Machine Learning, Fibre Bundles and Biological Morphology

Ingrid Daubechies   Tingran Gao

Department of Mathematics
Duke University

Feb 11, 2016

What this talk is about

Background

- motivation: mophometrics
- previous work: continuous Procrustes distance

Learning from Correspondences

- diffusion geometry
- fibre bundles
- horizontal random walks

Results

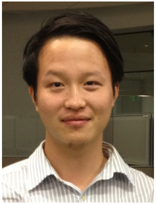
- biological applications
- mathematical theory

Collaborators


Rima Alaifari
ETH Zürich


Doug Boyer
Duke


Ingrid Daubechies
Duke


Tingran Gao
Duke


Yaron Lipman
Weizmann


Roi Poranne
ETH Zürich


Jesús Puente
J.P. Morgan


Robert Ravier
Duke

It all started with a conversation with biologists....

Doug Boyer

Jukka Jernvall

More Precisely: biological morphologists

↓

Study Teeth & Bones of

extant & extinct animals

still live today — fossils

First: project on "complexity" of teeth

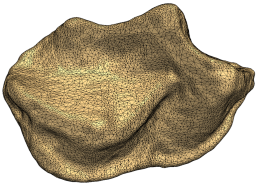
Then: find automatic way to compute Procrustes distances between surfaces — without landmarks

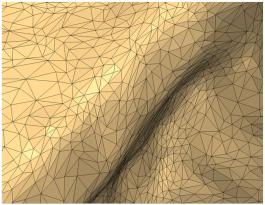
Landmarked Teeth $\longrightarrow$

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

Find way to compute a distance that does as well, for biological purposes, as Procrustes distance, based on expert-placed landmarks, automatically?

Examples: finely discretized triangulated surfaces



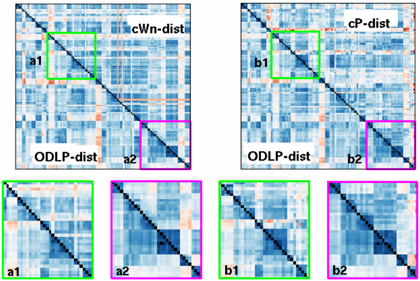


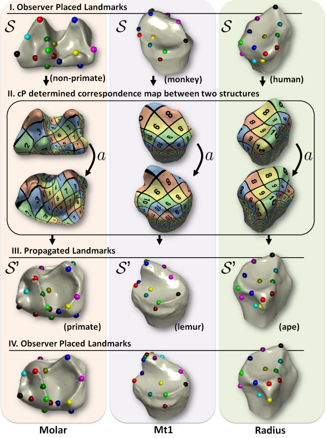
We defined 2 different distances

$d_{\mathrm{cWn}}(S_1, S_2)$: conformal flattening
comparison of neighborhood geometry
optimal mass transport

$d_{\mathrm{cP}}(S_1, S_2)$: continuous Procrustes distance

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



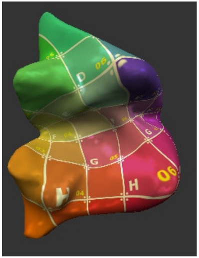


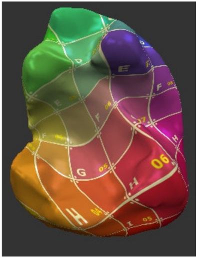
We defined 2 different distances

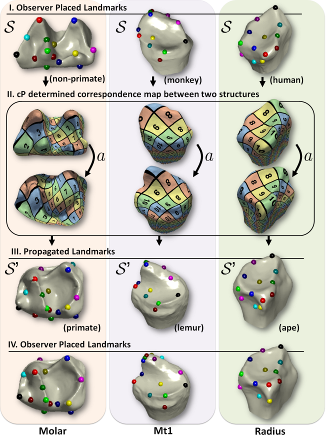
$d_{\mathrm{cWn}}(S_1, S_2)$: conformal flattening
comparison of neighborhood geometry
optimal mass transport

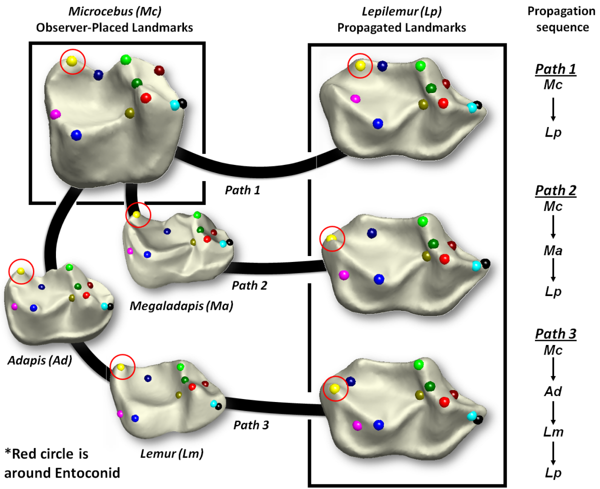
$d_{\mathrm{cP}}(S_1, S_2)$: continuous Procrustes distance

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









Even mistake made by d_{cP} were similar to biologists' mistakes

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

Biologists' "wish list" changed...

... as they learned our language and saw our methods

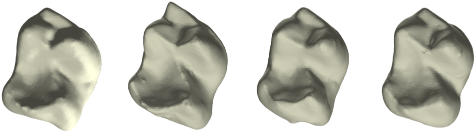
- mappings more important to them than distances ($\longrightarrow$ discussion of variability in individuals or between species, locally)
- no *holonomy!*

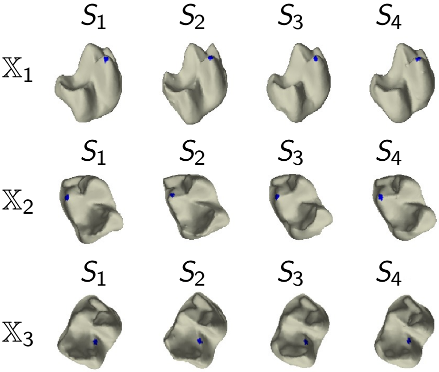
Our formulation of problem changed too

Tingran Gao $\longrightarrow$ reformulate as connection on fibre bundle
$+$ horizontal diffusion

Even before this...

biological content in large concatenated matrix





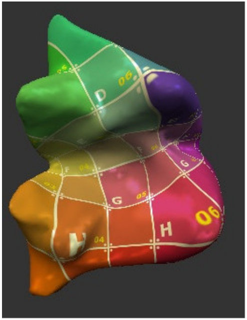
$$\begin{aligned} \min \quad & \|M\mathbb{X} - \mathbb{X}\|_2^2 + \lambda \|\mathbb{X}\|_1 \\ \text{s.t.} \quad & \|\mathbb{X}\|_2 = 1. \end{aligned}$$

Resulting minimizers $\mathbb{X}$ supported on union of 4 surfaces

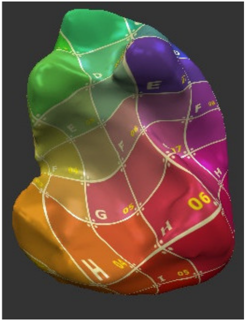
Use the Information in the Maps!

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

S_1



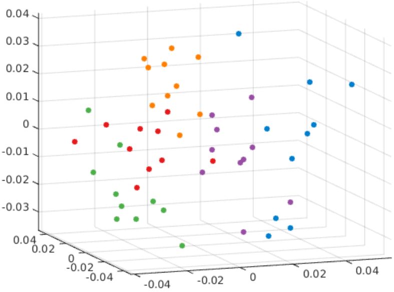
$\xrightarrow[f_{12}]{d_{12}}$

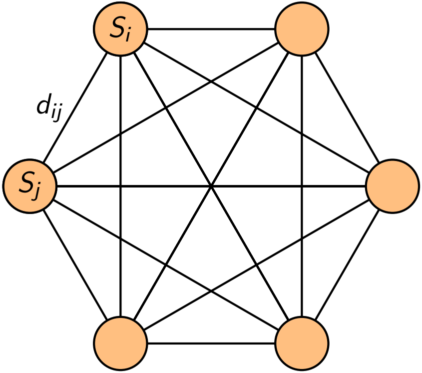


S_2

Learning from Distances

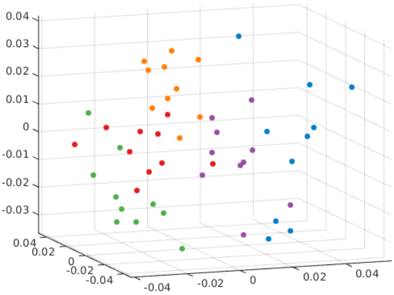
$$\begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1N} \\ d_{21} & d_{22} & \cdots & d_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \cdots & d_{NN} \end{pmatrix}$$

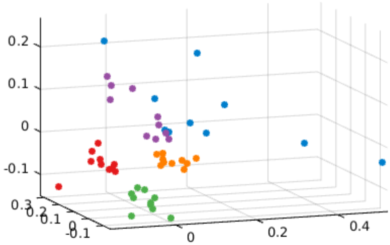




Learning from Distances

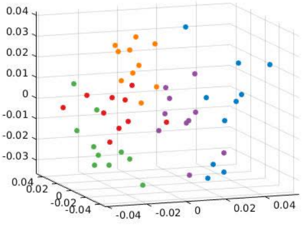
$$\begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1N} \\ d_{21} & d_{22} & \cdots & d_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \cdots & d_{NN} \end{pmatrix}$$



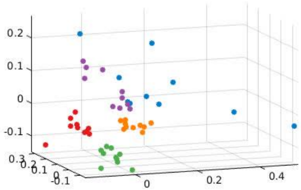


Diffusion Distance

MDS for CPD & DD

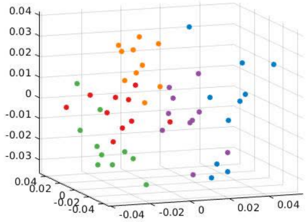


CPD

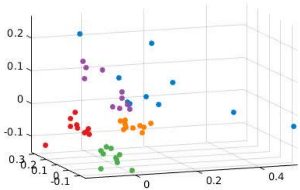


DD

MDS for CPD & DD

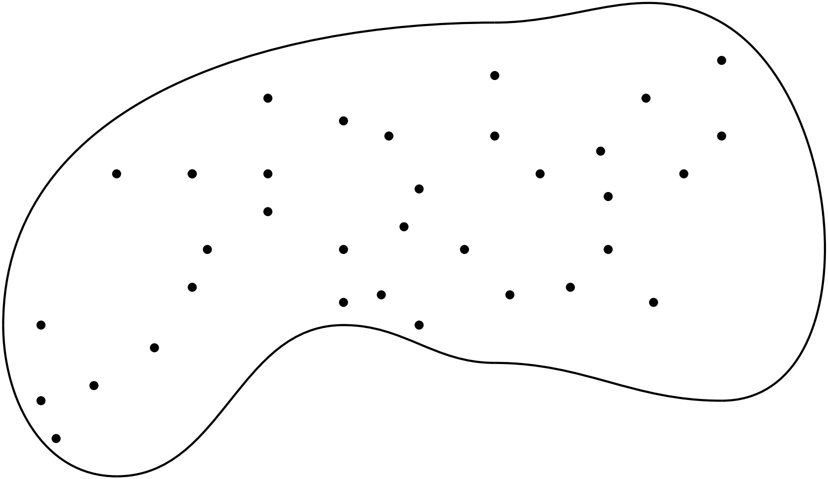


CPD

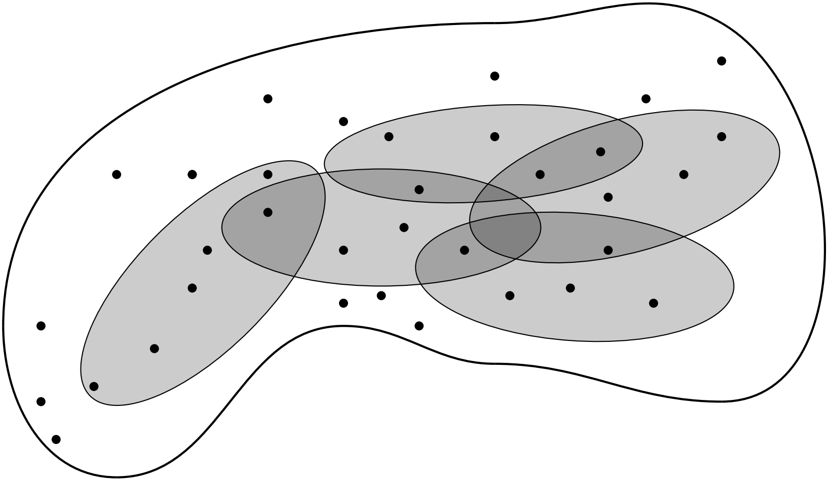


DD

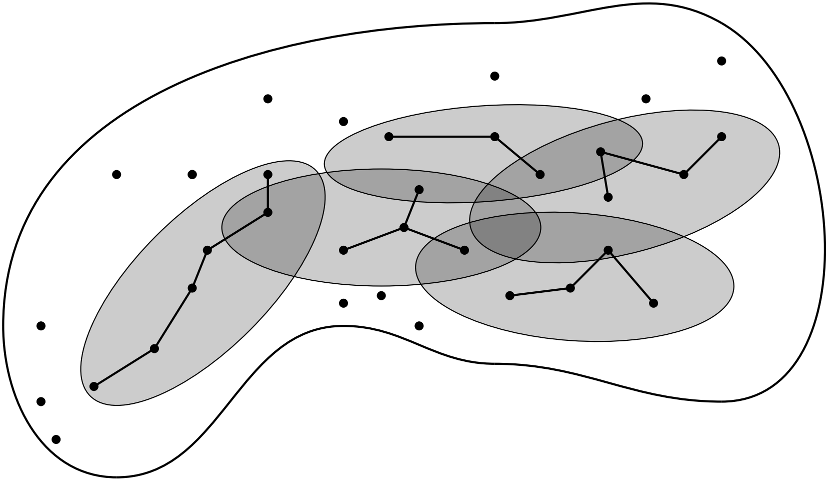
Diffusion Maps: “Knit Together” Local Geometry



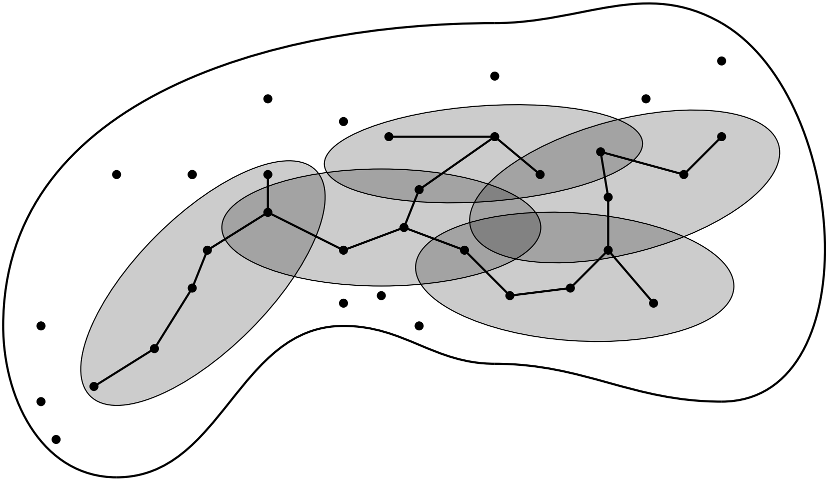
Diffusion Maps: “Knit Together” Local Geometry



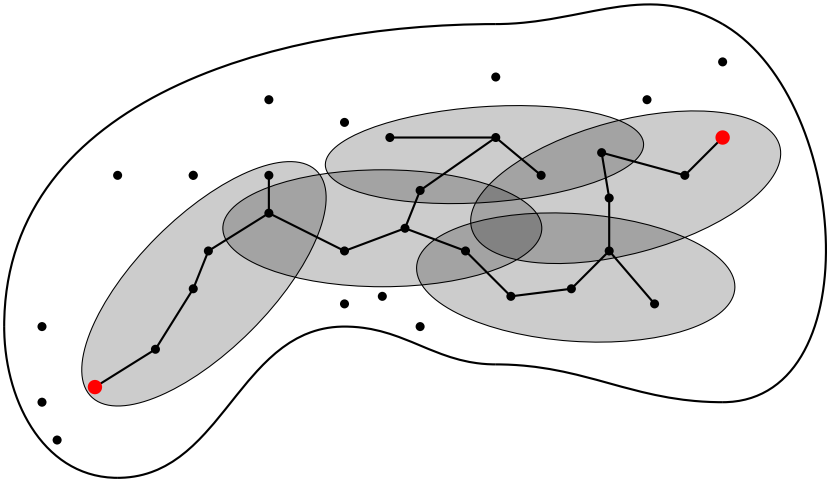
Diffusion Maps: “Knit Together” Local Geometry



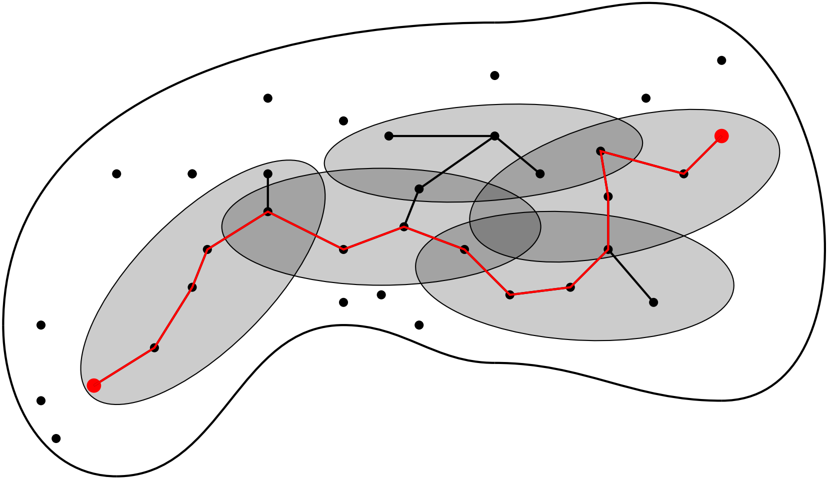
Diffusion Maps: "Knit Together" Local Geometry



Diffusion Maps: “Knit Together” Local Geometry

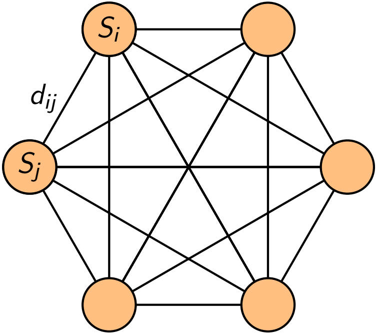


Diffusion Maps: "Knit Together" Local Geometry



Small distances are much more reliable!

Diffusion Maps: "Knit Together" Local Geometry



- $P = D^{-1}W$ defines a random walk on the graph
- Solve eigen-problem

$$Pu_j = \lambda_j u_j,\ j = 1, 2, \cdots, m$$

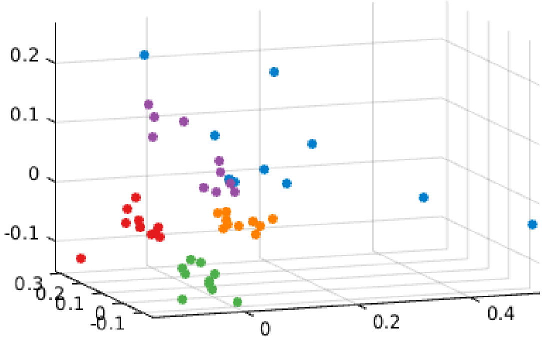
and represent each individual shape S_j as an m-vector

$$\left(\lambda_1^{t/2} u_1(j), \cdots, \lambda_m^{t/2} u_m(j)\right)$$

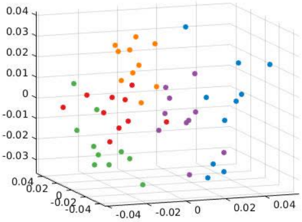
Diffusion Distance (DD)

Fix $1 \leq m \leq N,\ t \geq 0,$

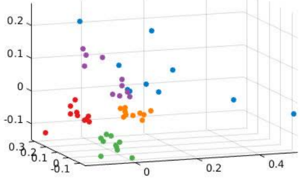
$$D_m^t\left(S_i, S_j\right) = \left(\sum_{k=1}^{m} \lambda_k^t \left(u_k\left(i\right) - u_k\left(j\right)\right)^2\right)^{\frac{1}{2}}$$



MDS for CPD & DD

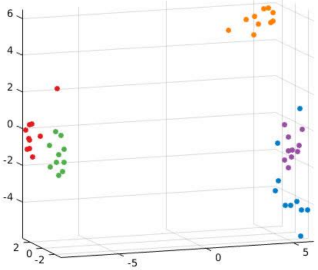


CPD

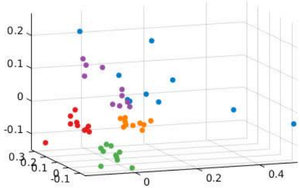


DD

Even Better: More Information!

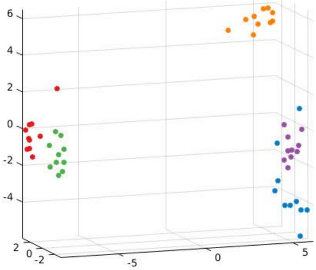


HBDD

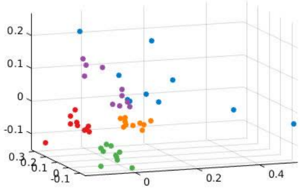


DD

Even Better: More Information!



HBDD

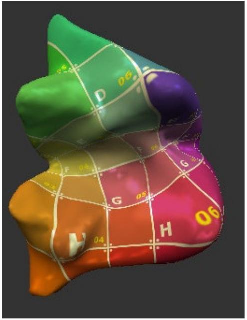


DD

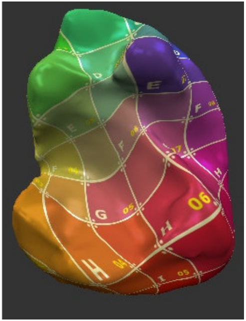
Use the Information in the Maps!

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

S_1

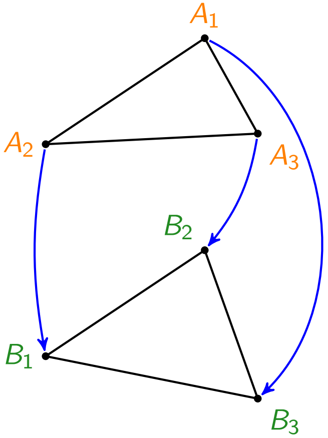


$\xrightarrow[f_{12}]{d_{12}}$



S_2

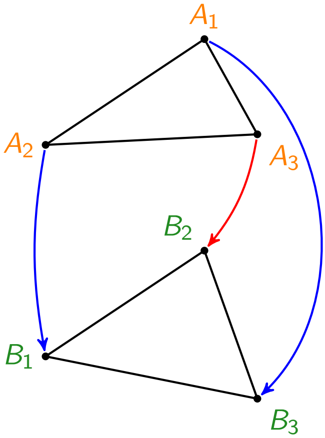
Correspondences Between Triangular Meshes



$$\xrightarrow{S_2}$$

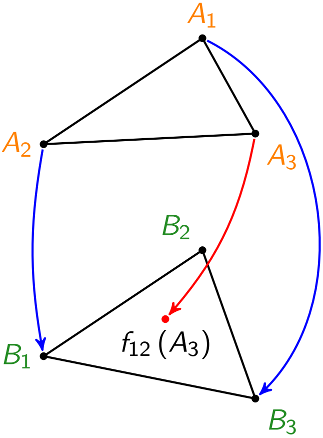
$$S_1 \downarrow \quad \begin{array}{c} \\ \vdots \\ A_1 \\ A_2 \\ A_3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} \cdots & B_1 & B_2 & B_3 & \cdots \end{array} \\ \left(\begin{array}{ccccc} & \vdots & \vdots & \vdots & \\ \cdots & 0 & 0 & 1 & \cdots \\ \cdots & 1 & 0 & 0 & \cdots \\ \cdots & 0 & 1 & 0 & \cdots \\ & \vdots & \vdots & \vdots & \end{array} \right) \end{array}$$

Correspondences Between Triangular Meshes



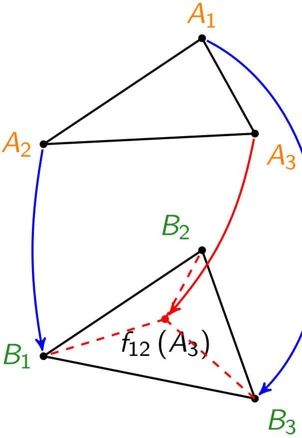
$$
\begin{array}{c c}
 & \xrightarrow{\quad s_2 \quad} \\
 & \begin{array}{ccccc} \cdots & B_1 & B_2 & B_3 & \cdots \end{array} \\
s_1 \downarrow \begin{array}{c} \vdots \\ A_1 \\ A_2 \\ A_3 \\ \vdots \end{array} & \begin{pmatrix} & \vdots & \vdots & \vdots & \\ \cdots & 0 & 0 & 1 & \cdots \\ \cdots & 1 & 0 & 0 & \cdots \\ \cdots & 0 & \boxed{1} & 0 & \cdots \\ & \vdots & \vdots & \vdots & \end{pmatrix}
\end{array}
$$

Correspondences Between Triangular Meshes



$$\underset{s_1 \downarrow}{} \quad \overset{s_2 \longrightarrow}{\begin{array}{c|ccccc} & \cdots & B_1 & B_2 & B_3 & \cdots \\ \vdots & & \vdots & \vdots & \vdots & \\ A_1 & \cdots & 0 & 0 & 1 & \cdots \\ A_2 & \cdots & 1 & 0 & 0 & \cdots \\ A_3 & \cdots & 0 & \boxed{?} & 0 & \cdots \\ \vdots & & \vdots & \vdots & \vdots & \end{array}}$$

Correspondences Between Triangular Meshes

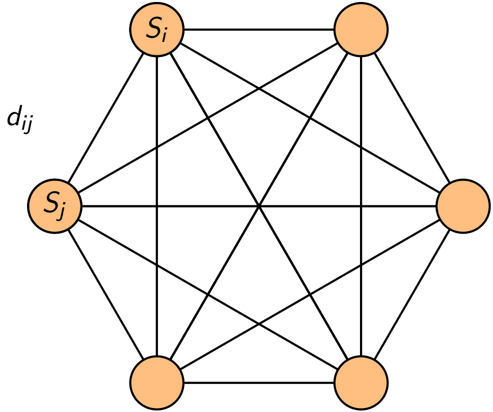


$$s_2 \longrightarrow$$

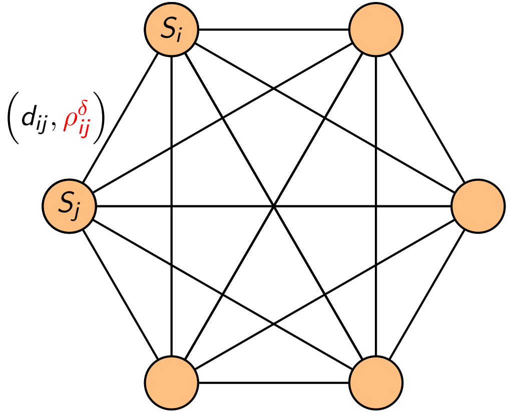
$$s_1 \downarrow \quad \begin{array}{c} \\ \vdots \\ A_1 \\ A_2 \\ A_3 \\ \vdots \end{array} \begin{pmatrix} \cdots & B_1 & B_2 & B_3 & \cdots \\ & \vdots & \vdots & \vdots & \\ \cdots & 0 & 0 & 1 & \cdots \\ \cdots & 1 & 0 & 0 & \cdots \\ \cdots & \boxed{0.91} & \boxed{0.95} & \boxed{0.88} & \cdots \\ & \vdots & \vdots & \vdots & \end{pmatrix}$$

$$\rho_{12}^{\delta}(r,s) = \exp\left(-\frac{\left\| f_{12}(A_r) - B_s \right\|^2}{\delta}\right)$$

Distance Graph



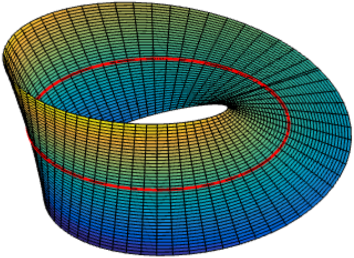
Augmented Distance Graph

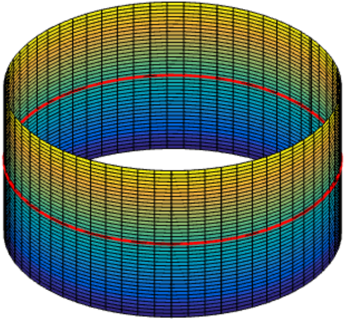


Horizontal Random Walk on a *Fibre Bundle*

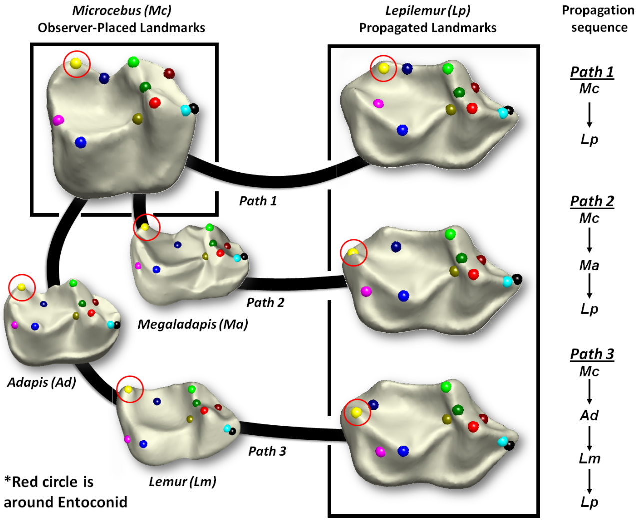
Fibre Bundle $\mathscr{E} = (E, M, F, \pi)$

- E: *total* manifold
- M: *base* manifold
- $\pi : E \to M$: smooth surjective map (*bundle projection*)
- F: *fibre* manifold
- *local triviality*: for "small" open set $U \subset M$, $\pi^{-1}(U)$ is diffeomorphic to $U \times F$





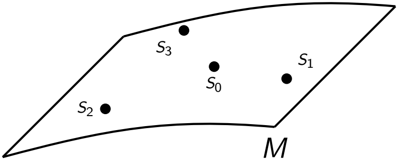
Shape Space is NOT a Trivial Fibre Bundle

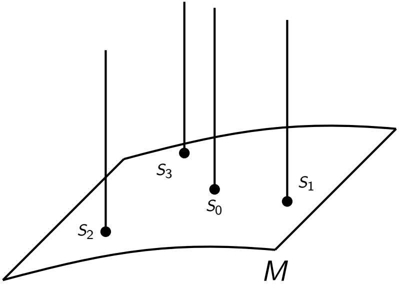


Horizontal Random Walk on a *Fibre Bundle*

Fibre Bundle $\mathscr{E} = (E, M, F, \pi)$

- E: *total* manifold
- M: *base* manifold
- $\pi : E \to M$: smooth surjective map (*bundle projection*)
- F: *fibre* manifold
- *local triviality*: for "small" open set $U \subset M$, $\pi^{-1}(U)$ is diffeomorphic to $U \times F$

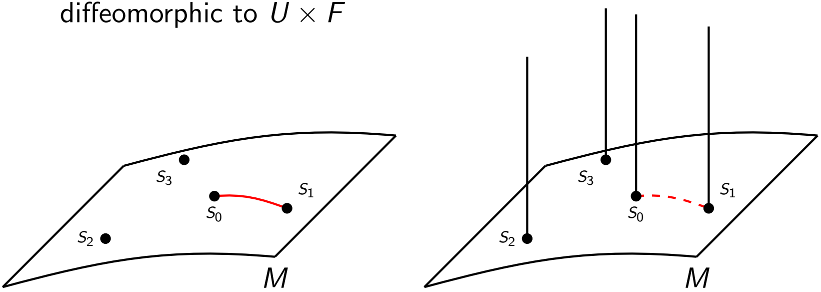




Horizontal Random Walk on a Fibre Bundle

Fibre Bundle $\mathscr{E} = (E, M, F, \pi)$

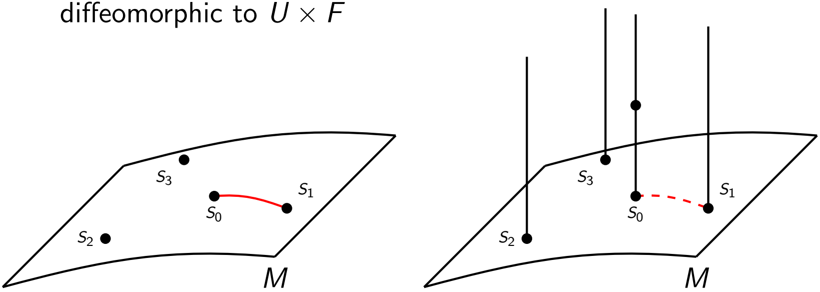
- E: *total* manifold
- M: *base* manifold
- $\pi : E \to M$: smooth surjective map (*bundle projection*)
- F: *fibre* manifold
- *local triviality*: for "small" open set $U \subset M$, $\pi^{-1}(U)$ is diffeomorphic to $U \times F$



Horizontal Random Walk on a *Fibre Bundle*

Fibre Bundle $\mathscr{E} = (E, M, F, \pi)$

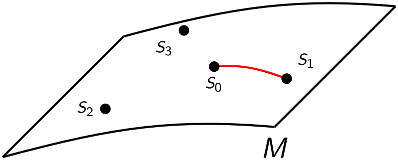
- E: *total* manifold
- M: *base* manifold
- $\pi : E \to M$: smooth surjective map (*bundle projection*)
- F: *fibre* manifold
- *local triviality*: for "small" open set $U \subset M$, $\pi^{-1}(U)$ is diffeomorphic to $U \times F$

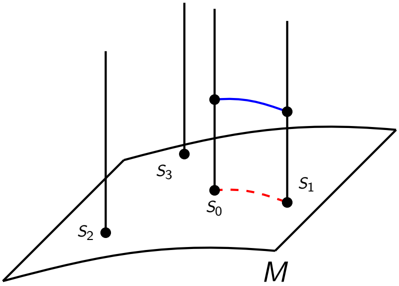


Horizontal Random Walk on a *Fibre Bundle*

Fibre Bundle $\mathscr{E} = (E, M, F, \pi)$

- E: *total* manifold
- M: *base* manifold
- $\pi : E \to M$: smooth surjective map (*bundle projection*)
- F: *fibre* manifold
- *local triviality*: for "small" open set $U \subset M$, $\pi^{-1}(U)$ is diffeomorphic to $U \times F$

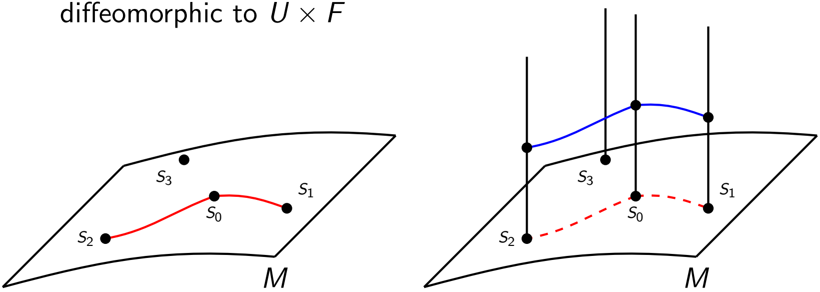




Horizontal Random Walk on a *Fibre Bundle*

Fibre Bundle $\mathscr{E} = (E, M, F, \pi)$

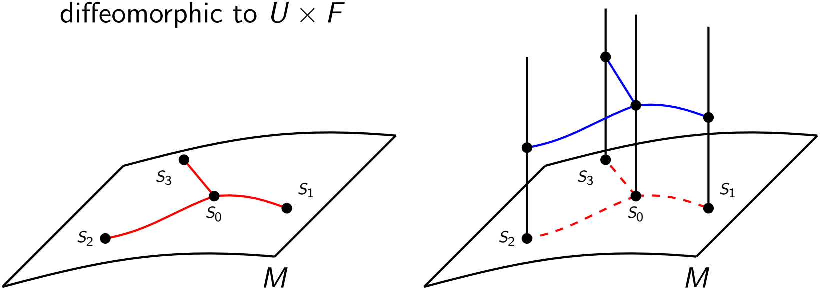
- E: *total* manifold
- M: *base* manifold
- $\pi : E \to M$: smooth surjective map (*bundle projection*)
- F: *fibre* manifold
- *local triviality*: for "small" open set $U \subset M$, $\pi^{-1}(U)$ is diffeomorphic to $U \times F$



Horizontal Random Walk on a *Fibre Bundle*

Fibre Bundle $\mathscr{E} = (E, M, F, \pi)$

- E: *total* manifold
- M: *base* manifold
- $\pi : E \to M$: smooth surjective map (*bundle projection*)
- F: *fibre* manifold
- *local triviality*: for "small" open set $U \subset M$, $\pi^{-1}(U)$ is diffeomorphic to $U \times F$



Towards *Horizontal* Diffusion Maps

Diffusion Maps

$$D^{-1}Wu_k = \lambda_k u_k, \quad 1 \leq k \leq N$$

$$D^{-1}\begin{pmatrix} & & \vdots & \\ & & \vdots & \\ \cdots & \cdots & e^{-d_{ij}^2/\epsilon} & \cdots \\ & & \vdots & \end{pmatrix}\begin{pmatrix} \vdots \\ \vdots \\ u_k\left(j\right) \\ \vdots \end{pmatrix} = \lambda_k \begin{pmatrix} \vdots \\ \vdots \\ u_k\left(j\right) \\ \vdots \end{pmatrix}$$

Towards *Horizontal* Diffusion Maps

Horizontal Diffusion Maps

$$\mathcal{D}^{-1}\mathcal{W}u_k = \lambda_k u_k, \quad 1 \leq k \leq \kappa$$

$$\mathcal{D}^{-1}\begin{pmatrix} & & \vdots & \\ & & \vdots & \\ \cdots & \cdots & e^{-d_{ij}^2/\epsilon}\rho_{ij}^{\delta} & \cdots \\ & & \vdots & \end{pmatrix}\begin{pmatrix} \vdots \\ \vdots \\ u_{k[j]} \\ \vdots \end{pmatrix} = \lambda_k \begin{pmatrix} \vdots \\ \vdots \\ u_{k[j]} \\ \vdots \end{pmatrix}$$

Towards *Horizontal* Diffusion Maps

Horizontal Diffusion Maps

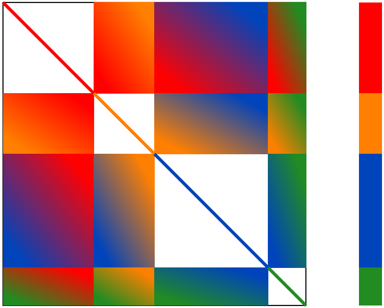
$$\mathcal{D}^{-1}\mathcal{W}u_k = \lambda_k u_k, \quad 1 \leq k \leq \kappa$$

$$\mathcal{D}^{-1}\begin{pmatrix} & & \vdots & \\ & & \vdots & \\ \cdots & \cdots & \boxed{e^{-d_{ij}^2/\epsilon}\rho_{ij}^{\delta}} & \cdots \\ & & \vdots & \end{pmatrix}\begin{pmatrix} \vdots \\ \vdots \\ \boxed{u_{k[j]}} \\ \vdots \end{pmatrix} = \lambda_k \begin{pmatrix} \vdots \\ \vdots \\ \boxed{u_{k[j]}} \\ \vdots \end{pmatrix}$$

Towards *Horizontal* Diffusion Maps

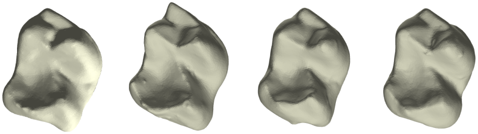
Horizontal Diffusion Maps

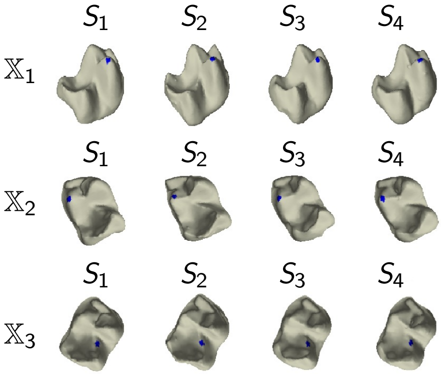
$$\mathcal{D}^{-1}\mathcal{W}u_k = \lambda_k u_k, \quad 1 \leq k \leq \kappa$$

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

Even before this...

biological content in large concatenated matrix





$$\begin{aligned} \min \quad & \|{\color{red}M}\mathbb{X} - \mathbb{X}\|_2^2 + \lambda \|\mathbb{X}\|_1 \\ \text{s.t.} \quad & \|\mathbb{X}\|_2 = 1. \end{aligned}$$

Resulting minimizers $\mathbb{X}$ supported on <u>union</u> of 4 surfaces

Towards *Horizontal* Diffusion Maps

Horizontal Diffusion Maps

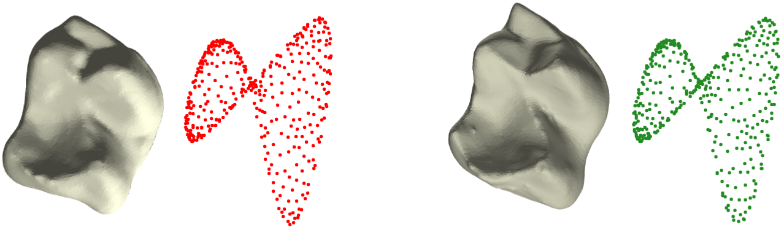
$$\mathcal{D}^{-1}\mathcal{W}u_k = \lambda_k u_k, \quad 1 \leq k \leq \kappa$$

$$\mathcal{D}^{-1}\begin{pmatrix} & & \vdots & \\ & & \vdots & \\ \cdots & \cdots & e^{-d_{ij}^2/\epsilon}\rho_{ij}^{\delta} & \cdots \\ & & \vdots & \end{pmatrix}\begin{pmatrix} \vdots \\ \vdots \\ u_{k[j]} \\ \vdots \end{pmatrix} = \lambda_k \begin{pmatrix} \vdots \\ \vdots \\ u_{k[j]} \\ \vdots \end{pmatrix}$$

Horizontal Diffusion Maps: For fixed $1 \leq m \leq \kappa$, $t \geq 0$, represent S_j as a $\kappa_j \times m$ *matrix*

$$\left(\lambda_1^{t/2}u_{1[j]}, \cdots, \lambda_m^{t/2}u_{m[j]}\right)$$

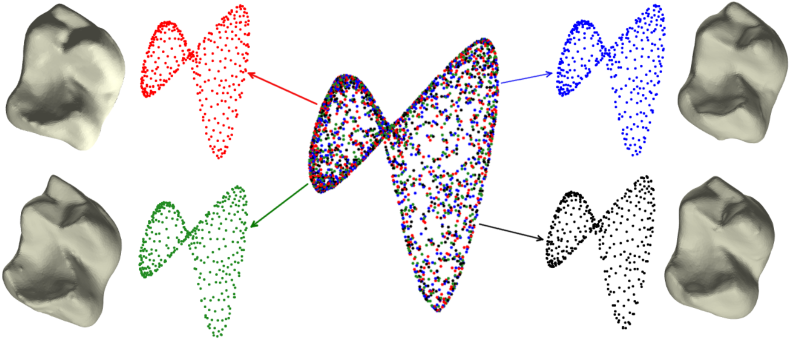
Horizontal Diffusion Maps



Horizontal Diffusion Maps: For fixed $1 \leq m \leq \kappa$, $t \geq 0$, represent S_j as a $\kappa_j \times m$ *matrix*

$$\left(\lambda_1^{t/2} u_{1[j]}, \cdots, \lambda_m^{t/2} u_{m[j]}\right)$$

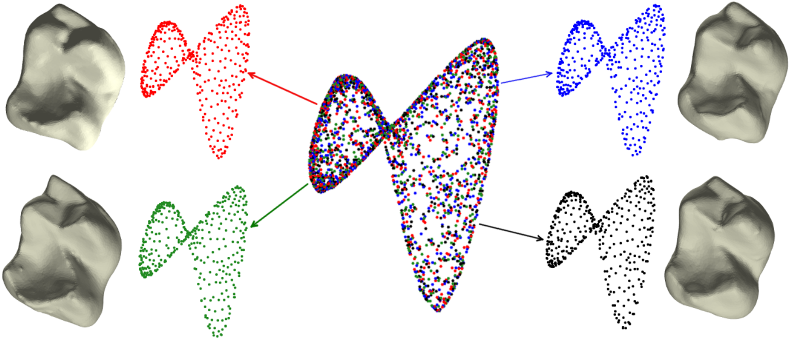
Horizontal Diffusion Maps



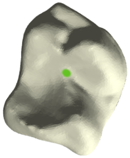
HDM: Application in Geometric Morphometrics

1. Global Registration
2. Automatic Landmarking
3. Species Classification

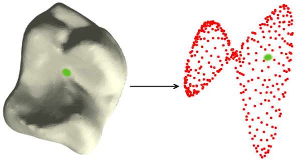
1. Global Registration



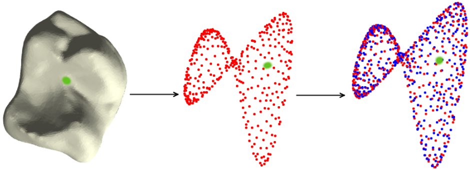
1. Global Registration



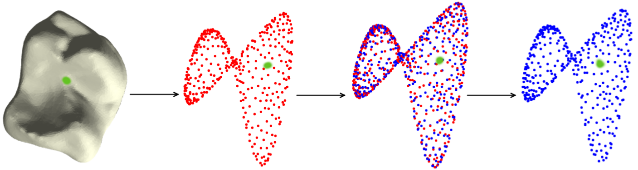
1. Global Registration



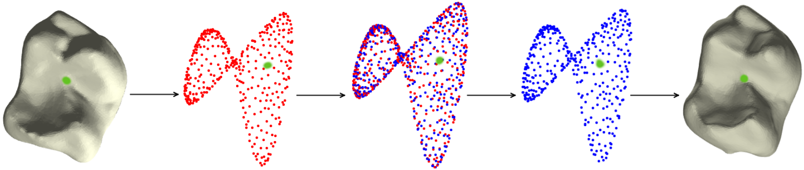
1. Global Registration



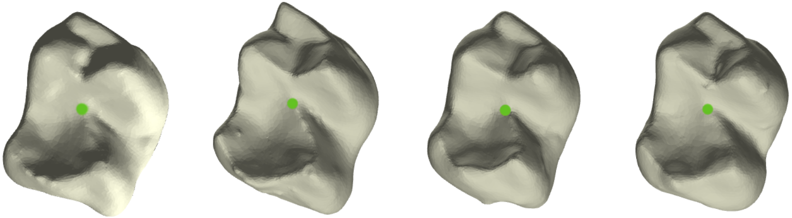
1. Global Registration



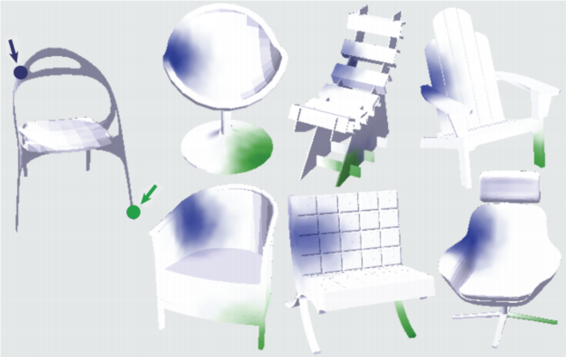
1. Global Registration



1. Global Registration

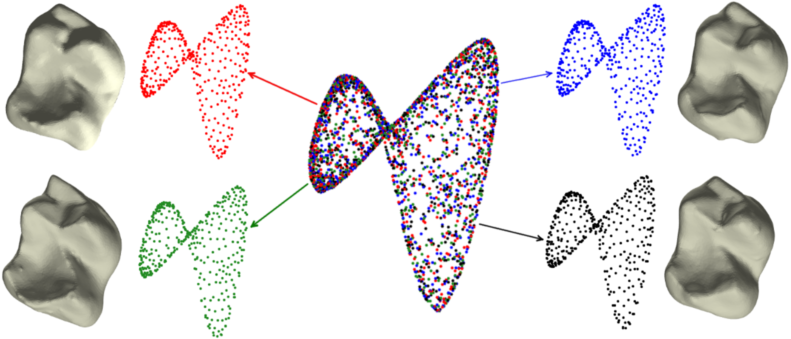


1. Global Registration: A Tool of Visual Exploration

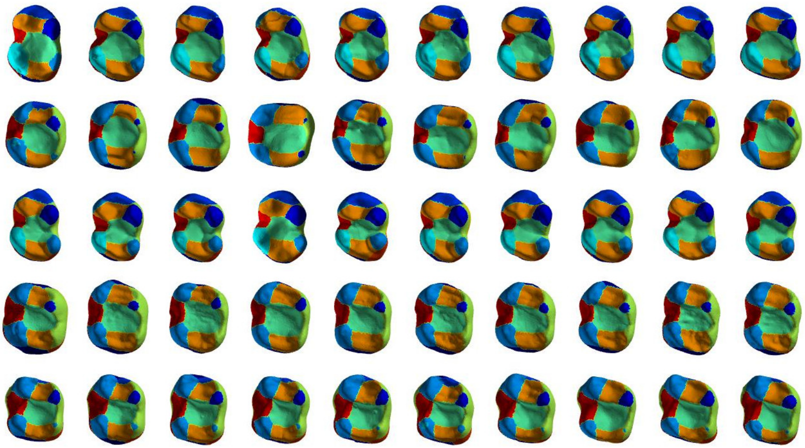


Kim et al. *Exploring Collections of 3D Models using Fuzzy Correspondences.* SIGGRAPH 2012.

2. Automatic Landmarking: *Spectral Clustering*



2. Automatic Landmarking: *Spectral Clustering*



3. Species Classification: HBDM & HBDD

Horizontal Diffusion Maps (HDM): For fixed $1 \leq m \leq \kappa$, $t \geq 0$, represent S_j as a $\kappa_j \times m$ *matrix*

$$\left(\lambda_1^{t/2} u_{1[j]}, \cdots, \lambda_m^{t/2} u_{m[j]}\right)$$

Horizontal Base Diffusion Maps (HBDM): For fixed $1 \leq m \leq \kappa$, $t \geq 0$, represent S_j as a $\binom{m}{2}$-dimensional *vector*

$$\left(\lambda_\ell^{t/2} \lambda_k^{t/2} \left\langle u_{\ell[j]}, u_{k[j]} \right\rangle \right)_{1 \leq \ell < k \leq m}$$

3. Species Classification: HBDM & HBDD

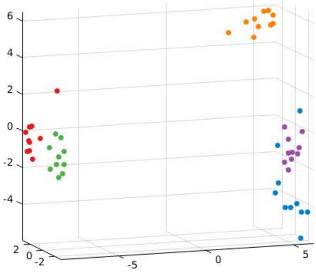
Horizontal Base Diffusion Distance (HBDD): For fixed $1 \leq m \leq \kappa$, $t \geq 0$,

$$D^t_{HB}(S_i, S_j) = \left(\sum_{1 \leq \ell < k \leq m} \lambda^t_\ell \lambda^t_k \left(\left\langle u_{\ell[i]}, u_{k[i]} \right\rangle - \left\langle u_{\ell[j]}, u_{k[j]} \right\rangle \right)^2 \right)^{\frac{1}{2}}$$

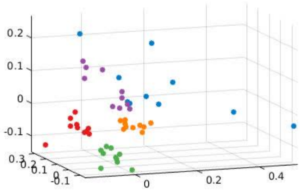
Horizontal Base Diffusion Maps (HBDM): For fixed $1 \leq m \leq \kappa$, $t \geq 0$, represent S_j as a $\binom{m}{2}$-dimensional *vector*

$$\left(\lambda^{t/2}_\ell \lambda^{t/2}_k \left\langle u_{\ell[j]}, u_{k[j]} \right\rangle \right)_{1 \leq \ell < k \leq m}$$

3. Species Classification: HBDM & HBDD

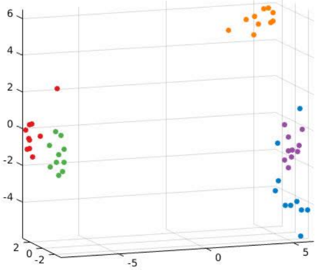


HBDD

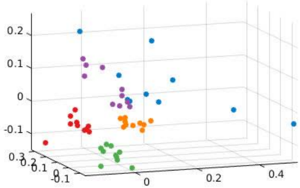


DD

3. Species Classification: HBDM & HBDD

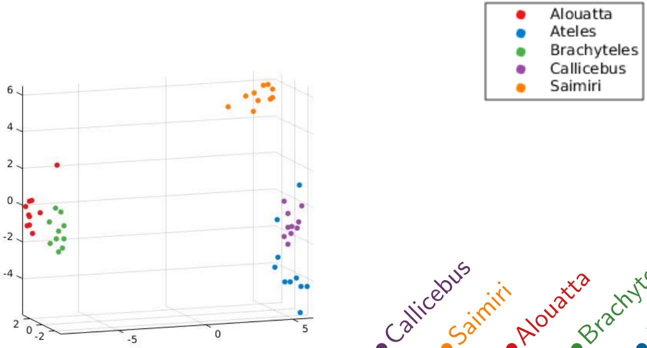


HBDD

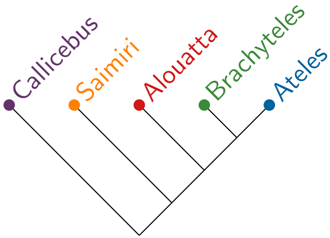


DD

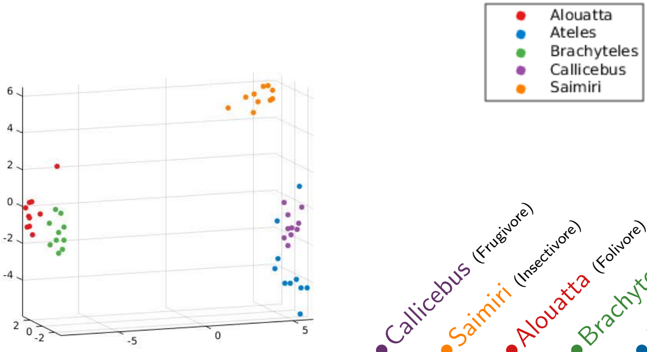
3. Species Classification: HBDM & HBDD



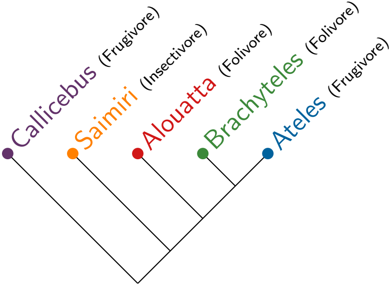
HBDD



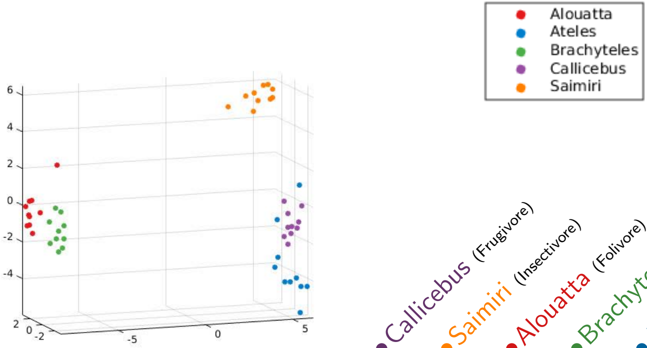
3. Species Classification: HBDM & HBDD



HBDD

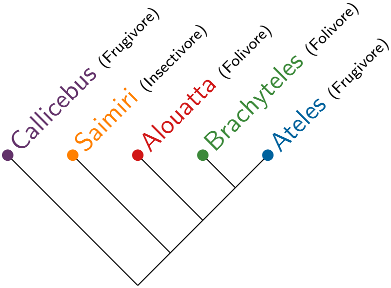


3. Species Classification: HBDM & HBDD



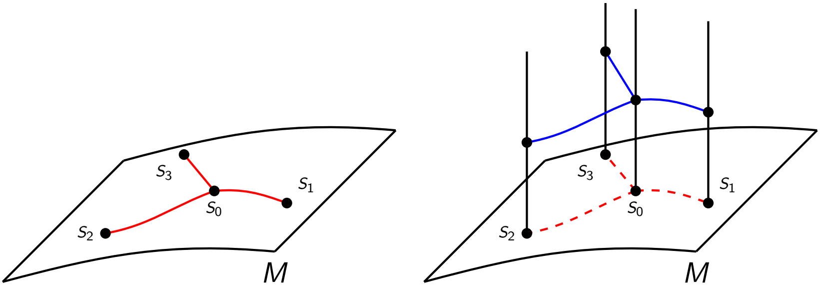
HBDD

"Form Follows Function"



HDM: Horizontal Random Walk on a *Fibre Bundle*

$$P_{\epsilon}^{(\alpha)} = \left(D_{\epsilon}^{(\alpha)}\right)^{-1} W_{\epsilon}^{(\alpha)} \qquad H_{\epsilon,\delta}^{(\alpha)} = \left(\mathscr{D}_{\epsilon,\delta}^{(\alpha)}\right)^{-1} \mathscr{W}_{\epsilon,\delta}^{(\alpha)}$$



Asymptotic Theory for Diffusion Maps

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 in probability

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

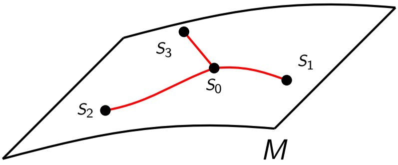
Theorem (Coifman-Lafon 2006). As $\epsilon \to 0$, for any $f \in C^\infty(M)$ and $x \in M$,

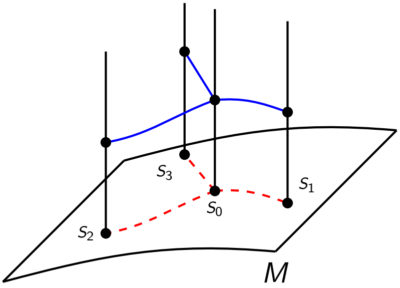
$$\begin{aligned} & P_\epsilon^{(\alpha)} f(x) \\ & = f(x) + \epsilon \frac{m_2}{2m_0} \left[\frac{\Delta_M \left[f p^{1-\alpha} \right](x)}{p^{1-\alpha}(x)} - f(x) \frac{\Delta_M p^{1-\alpha}(x)}{p^{1-\alpha}(x)} \right] + O\left(\epsilon^2\right). \end{aligned}$$

HDM: Horizontal Random Walk on a *Fibre Bundle*

$$P_{\epsilon}^{(\alpha)} = \left(D_{\epsilon}^{(\alpha)}\right)^{-1} W_{\epsilon}^{(\alpha)}$$

$$H_{\epsilon,\delta}^{(\alpha)} = \left(\mathscr{D}_{\epsilon,\delta}^{(\alpha)}\right)^{-1} \mathscr{W}_{\epsilon,\delta}^{(\alpha)}$$





Asymptotic Theory for HDM on (E, M, F, π)

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.* arXiv:1602.02330, 2016

Asymptotic Theory for HDM on (E, M, F, π)

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 $\epsilon \to 0$,

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

- 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*

HDM on Unit Tangent Bundles ($\pi : UTM \to M$)

Corollary. If $\alpha = 1$, then

(i) If $\delta = O(\epsilon)$ as $\epsilon \to 0$, then for any $f \in C^\infty(UTM)$, as $\epsilon \to 0$,

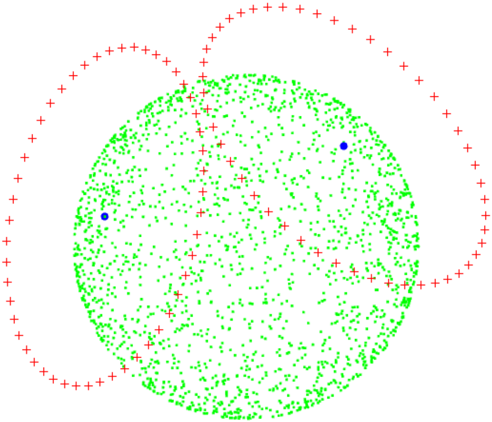
$$H^1_{\epsilon,\delta} f(x,v) = f(x,v) + \epsilon \frac{m_{21}}{2m_0} \Delta^H_{UTM} f(x,v) + \delta \frac{m_{22}}{2m_0} \Delta^V_{UTM} f(x,v) + O\left(\epsilon^2 + \epsilon\delta + \delta^2\right);$$

(ii) If $\delta = \gamma\epsilon$, then for any $f \in C^\infty(UTM)$,

$$\lim_{\gamma\to\infty} H^1_{\epsilon,\gamma\epsilon} f(x,v) = \langle f \rangle(x) + \epsilon \frac{m_2'}{2m_0'} \Delta_M \langle f \rangle(x) + O\left(\epsilon^2\right),$$

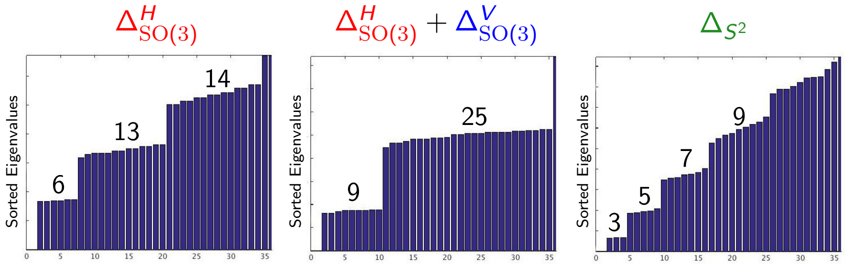
where $\langle f \rangle \in C^\infty(M)$ is the *average of f along the fibres*.

HDM on Unit Tangent Bundles: Validation on $\mathrm{SO}(3)$



$\mathrm{SO}(3)$ as the unit tangent bundle of $S^2 \subset \mathbb{R}^3$

HDM on Unit Tangent Bundles: Validation on $\mathrm{SO}(3)$



(a) $\delta = 0.002$ (b) $\delta = 0.015$ (c) $\delta = 20$

Bar plots of the smallest 36 eigenvalues of *horizontal*, *total*, and *base* Laplacians on $\mathrm{SO}(3)$, with fixed $\epsilon = 0.2$ and varying δ

Tingran Gao. *The Diffusion Geometry of Fibre Bundles.* arXiv:1602.02330, 2016

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 $H^{\alpha}_{\epsilon,\delta}$ converges to its smooth limit at rate

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

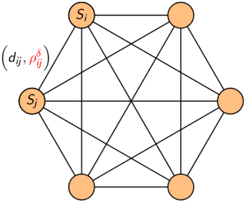
where

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

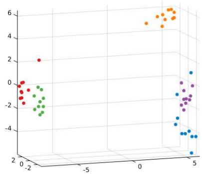
Tingran Gao. *The Diffusion Geometry of Fibre Bundles.* arXiv:1602.02330, 2016

Summary and Future Work

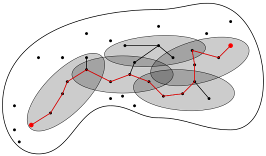
Augmented Distance Graph



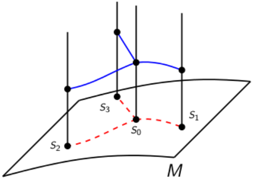
Horizontal Base Diffusion Distance



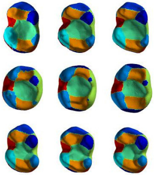
"Knit Together" Local Geometry



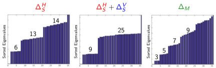
Random Walk on a Fibre Bundle



Biological Application



Mathematical Theory

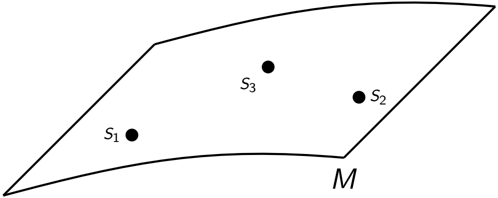


$$O\left(\theta_*^{-1} N_B^{-\frac{1}{2}} \epsilon^{-\frac{d}{4}}\right),$$

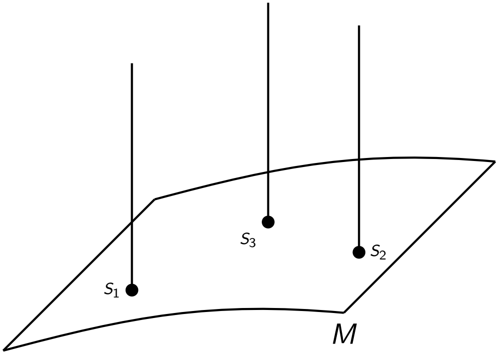
$$\theta_* = 1 - \frac{1}{1 + \epsilon^{\frac{d}{4}} \delta^{\frac{d-1}{4}} \sqrt{\frac{N_F}{N_B}}}.$$

Tingran Gao. *The Diffusion Geometry of Fibre Bundles.* arXiv:1602.02330, 2016

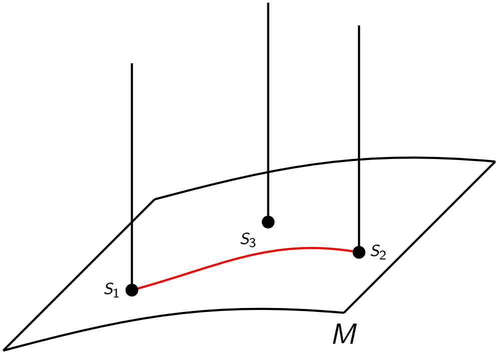
HDM: Random Walk on a *Fibre Bundle*



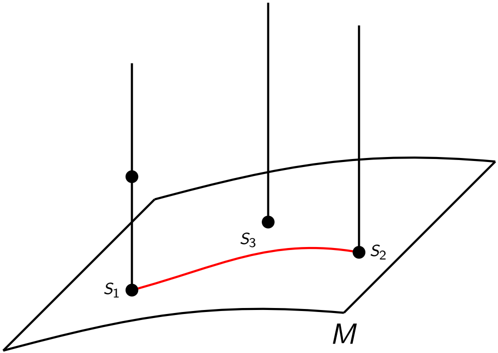
HDM: Random Walk on a *Fibre Bundle*



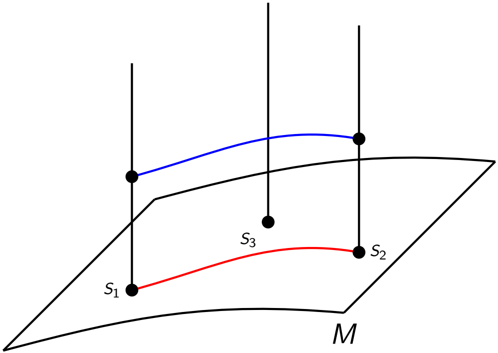
HDM – Parallel Transport



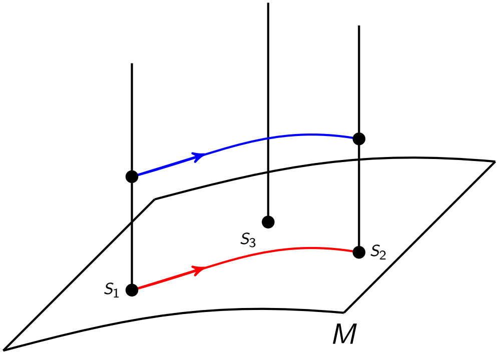
HDM – Parallel Transport



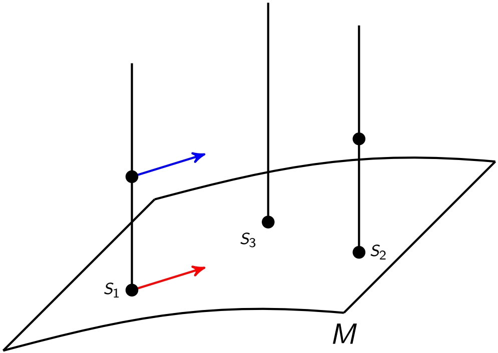
HDM – Parallel Transport



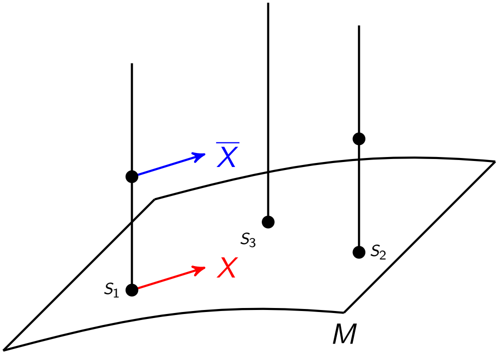
HDM – Connection



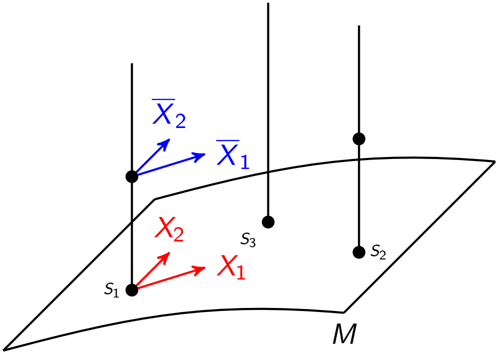
HDM – Connection



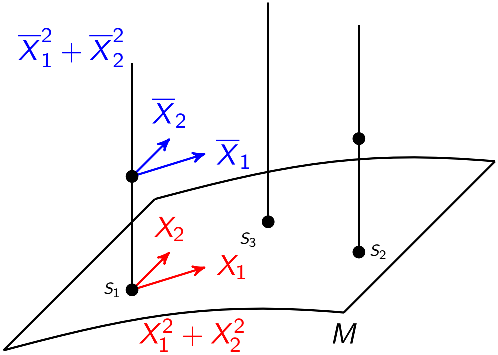
HDM – Connection



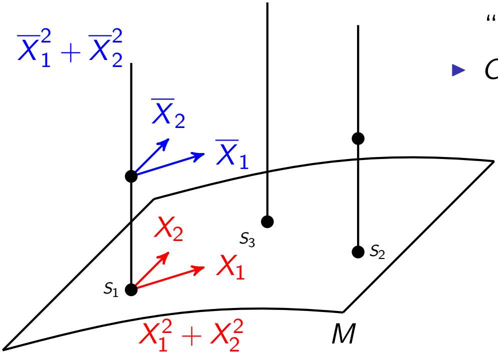
HDM – Connections



HDM – Hypoellipticity



HDM – Hypoellipticity

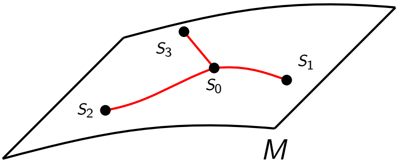


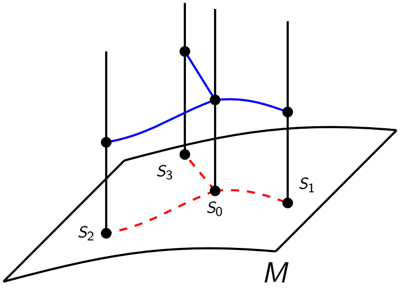
- *Hörmander's Condition* or "Bracket Generating"
- *Chow-Rashevskii Theorem*

HDM: Random Walk on a *Fibre Bundle*

$$P_{\epsilon,\alpha} = D_{\epsilon,\alpha}^{-1} W_{\epsilon,\alpha}$$

$$H_{\epsilon,\delta}^{(\alpha)} = \left(\mathscr{D}_{\epsilon,\delta}^{(\alpha)}\right)^{-1} \mathscr{W}_{\epsilon,\delta}^{(\alpha)}$$





Asymptotic Theory for HDM on (E, M, F, π)

Definition (G. 2015). For any $f \in C^{\infty}(E)$, $(x, v) \in E$,

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

where

$$K_{\epsilon,\sigma}^{(\alpha)}(x,v;y,w) = \frac{K_{\epsilon,\delta}(x,v;y,w)}{p_{\epsilon,\delta}^{\alpha}(x,v)\, p_{\epsilon,\delta}^{\alpha}(y,w)},$$

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