)}{p^{1-\alpha}(x, v)} \right] \\
&\quad + \delta \frac{m_{22}}{2m_0} \left[\frac{\Delta_E^V \left(f p^{1-\alpha} \right)(x, v)}{p^{1-\alpha}(x, v)} - f(x, v) \frac{\Delta_E^V p^{1-\alpha}(x, v)}{p^{1-\alpha}(x, v)} \right] \\
&\quad + O\left(\epsilon^2 + \epsilon\delta + \delta^2 \right).
\end{aligned}
$$

- Δ_E^V is the vertical Laplacian on E
- Δ_H is the Bochner horizontal Laplacian on E
- In general $\Delta_H + \Delta_E^V \neq \Delta_E$, true if and only if π is *harmonic*
- $\Rightarrow$ HDM is not applying diffusion maps to the total space!

The Convergence Rate: Diffusion Maps

Theorem (Singer, 2006). Suppose N points are i.i.d. uniformly sampled from a d-dimensional Riemannian manifold M. The graph diffusion operator $P_{\epsilon,\alpha}$ converges to its smooth limit at rate

$$O\left(N^{-\frac{1}{2}}\epsilon^{\frac{1}{2}-\frac{d}{4}}\right).$$

Corollary. Under the same assumption, non-uniform sampling has convergence rate

$$O\left(N^{-\frac{1}{2}}\epsilon^{-\frac{d}{4}}\right).$$

The Convergence Rate: HDM on Unit Tangent Bundles

Theorem (G., 2015). Suppose N_B points are i.i.d. sampled from a d-dimensional Riemannian manifold M, and N_F unit tangent vectors are i.i.d. sampled at each of the N_B samples. The graph horizontal diffusion operator $H_{\epsilon,\delta}^{(\alpha)}$ converges to its smooth limit at rate

$$O\left(\theta_*^{-1} N_B^{-\frac{1}{2}} \epsilon^{-\frac{d}{4}}\right),$$

where

$$\theta_* = 1 - \frac{1}{1 + \epsilon^{\frac{d}{4}} \delta^{\frac{d-1}{4}} \sqrt{\frac{N_F}{N_B}}}.$$

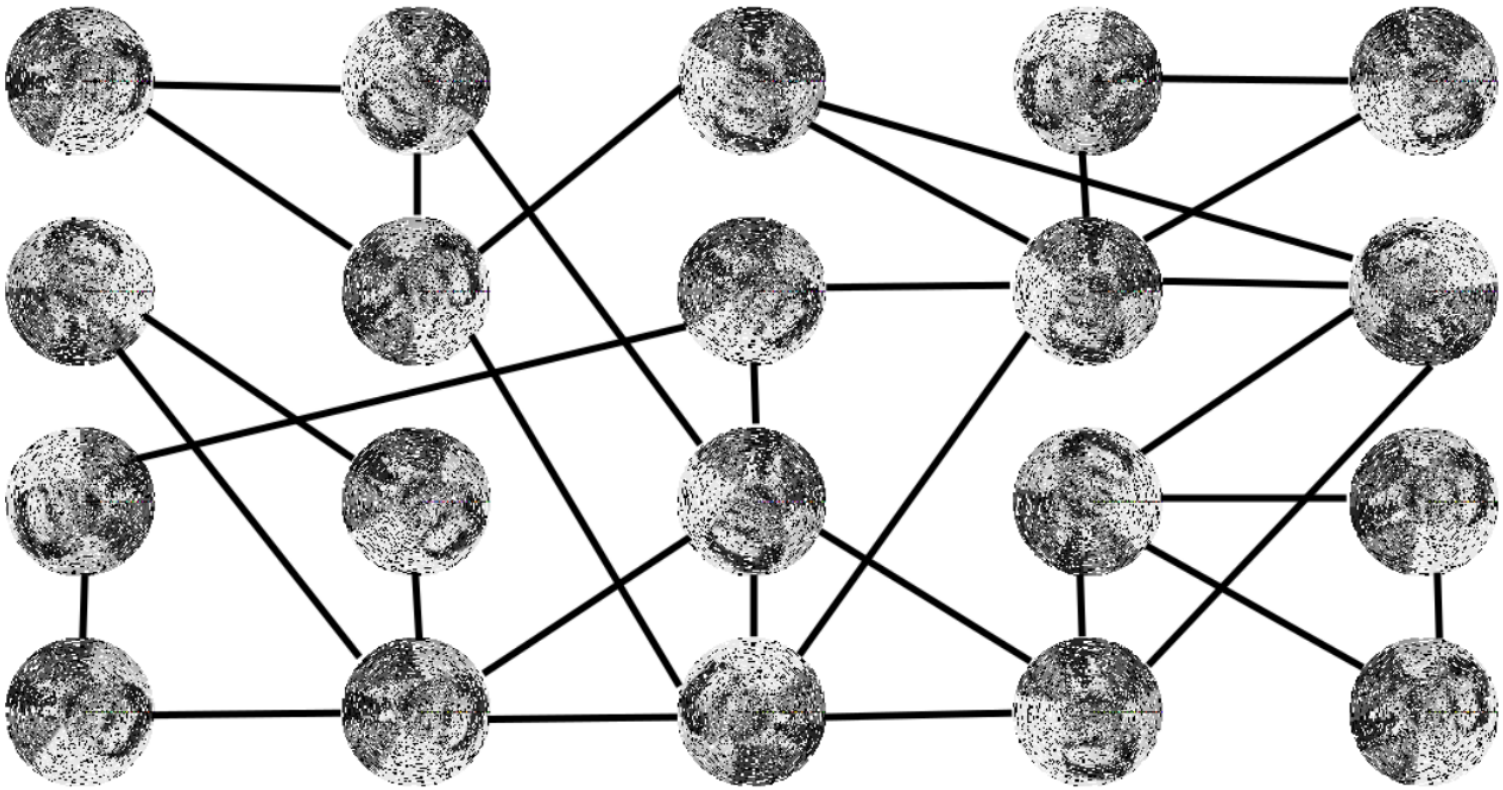
Remark. Diffusion maps on the total manifold in this setting has convergence rate of $O\left(N_F^{-\frac{1}{2}} N_B^{-\frac{1}{2}} \epsilon^{-\frac{d}{4}}\right)$, by (Singer, 2006)

Tingran Gao, "Hypoelliptic Diffusion Maps and Their Applications in Automated Geometric Morphometrics," PhD Thesis, Duke University (2015)

Outline

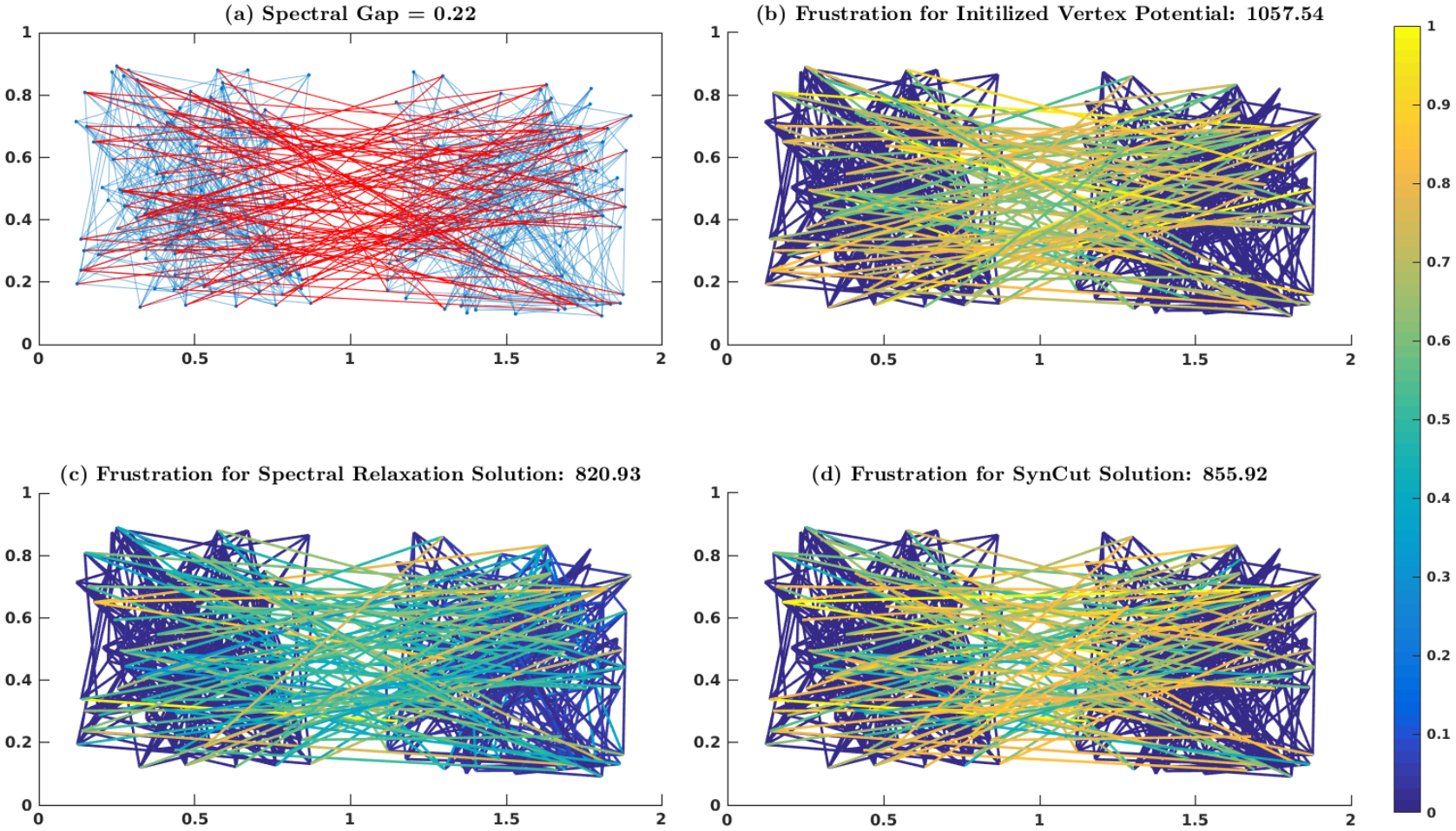
Synchronization Problems

$$y_i = R_i x + \xi_i$$
$$R_i \in O(d), \quad \xi_i \sim \text{i.i.d. noise}$$



Afonso S. Bandeira. "Ten Lectures and Forty-Two Open Problems in the Mathematics of Data Science." (2015).

Average Out the Variance, but Respect the Bias



Tingran Gao, Jacek Brodzki, Sayan Mukherjee. "The Geometry of Synchronization Problems and Learning Group Actions." *Discrete & Computational Geometry*, to appear (2019)

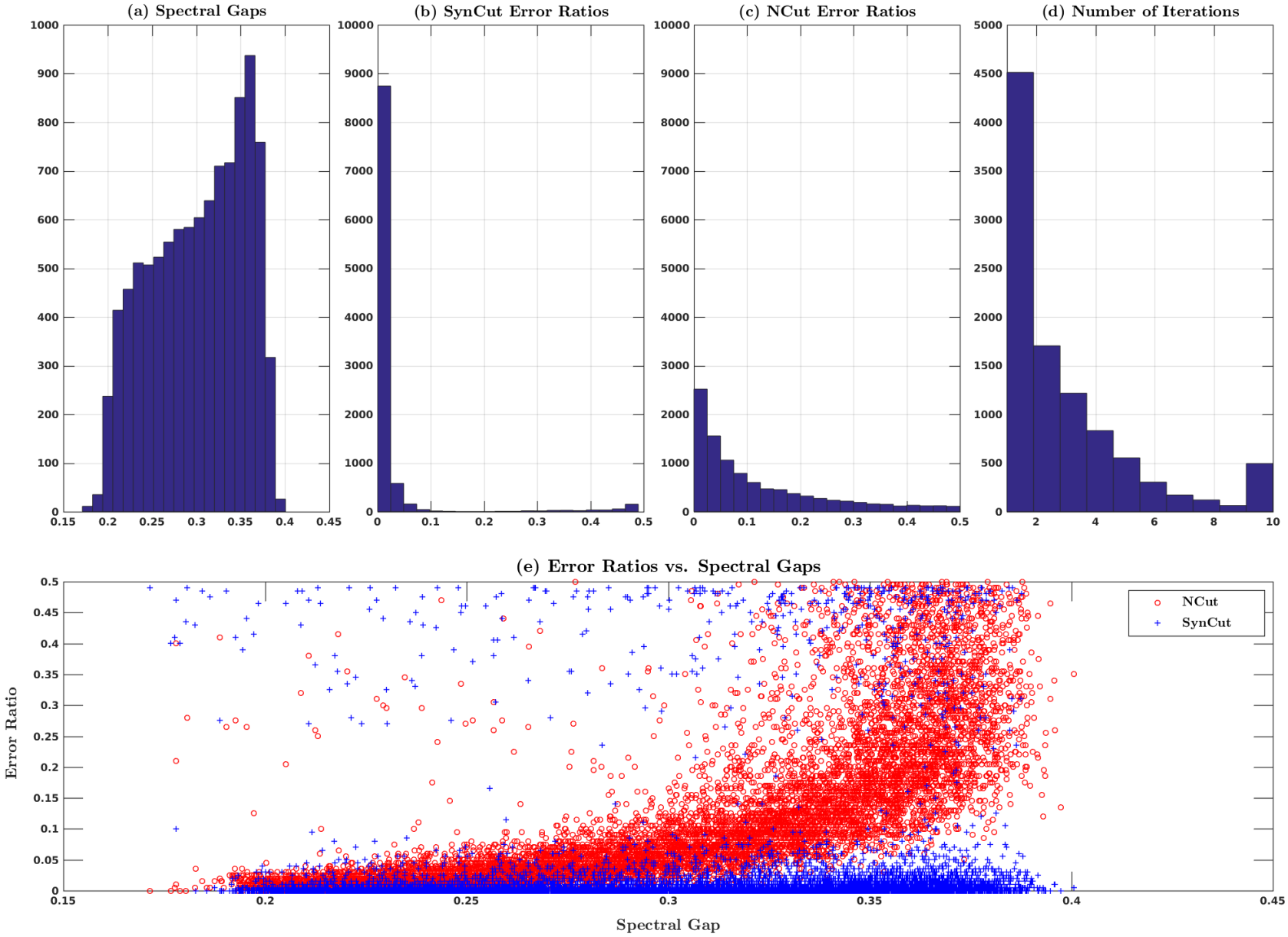
Algorithm: SynCut

Input: $\Gamma = (V, E, w)$, $\rho \in C^1(\Gamma; G)$, number of partitions K
Output: Partitions $\{S_1, \cdots, S_K\}$

1. Solve synchronization problem over Γ for ρ, obtain $f \in C^0(\Gamma; G)$
2. Compute $d_{ij} = \exp(-w_{ij} \|f_i - \rho_{ij} f_j\|)$ on all edges $(i, j) \in E$
3. Spectral clustering on weighted graph (V, E, d) to get $\{S_1, \cdots, S_k\}$
4. Solve synchronization problem within each partition S_j, "glue up" the local solutions to obtain $f_* \in C^0(\Gamma; G)$
5. $f \leftarrow f_*$, repeat from Step 2

Tingran Gao, Jacek Brodzki, Sayan Mukherjee. "The Geometry of Synchronization Problems and Learning Group Actions." *Discrete & Computational Geometry*, to appear (2019)

Compare with Graph-based Spectral Clustering



Stochastic Block + Random Corruption Model

- Consider graph $\Gamma = (V, E)$, where $V = (v_1, \cdots, v_n)$ are separated into 2 communities
- Fix orthogonal matrices $R_1, \cdots, R_n \in O(d)$, one at each corresponding vertex
- Fix $0 \leq q < p \leq 1$. Consider the following $\rho \in C^1(\Gamma; O(d))$:
  - If v_i, v_j belong to the same community,

$$\rho_{ij} = \begin{cases} R_i^\top R_j & \text{with probability } p \\ \mathrm{Unif}(O(d)) & \text{with probability } 1-p \end{cases}$$

  - If v_i, v_j belong to different communities

$$\rho_{ij} = \begin{cases} R_i^\top R_j & \text{with probability } q \\ \mathrm{Unif}(O(d)) & \text{with probability } 1-q \end{cases}$$

Stochastic Block + Random Corruption Model

- Fix $0 \leq q < p \leq 1$. Consider the following $\rho \in C^1(\Gamma; O(d))$:
  - If v_i, v_j belong to the same community,

$$\rho_{ij} = \begin{cases} R_i^\top R_j & \text{with probability } p \\ \mathrm{Unif}(O(d)) & \text{with probability } 1-p \end{cases}$$

  - If v_i, v_j belong to different communities

$$\rho_{ij} = \begin{cases} R_i^\top R_j & \text{with probability } q \\ \mathrm{Unif}(O(d)) & \text{with probability } 1-q \end{cases}$$

- "Clean" case ($p = q = 0$): synchronization solution

$$\tilde{\rho}_{ij} = R_i^\top R_j$$

Stochastic Block + Random Corruption Model

- Fix $0 \leq q < p \leq 1$. Consider the following $\rho \in C^1(\Gamma; O(d))$:
  - If v_i, v_j belong to the same community,

$$\rho_{ij} = \begin{cases} R_i^\top R_j & \text{with probability } p \\ \mathrm{Unif}(O(d)) & \text{with probability } 1-p \end{cases}$$

  - If v_i, v_j belong to different communities

$$\rho_{ij} = \begin{cases} R_i^\top R_j & \text{with probability } q \\ \mathrm{Unif}(O(d)) & \text{with probability } 1-q \end{cases}$$

- "Perturbed case $(q > p > 0)$": synchronization solution

$$\tilde{\rho}_{ij} \approx R_i^\top R_j$$

Stochastic Block + Random Corruption Model

- Fix $0 \leq q < p \leq 1$. Consider the following $\rho \in C^1(\Gamma; O(d))$:
  - If v_i, v_j belong to the same community,

$$\rho_{ij} = \begin{cases} R_i^\top R_j & \text{with probability } p \\ \mathrm{Unif}(O(d)) & \text{with probability } 1-p \end{cases}$$

  - If v_i, v_j belong to different communities

$$\rho_{ij} = \begin{cases} R_i^\top R_j & \text{with probability } q \\ \mathrm{Unif}(O(d)) & \text{with probability } 1-q \end{cases}$$

- "Perturbed case $(q > p > 0)$": synchronization solution

$$\tilde{\rho}_{ij} \approx R_i^\top R_j$$

- But $\|\tilde{\rho}_{ij} - \rho_{ij}\|$ deviates from 0 more frequently on edges connecting different communities!

Stochastic Block + Random Corruption Model

- Fix $0 \leq q < p \leq 1$. Consider the following $\rho \in C^1(\Gamma; O(d))$:
  - If v_i, v_j belong to the same community,

$$\rho_{ij} = \begin{cases} R_i^\top R_j & \text{with probability } p \\ \mathrm{Unif}(O(d)) & \text{with probability } 1-p \end{cases}$$

  - If v_i, v_j belong to different communities

$$\rho_{ij} = \begin{cases} R_i^\top R_j & \text{with probability } q \\ \mathrm{Unif}(O(d)) & \text{with probability } 1-q \end{cases}$$

- "Perturbed case $(q > p > 0)$": synchronization solution

$$\tilde{\rho}_{ij} \approx R_i^\top R_j$$

- But $\|\tilde{\rho}_{ij} - \rho_{ij}\|$ deviates from 0 more frequently on edges connecting different communities!
- suggesting normalized graph cut with new weights $\|\tilde{\rho}_{ij} - \rho_{ij}\|$

Simultaneous Matching and Clustering (SMAC)

- Established results for permutation groups in the context of collection shape matching
- Stochastic Block + Random Corruption for the maps, Erdős-Rényi for the observation graph, $K \geq 2$ balanced clusters, synchronization with spectral methods
- The recovery rate for the underlying clusters matches the information theoretic lower bound in [Chen et al. 2017]

• Yuxin Chen, Changho Suh, and Andrea J. Goldsmith. "Information Recovery from Pairwise Measurements." *IEEE Transactions on Information Theory* 62, no. 10 (2016): 5881-5905.
• Chandrajit Bajaj, **Tingran Gao**, Zihang He, Qixing Huang, and Zhenxiao Liang, "SMAC: Simultaneous Mapping and Clustering Using Spectral Decompositions," *Proceedings of the 35th International Conference on Machine Learning, PMLR* 80:324-333, 2018

Algorithm 1 PermSMAC: Simultaneously mapping and clustering

Input: Observation graph $\mathcal{G} = (\mathcal{S}, \mathcal{E})$ and initial pairwise maps $X_{ij}^{in}, (i,j) \in \mathcal{E}$

Output: Underlying clusters $\mathcal{S} = c_1 \cup \cdots \cup c_k$ and optimized pairwise maps $X_{ij}, 1 \leq i, j \leq n$

1: {**Step 1**} Simultaneously compute the intra-cluster maps and extract the underlying clusters:

2: {**Step 1.1**} Form data matrix based on (1).

3: {**Step 1.2**} Compute the critical value $r = \underset{2 \leq i \leq nm}{\operatorname{argmax}} \frac{\lambda_i - \lambda_{i+1}}{\lambda_i + \lambda_{i+1}}$.

4: {**Step 1.3**} Let $U \in \mathbb{R}^{nm \times r}$ store the leading r eigenvectors of X. Compute pair-wise maps $X_{ij}^{\star}$ by solving (2)

5: {**Step 1.4**} Use $f_{ij}(X_{ij}^{\star})$ as the affinity score and apply single-linkage clustering to obtain the underlying clusters

6: {**Step 2**} compute the inter-cluster maps by solving (6)

SMAC: Exact Recovery Conditions

Theorem (Bajaj et al. 2018). Under mild assumptions, `PermSMAC` recovers the underlying clusters and intra-cluster maps with high probability provided

$$p - q \geq ck\sqrt{\frac{\log n}{nt}}$$

for some constant $c > 0$, where $t = \Omega\left(\log\left(n\right)/n\right)$. The inter-cluster maps are recovered with high probability provided

$$q \geq c'k\sqrt{\frac{\log n}{n^2 t}}$$

for some constant $c' > 0$.

• Chandrajit Bajaj, **Tingran Gao**, Zihang He, Qixing Huang, and Zhenxiao Liang, "SMAC: Simultaneous Mapping and Clustering Using Spectral Decompositions," *Proceedings of the 35th International Conference on Machine Learning, PMLR* 80:324-333, 2018

Collaborators

Doug Boyer
Duke

Ingrid Daubechies
Duke

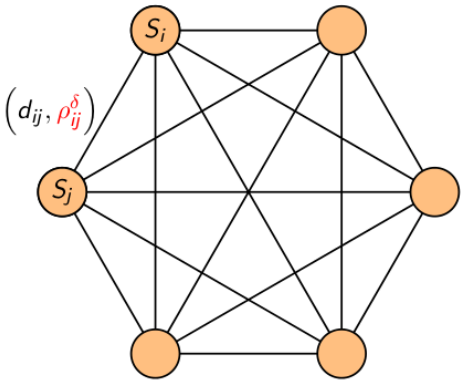
Sayan Mukherjee
Duke

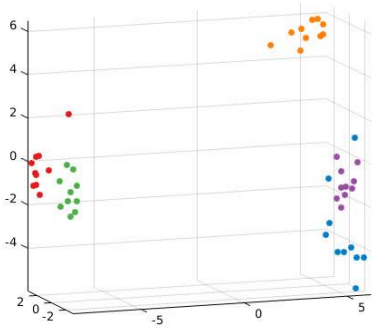
Jacek Brodzki
Southamption

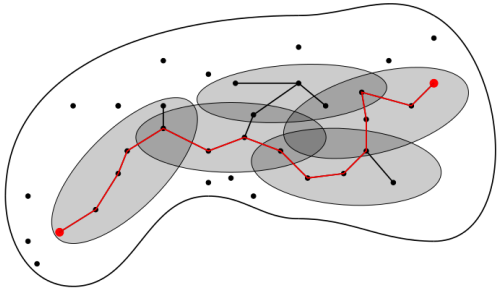
Qixing Huang
UT Austin

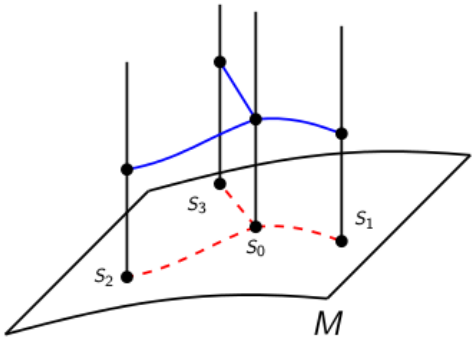
Chandrajit Bajaj
UT Austin

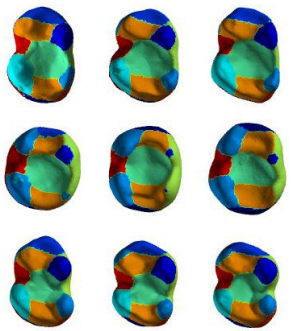
Thank You!

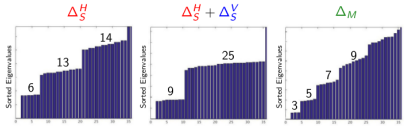












$$O\left(\theta_*^{-1} N_B^{-\frac{1}{2}} \epsilon^{-\frac{d}{4}}\right),$$

$$\theta_* = 1 - \frac{1}{1 + \epsilon^{\frac{d}{4}} \delta^{\frac{d-1}{4}} \sqrt{\frac{N_F}{N_B}}}.$$

- **Tingran Gao**, "Hypoelliptic Diffusion Maps and Their Applications in Automated Geometric Morphometrics," PhD Thesis, Duke University (2015)
- **Tingran Gao**, "The Diffusion Geometry of Fibre Bundles: Horizontal Diffusion Maps." arXiv:1602.02330
- **Tingran Gao**, Jacek Brodzki, Sayan Mukherjee. "The Geometry of Synchronization Problems and Learning Group Actions." *Discrete & Computational Geometry*, to appear (2019)
- Chandrajit Bajaj, **Tingran Gao**, Zihang He, Qixing Huang, and Zhenxiao Liang, "SMAC: Simultaneous Mapping and Clustering Using Spectral Decompositions," *Proceedings of the 35th International Conference on Machine Learning, PMLR* 80:324-333, 2018