Principal Sub-manifolds

Zhigang Yao

Department of Statistics and Applied Probability
National University of Singapore

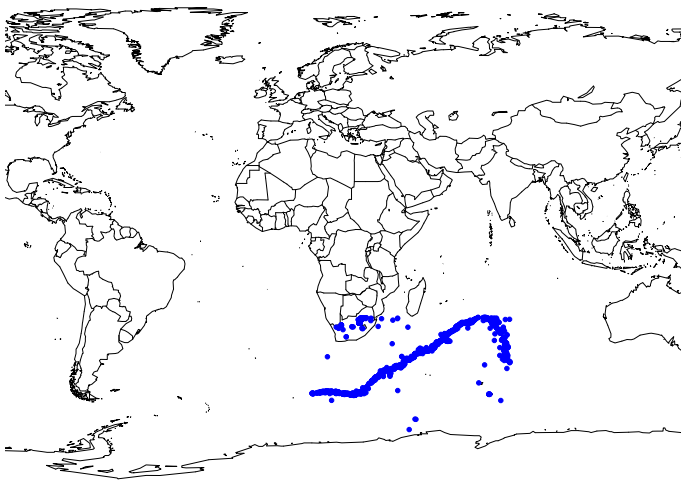
Collaborator: Tung Pham (U of Melbourne)

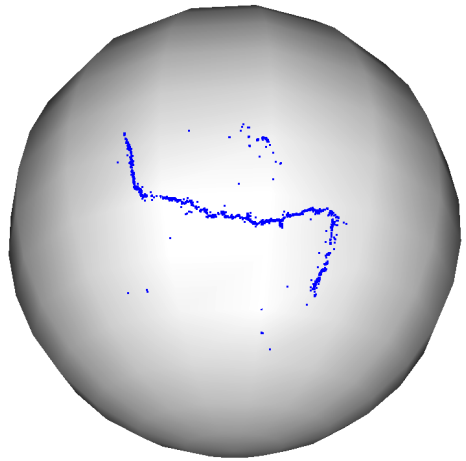
Data on Manifolds

Data on manifolds[1] may arise in (at least) two ways:

(1) Manifold is actual physical space where data reside

↪ Usually sphere; from geophysics to marine biology





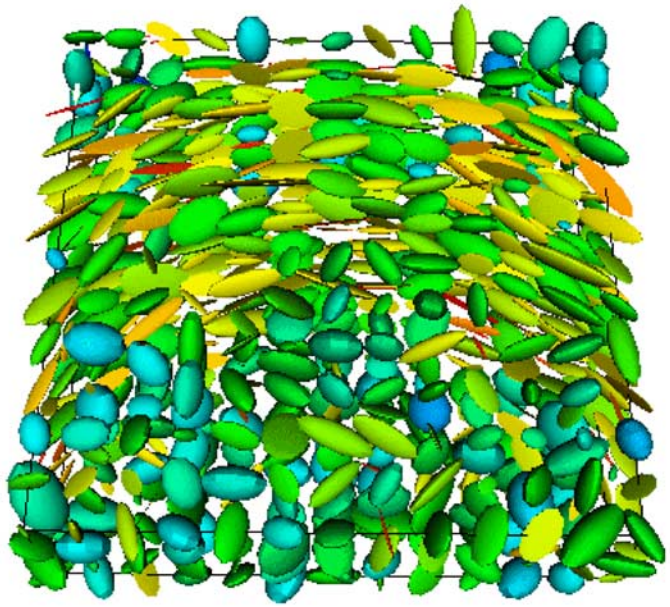
[1] contrast to manifolds in data analysis, or manifold-valued data

Data on Manifolds

Data on manifolds may arise in (at least) two ways:

(2) Multivariate data under non-linear constraints, thus being forced onto manifold

↪ e.g. cones for positive-def matrices or Stiefel manifolds for ordered bases



Data on Manifolds

$$D = \begin{pmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{pmatrix} : v^{\mathsf{T}} D v > 0, v \in \mathbb{R}^3$$

- let $\mathcal{P}_3$ be a space of all symmetric positive definite 3 by 3 matrices
- $\mathcal{P}_3 \subset \mathbb{R}^6$
- $\mathcal{P}_3$ is convex but not linear in $\mathbb{R}^6$: $\frac{3}{2} D_1 - \frac{1}{2} D_2$ might not in $\mathcal{P}_3$
- Its actual space is a cone in $\mathbb{R}^4$

Data on Manifolds

A k-ad (k landmarks)

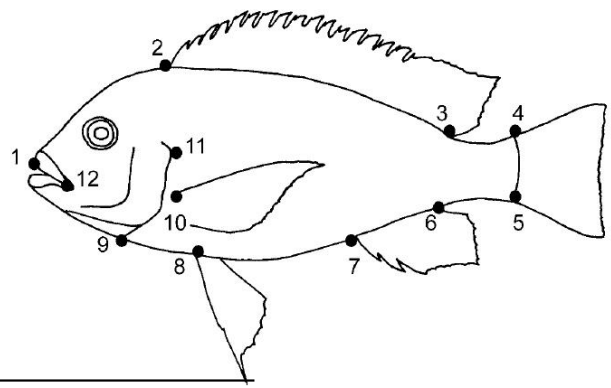
$$\mathbf{z} = \{(x_j, y_j), 1 \leq j \leq k\}$$

Kendall's shape space Σ_2^k

$$\mathbf{z} - \langle \mathbf{z} \rangle \in \mathbb{R}^{2k-2} = \{(x_j, y_j)_{1 \leq j \leq k} : \Sigma x_j = 0, \Sigma y_j = 0\}$$

$$preshape : \mathbf{w} = \frac{\mathbf{z} - \langle \mathbf{z} \rangle}{|\mathbf{z} - \langle \mathbf{z} \rangle|} \in S^{2k-3} \subset \mathbb{R}^{2k-2} (\text{unit sphere}^2)$$

$$shape : e^{i\theta}\mathbf{w} \in S^{2k-3}/S^1$$



[2] $S^n = \left\{x \in \mathbb{R}^{n+1} : |x| = 1\right\}$

Data on Manifolds: Some Challenges

Unexpected challenges: statistics more rooted in linearity than may originally meet the eye.

Even the simplest statistical notion (an average) is a highly non-trivial object.

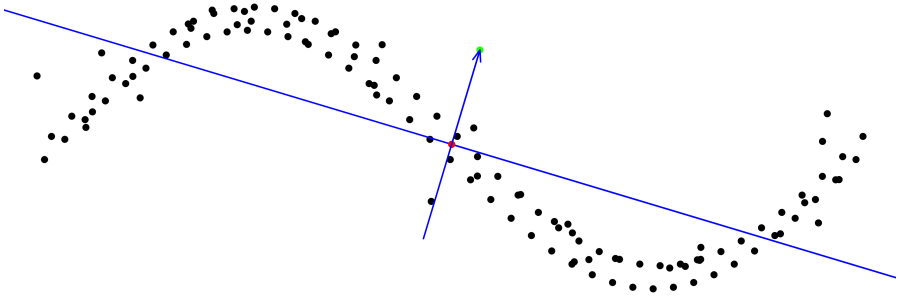
- May admit several "canonical" definitions, some of which do not guarantee existence, others may not guarantee uniqueness

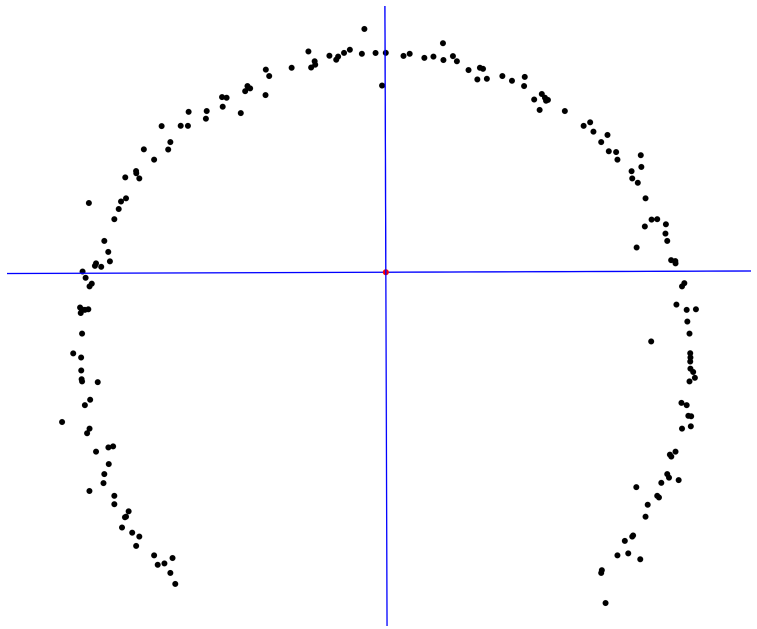
e.g. Fréchet mean μ for random variable X on $\mathcal{M}$

$$\mathbb{E}d^2(\mu, X) \leq \mathbb{E}d^2(x, X) \qquad \forall x \in \mathcal{M}.$$

Two Toy Data Structures

"S-shape" and "C-shape":





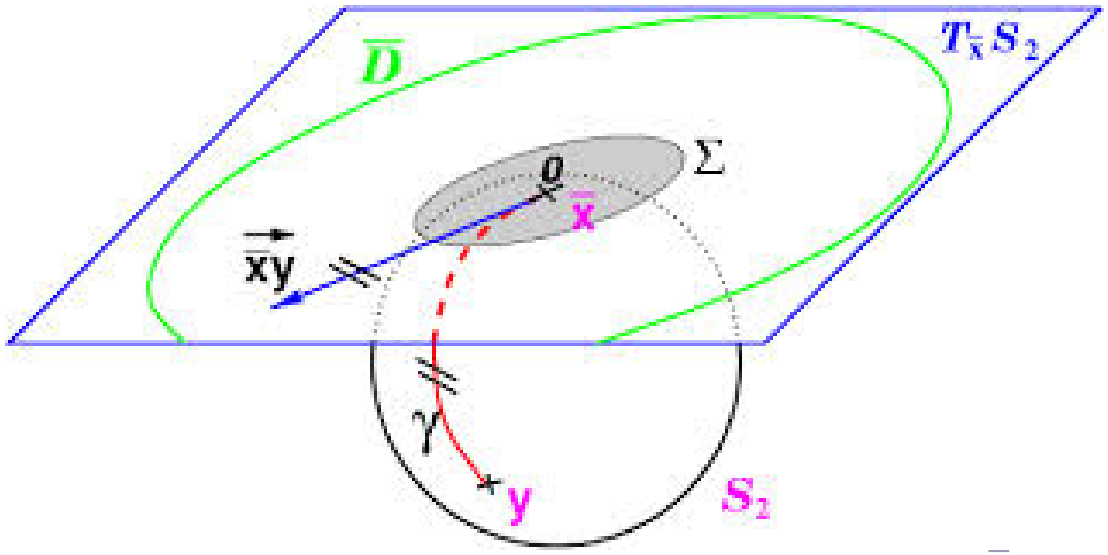
exp and log map

Exponential map: tangent space $\longmapsto$ manifold

$$\text{ie } \exp_{\bar{x}}(\vec{\bar{x}y}) = y$$

Log map: manifold $\longmapsto$ tangent space

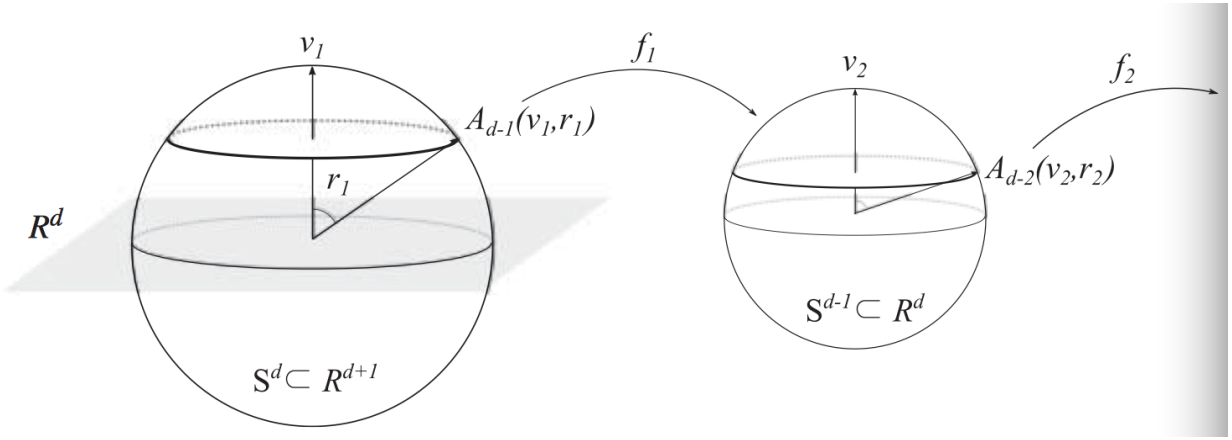
$$\text{ie } \log_{\bar{x}}(y) = \vec{\bar{x}y}$$



1. Principal nested spheres (spherical case) paralleling nested subspaces (Jung/Dryden/Marron, 2012)

$$\mathbb{U}_1 \subset \mathbb{U}_2 \subset \cdots \mathbb{U}_{d-1} \subset \mathbb{U}_d = \mathbb{S}^d$$

$$\mathbb{U}_k = \arg\min_{\mathbb{S}^k \subset \mathbb{U}_{k+1}} \sum_{i=1}^{n} d_m^2(x_i, \mathbb{S}^k)$$



A view on Tangent PCA[3]

1. Idea: local tangent space PCA at every point yields direction of local maximal variation at that point
   - For any $x \in \mathcal{M}$, take tangent space T_x, 'lift' data, take

$$\Sigma_h(\bar{x}) = \frac{1}{\sum_i \kappa_h(x_i, \bar{x})} \sum_i \left(\log_{\bar{x}}(x_i) \otimes \log_{\bar{x}}(x_i) \right) \kappa_h(x_i, \bar{x}) .$$

   - Consider (orientation) field $\{\lambda_1(\bar{x}) e_1(\bar{x})\}_{\bar{x} \in \mathcal{M}}$ of 1st eigenvectors times 1st eigenvalues of $\Sigma_h(\bar{x})$.
   - Aim to capture non-geodesic variation
   - Can make this well-defined and smooth ("same direction")
2. Question: starting at a point (e.g. Fréchet mean) is there an integral curve of 'maximal variation' along this field?

[3] (will not discuss regularity conditions)

Vector Field

- Given a staring point $\bar{x}$, transform the orientation field

$$V(\bar{x}) = \{-\lambda_1(\bar{x})e_1(\bar{x}), \lambda_1(\bar{x})e_1(\bar{x})\}$$

from $\Sigma_h(x)$ into a vector field W such that $\forall x \in N(\bar{x})$

$$\Sigma_h(x)\,W(x) = \lambda_1(x)\,W(x) \;\; (\text{i.e.}, W(x) \in V(x))$$

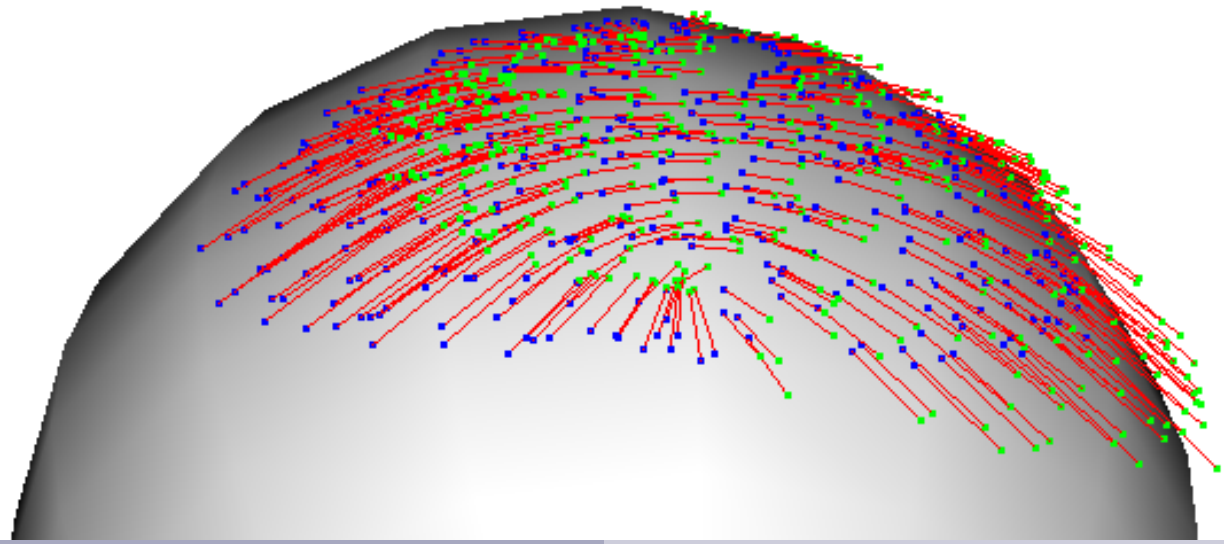
- At least locally, within an open neighborhood N, we can pick the eigenvectors $\{e_1(x) : x \in N\}$ to be *pointing in the same direction*

A Principal Flow?

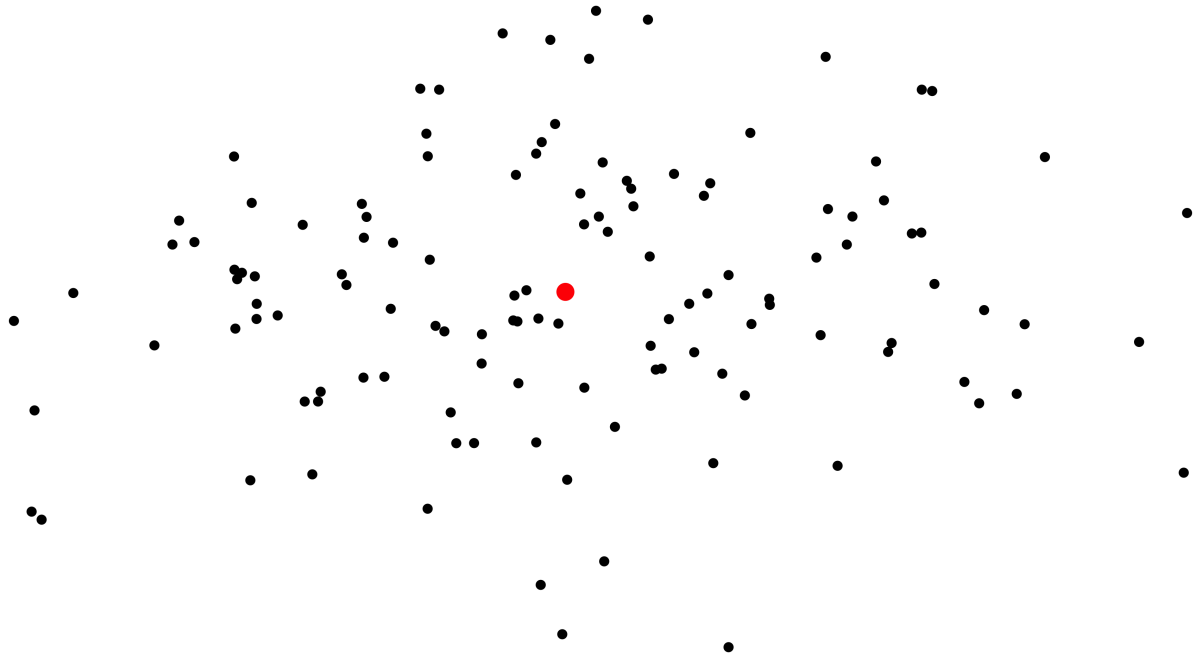
What does "*curve of maximal variation*" mean?

Would like a reasonably smooth curve $\gamma(x)$ whose derivative $\dot{\gamma}(x) \in T_x M$ at any point $x \in M$ is $\approx$ tangent to $\lambda_1(x) e_1(x)$

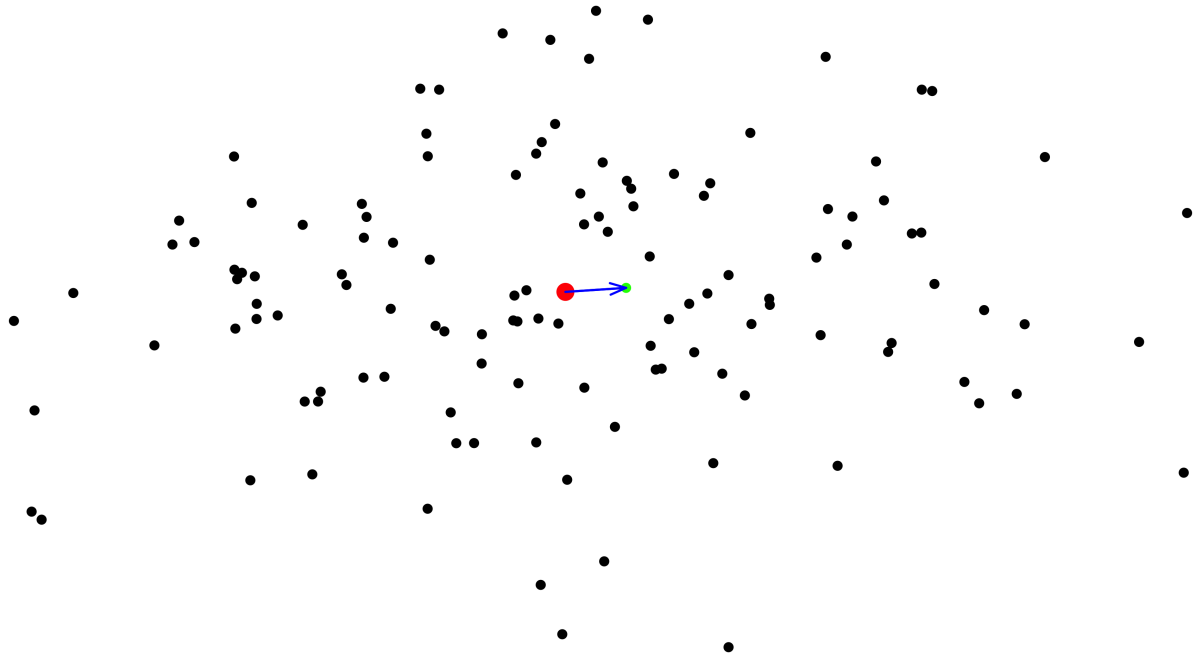
...AND maximizes the work done by the field on a particle traveling along its path



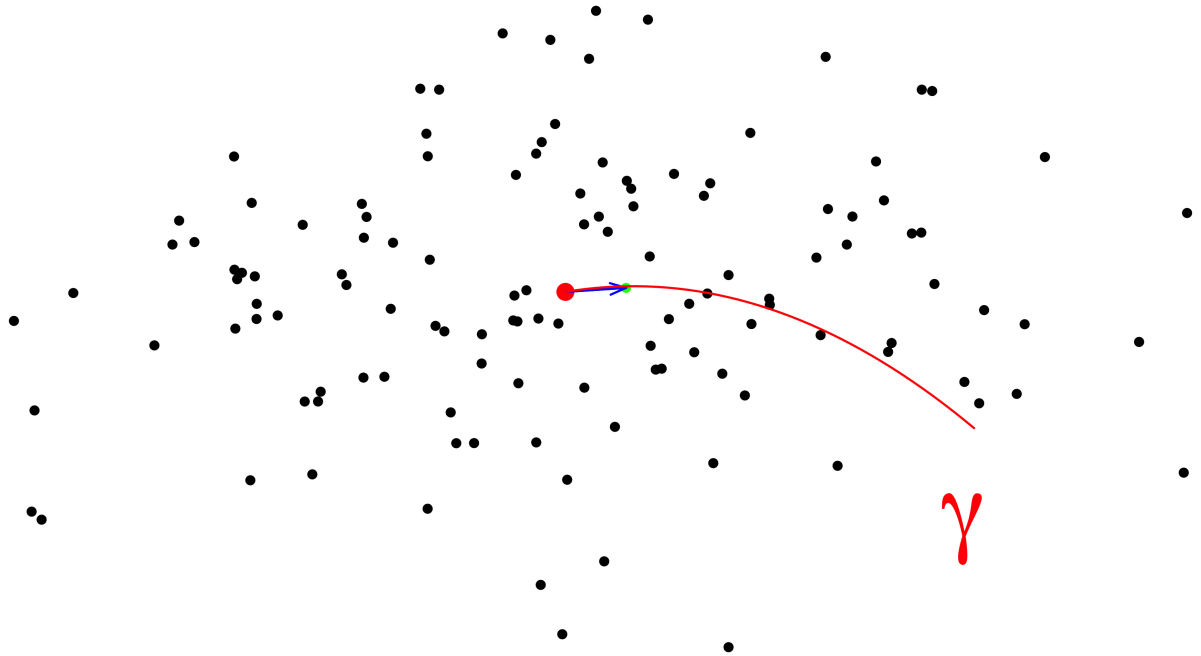
Illustration



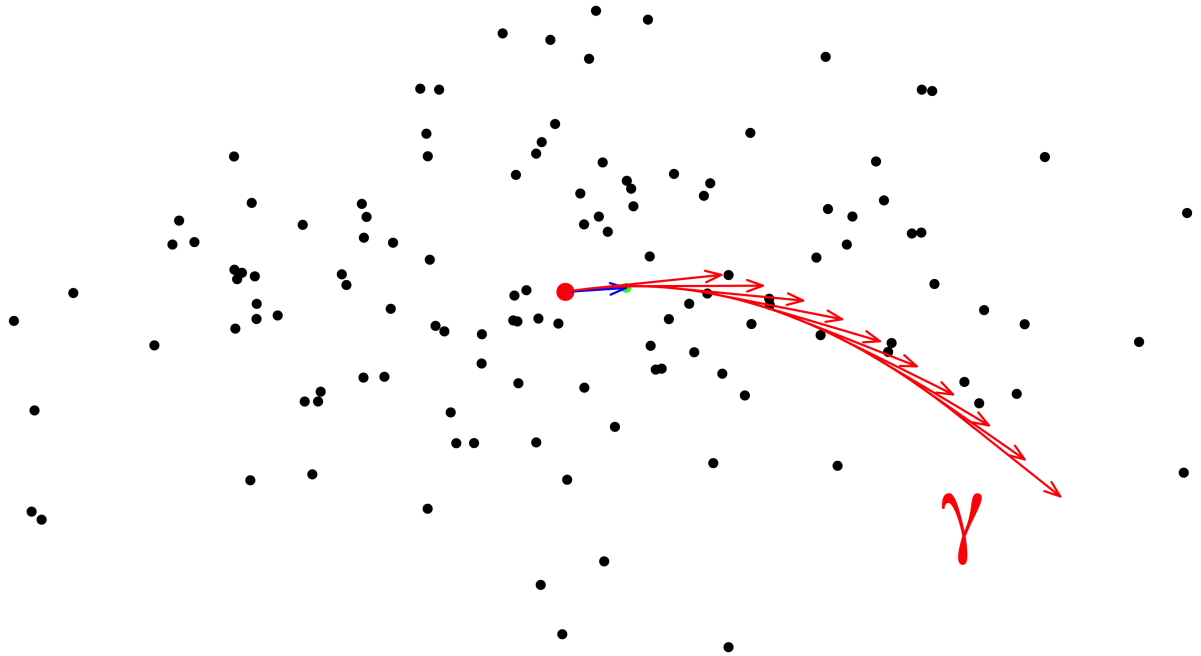
Illustration



Illustration



Illustration



A Principal Flow (Panaretos/Pham/Yao *JASA*, 2014)

(mod technicalities)[4] Curve with midpoint $\bar{x}$, maximizing

$$\int \left| \left\langle \dot{\gamma}(t), W(\gamma(t)) \right\rangle \right| dt$$

$$\text{SubM}(A, v, \mathcal{M}) = \Big\{ \gamma : [0, r] \to \mathcal{M}, \gamma \in C^2(\mathcal{M}), \gamma(s) \neq \gamma(s') \text{ for } s \neq s',$$
$$\gamma(0) = A, \dot{\gamma}(0) = v, \ell(\gamma[0, t]) = t \text{ for all } 0 \leq t \leq r \leq 1 \Big\}.$$

- Answer: yes, reformulate to Euler-Lagrange equations
- $\exists$ unique solution under mild conditions on manifold+field
- Requires geodesics and second fundamental tensor
- Numerically Feasible for many "standard" manifolds
- Canonical: reduces to ordinary PCA in Euclidean spaces

[4] (several technical issues will not be discussed)

Principal Curve (Hastie/Stuetzle (1989)) and Examples [5]

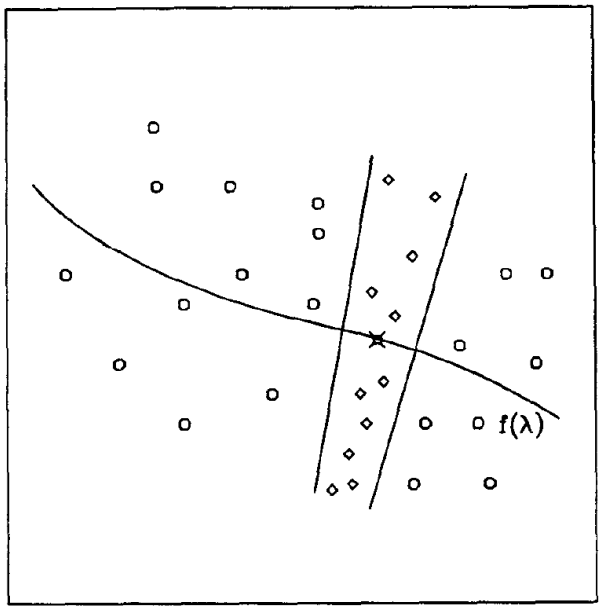
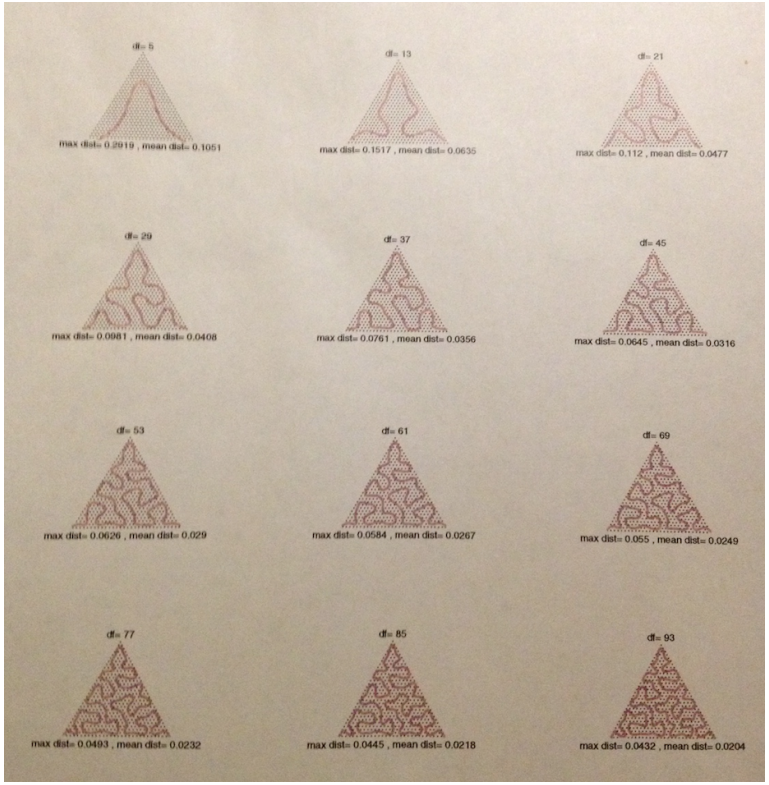
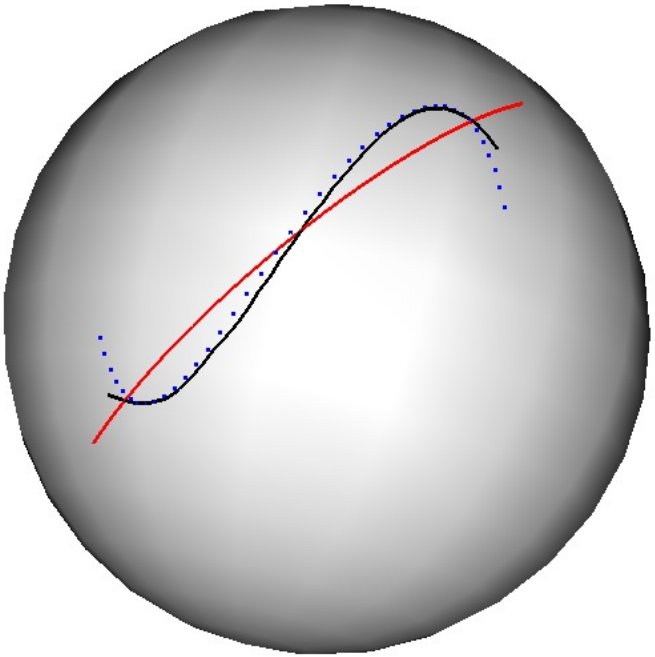


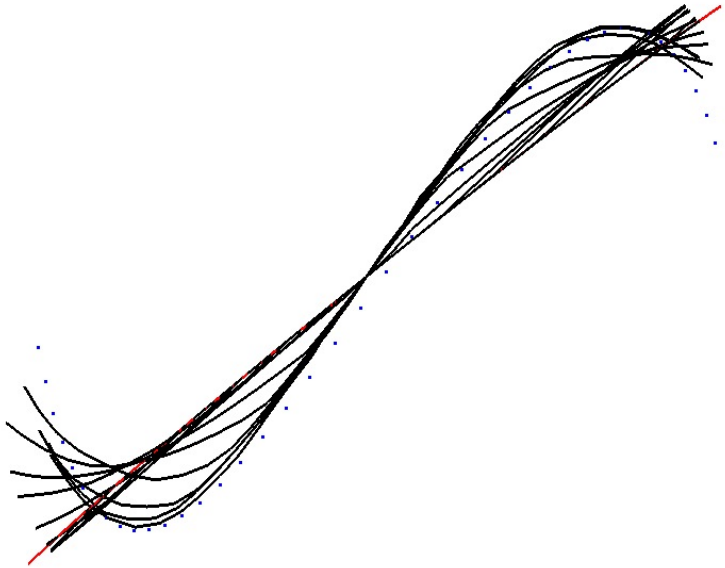
Figure (3.2) Each point on a principal curve is the average of the points that project there.



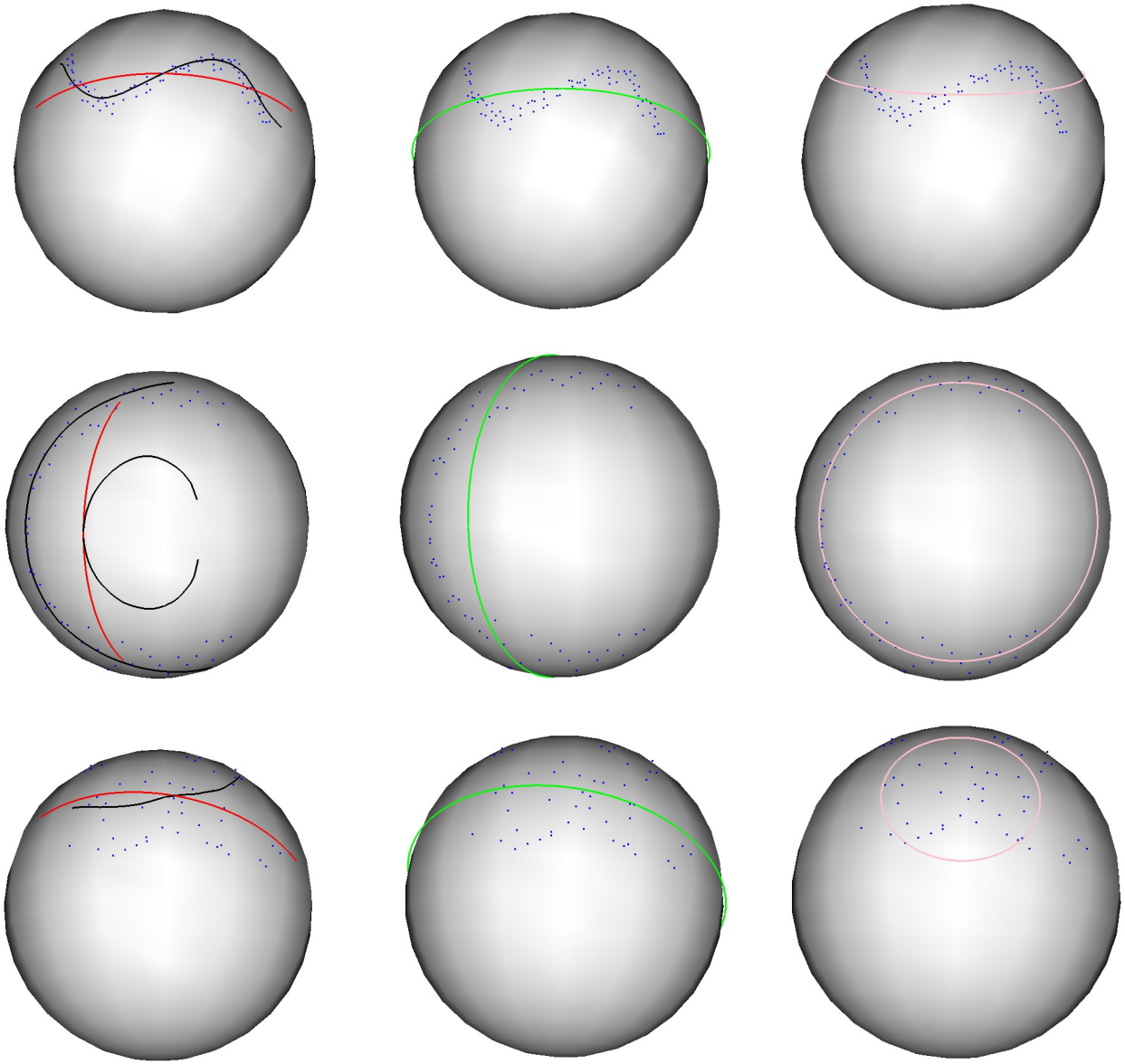
[5] Thanks to Trevor Hastie for sharing the exaples)

Quick Illustration (varying scale parameter h)





Simulation-Sphere (S-shape, C-shape, Diffusion)



Some notations

- For any point x in $\mathcal{M}$ and $\mathcal{N}$,

$$\Sigma_{x,\mathcal{M}} = \frac{1}{n}\sum_{i=1}^{n} \log_{x,\mathcal{M}}(x_i) \otimes \log_{x,\mathcal{M}}(x_i) \text{ and}$$

$$\Sigma_{x,\mathcal{N}} = \frac{1}{n}\sum_{i=1}^{n} \log_{x,\mathcal{N}}(x_i) \otimes \log_{x,\mathcal{N}}(x_i).$$

- Let $\{e_1(x,\mathcal{M}),\ldots,e_k(x,\mathcal{M})\}$ be the first k eigenvectors of $\Sigma_{x,\mathcal{M}}$. k eigenvectors $\{e_1(x,\mathcal{N}),\ldots,e_k(x,\mathcal{N})\}$ for $\Sigma_{x,\mathcal{N}}$.
- Let $H_k(x,\mathcal{M})$ be the hyperplane on $\mathcal{M}$ spanned by $\{e_1(x,\mathcal{M}),\ldots,e_k(x,\mathcal{M})\}$; $H_k(x,\mathcal{N})$ be the hyperplane on $\mathcal{N}$ spanned by$\{e_1(x,\mathcal{N}),\ldots,e_k(x,\mathcal{N})\}$.

Some notations-cont'd

- $$\mathbf{B}(x, \mathcal{N}, \epsilon) = \Big\{ y \in \mathcal{N} : d_{\mathcal{N}}(x, y) \leq \epsilon \Big\}.$$

  where $d_{\mathcal{N}}(x, y)$ is the distance of x and y on $\mathcal{N}$.
- For any positive integer number $k < m$, and any point $x \in \mathcal{M}$, let $\mathrm{SubM}(x, \epsilon, k, \mathcal{M})$ be the set of all k-dimensional sub-manifolds of $\mathbf{B}(x, \mathcal{M}, \epsilon)$.
- For a given k, the main idea of the principal sub-manifold is: at each point B of the sub-manifold $\mathcal{N}$, the sub-manifold should be able to explain the manifold variation as much as possible.

An ideal principal sub-manifold (Yao/Pham, 2016)

(Ideal principal sub-manifold)[6] k-th dimensional principal sub-manifold

$$\arg \sup_{\mathcal{N} \in \mathrm{SubM}\left(A, \epsilon, k, \mathcal{M}\right)} \int_{B \in \mathcal{N}} \left(\cos(\alpha_B) \times \sum_{j=1}^{k} \lambda_j(B, \mathcal{M}) \right) d\mu_{\mathcal{N}},$$

where $\mu_{\mathcal{N}}$ is the measure on $\mathcal{N}$

- To measure the degree of variation, we use the angle α_B between the two hyperplanes, $H_k(B, \mathcal{M})$ and $H_k(B, \mathcal{N})$.
- Theoretically, if $\alpha_B = 0$ for every B, then $H_k(B, \mathcal{M}) = H_k(B, \mathcal{N})$. For general cases, one would hope α_B is as small as possible.

[6] (subject to modification)

How to get a concrete sub-manifold

- The idea is to apply the mapping to map $\mathcal{N}$ into a ball of radius ϵ in its tangent space, and then we can use the polar coordinates of that ball in the tangent space.
- Denote $L(\mathcal{N}, \epsilon) = \mathbf{log}_p(\mathcal{N})$ to be the image of the sub-manifold $\mathcal{N}$ at p under the logarithm map.

Principal sub-manifold (Yao/Pham, 2016)

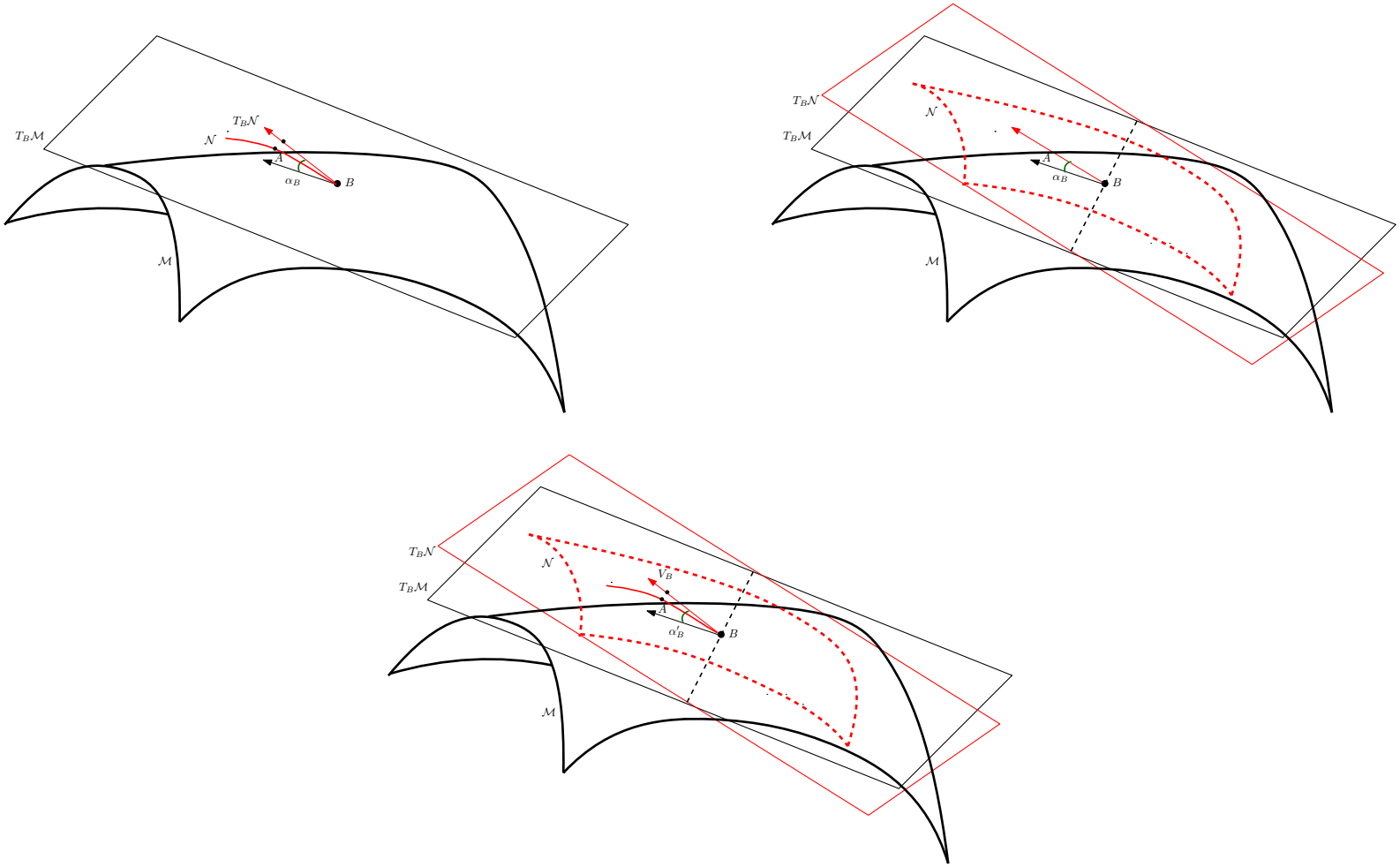
(principal sub-manifoly, slightly changed) k-th dimensional principal sub-manifold

$$\arg \sup_{\mathcal{N} \in \mathrm{SubM}\left(A,\epsilon,k,\mathcal{M}\right)} \int_{\log_A(B) \in \log_A(\mathcal{N})} \left(\cos(\alpha'_B) \times \sum_{j=1}^{k} \lambda_j(B, \mathcal{M}) \right) d\mu_k,$$

where μ_k is the measure on the ball of the k-dimensional space of radius ϵ.

- Let v_B be the tangent vector field of a geodesic curve on the sub-manifold $\mathcal{N}$ from A to B, denote α'_B to be the angle between v_B and hyperplane $H_k(B, \mathcal{M})$
- α'_B is chosen to replace α_B: this modification not only measures how $\mathcal{N}$ differs from $\mathcal{M}$ at B, but also provides a convenient construction of a surface of maximal "cumulative variation."

Principal flow and principal submanifold



Principal sub-manifold in Euclidean Space

Theorem

Assume that $\mathcal{M} = \mathbb{R}^d$ then

$$\arg \sup_{\mathcal{N} \in \mathtt{SubM}\left(A, \epsilon, k, \mathcal{M}\right)} \int_{\log_A(B) \in \log_A(\mathcal{N})} \left(\cos(\alpha_B) \times \sum_{j=1}^{k} \lambda_j(B, \mathcal{M}) \right) d\mu_k$$

$$= \text{Hyperplane spanned by } \{e_1(A, \mathcal{M}), e_2(A, \mathcal{M}), \ldots, e_k(A, \mathcal{M})\}.$$

Algorithm 1: two-dimensional principal sub-manifold

1. At a point A (mean), use the log map: $\mathbf{log}_A(x_i) = y_i$.
2. Find the covariance matrix from $y_1, \ldots, y_n$

$$\Sigma_A = (y_i - A)^T \times (y_i - A).$$

3. Let $e_1(A)$ and $e_2(A)$ be the first and second eigenvectors of Σ_A. Define

$$Z_l = \epsilon \times \left[\cos(2l\pi/180)\, e_1(A) + \sin(2l\pi/180)\, e_2(A)\right],$$

with $l = 1, \ldots, 180$.

4. Use exponential map to map Z_l on the manifold so we get a set of new points $\mathbf{exp}_A(Z_l) = A_l$.
5. Assume that we stay at point $A_{l,i}$, we are going to find $A_{l,i+1}$ ($A_{l,0} = A$ and $A_{l,1} = A_l$)
   - 5.1 Find $\Sigma_{A_{l,i}}$.
   - 5.2 Find $e_1(A_{l,i})$ and $e_2(A_{l,i})$.
   - 5.3 Find $\mathbf{log}_{A_{l,i}}(A_{l,i-1}) = v_{l,i}$.
   - 5.4 Find

$$u_{l,i} = \left\langle v_{l,i}, e_1(A_{l,i}) \right\rangle \times e_1(A_{l,i}) + \left\langle v_{l,i}, e_2(A_{l,i}) \right\rangle \times e_2(A_{l,i})$$

   where $\langle a, b \rangle = \sum_{i=1}^n a_i b_i$ with $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$.
   - 5.5 Find

$$r_{l,i} = \epsilon \times \frac{u_{l,i}}{\|u_{l,i}\|}.$$

   - 5.6 Then

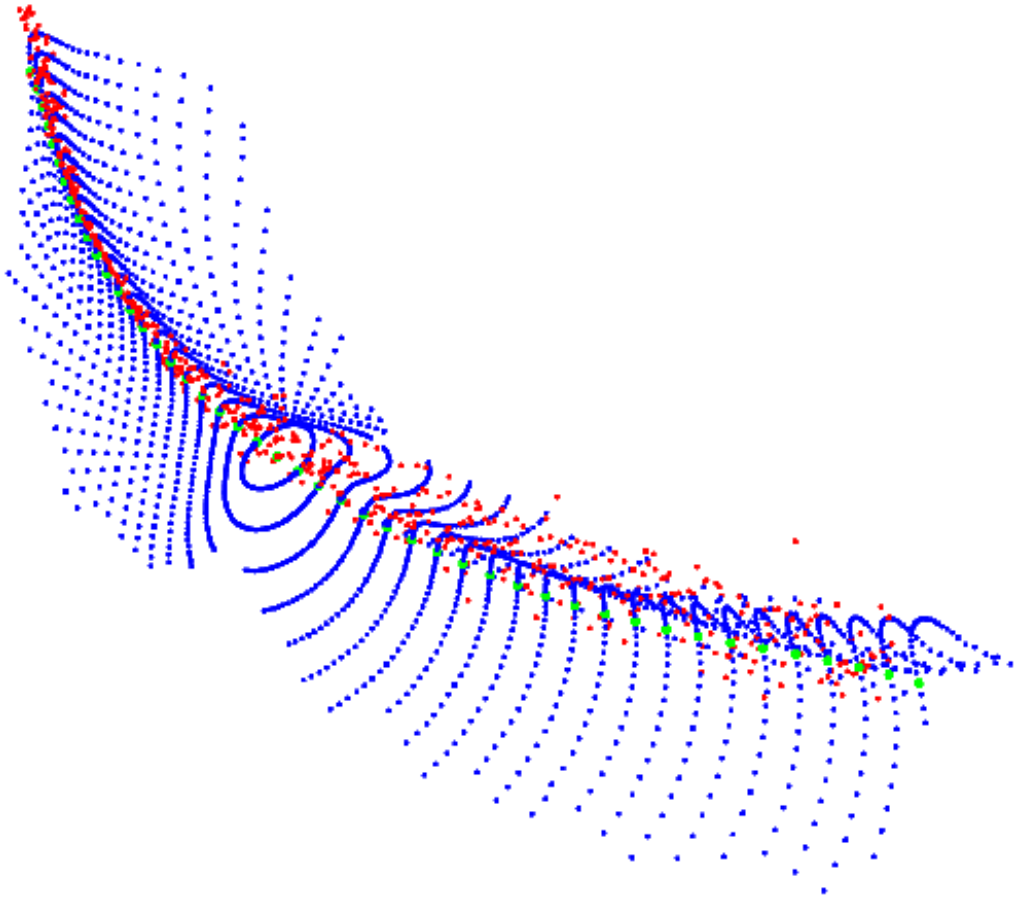
$$A_{l,i+1} = \mathbf{exp}_{A_{l,i}}(-r_{l,i}).$$

   - 5.7 Stop at $A_{l,i+1}$ when

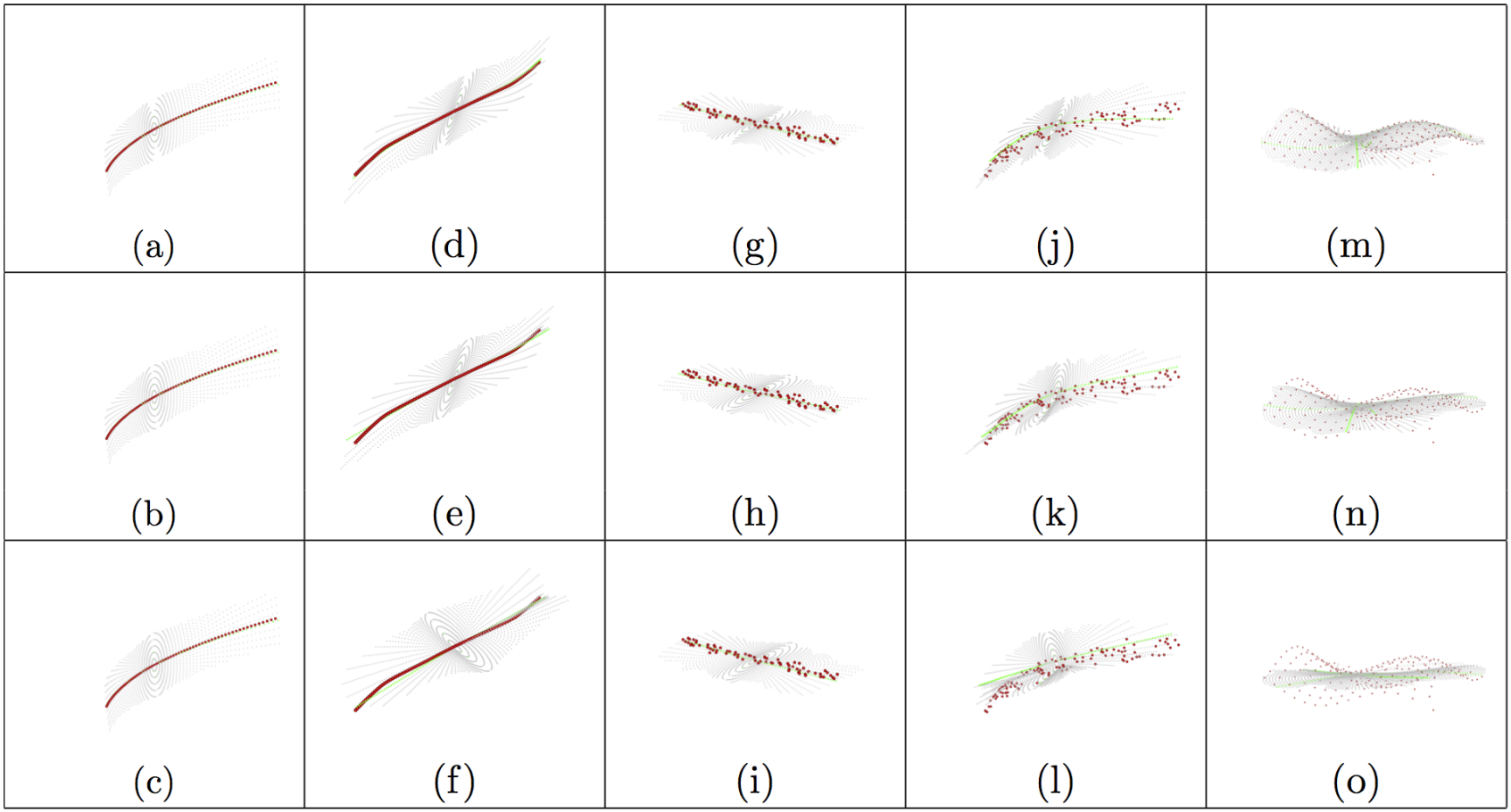
$$\left\langle \mathbf{log}_{A_{l,i+1}}(A_{l,i}), \mathbf{log}_{A_{l,i+1}}(x_j) \right\rangle \geq 0.$$

   for all $j = 1, \ldots, n$.
6. For every $l = 1, \ldots, 180$, connect $A_{l,i}$ with $A_{l,i+1}$ for all i we get a net of principal sub-manifold.
7. **Output**: $\mathcal{A}_l$ for $1 \leq l \leq 180$.

A view of the projected two-dimensional sub-manifold



Visualization of the projected sub-manifold for data on S^3



Principal sub-manifolds for four sea wave sets of data with noise on S^3

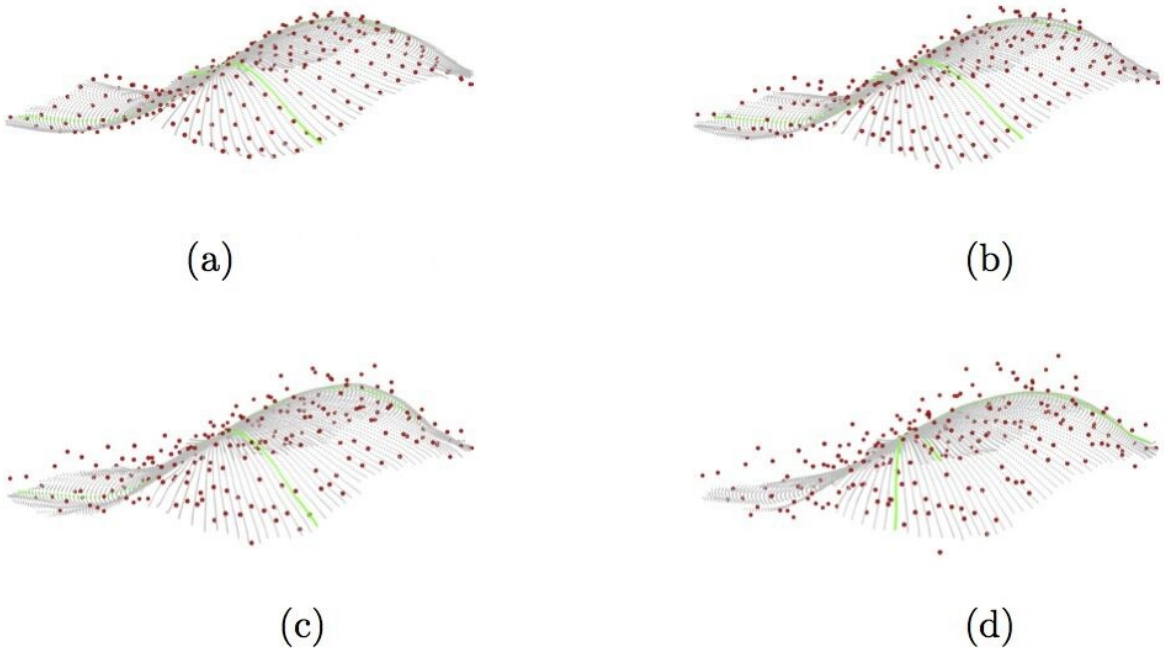


Figure 6: Principal sub-manifolds (with superimposed principal directions) for four sea wave sets of data with noise on S^3. (a) Principal sub-manifolds with no noise added. (b), (c) and (d) provide the same information for three different noise levels.

Principal variation of leaf growth-data

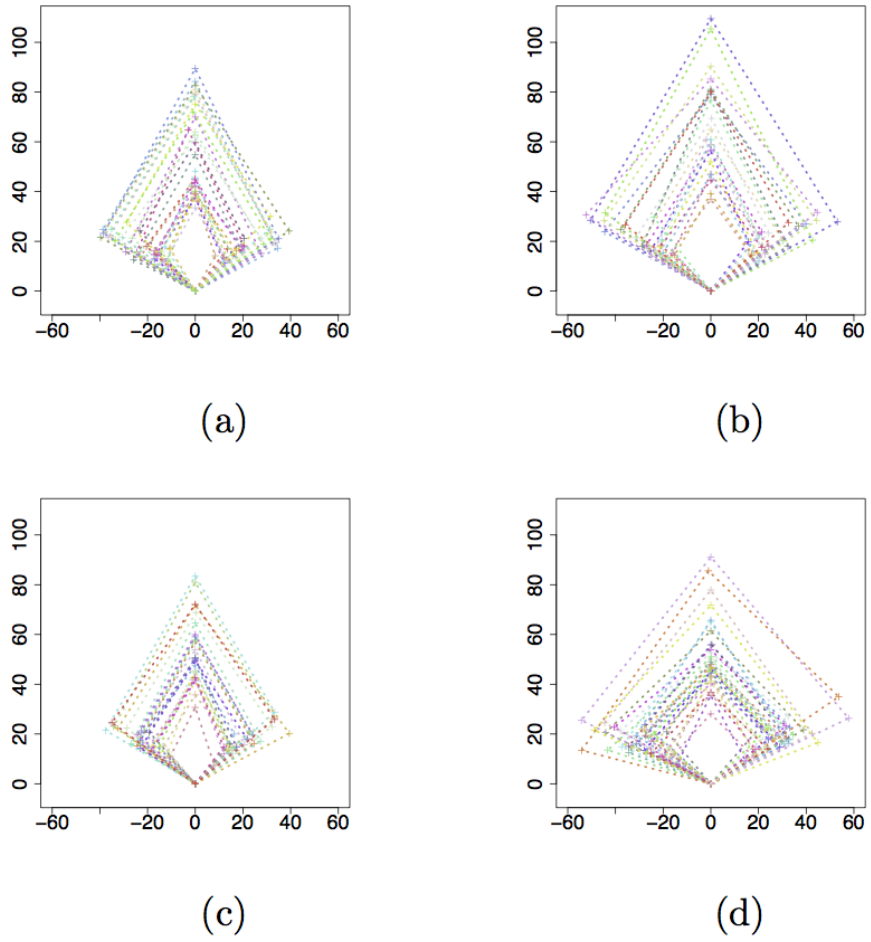


Figure 9: Leaf growth over a growing period of Clone 1 (a), Clone 2 (b), Clone 3 (c), and a reference tree (d). (a) Four landmarks on the leaf of clone 1 have been connected and represented by a polygon at each growing period (27 polygons totally); (b)-(d) provide the same information for Clone 2 (22 polygons), Clone 3 (24 polygons) and the reference tree (31 polygons).

Principal variation of leaf growth-result

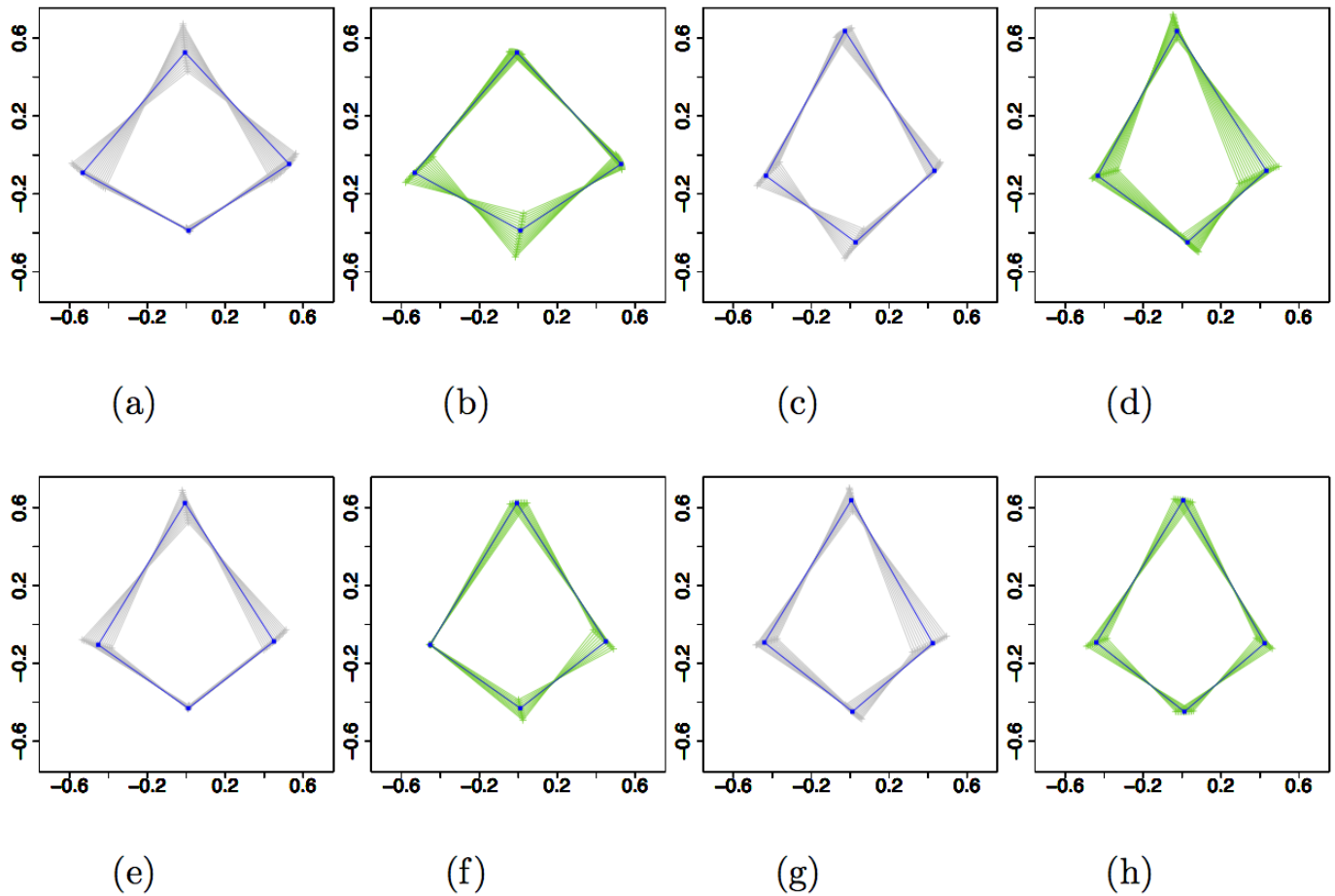


Figure 12: Principal sub-manifolds of the leaf growth data. (a) First principal direction obtained from the combined leaves at breast height and the crown of the reference tree; (b) Second principal direction obtained from the combined leaves at breast height and the crown of the reference tree. (c)-(h) provide the same information for clone 1, 2 and 3.

Take-Home Message

1. A manifold extension of principal components (retain canonical interpretation+ allow for more flexible reduction of non-geodesic variation)
2. Study of notions of local covariance on manifolds?
   - Behaviour of theoretical version of $\Sigma_h(x)$ as process over x or h or both?
3. Asymptotics for empirical flows/sub-manifolds?
4. More generally? Notions of covariance for manifolds?

References

- Panaretos, V., Pham, T., and Yao, Z. (2014). Principal Flows. *JASA*.
- Yao, Z. and Pham, T. (2016). Principal Sub-manifolds (2016). *Manuscript.*
- Liu, H., Yao, Z., Leung, S. and Chan, T. F. (2016). A Level Set Based Variational Principal Flow Method for Nonparametric Dimension Reduction on Riemannian Manifolds. *Manuscript.*