Geometric Whitney problem and inverse problems

Matti Lassas

in collaboration with

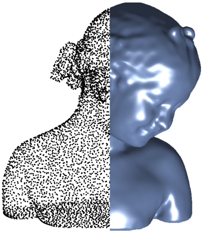
Charles Fefferman,
Sergei Ivanov,
Yaroslav Kurylev
Hariharan Narayanan

Finnish Centre of Excellence in
Inverse Modelling and Imaging
2018-2025

UNIVERSITY OF HELSINKI

Outline:

- Classical and geometric Whitney problems
- Surface interpolation
- Riemannian manifolds in inverse problems and other applications
- Manifold interpolation: Construction of a manifold from distances with small errors
- Learning a manifold from distances with large random noise

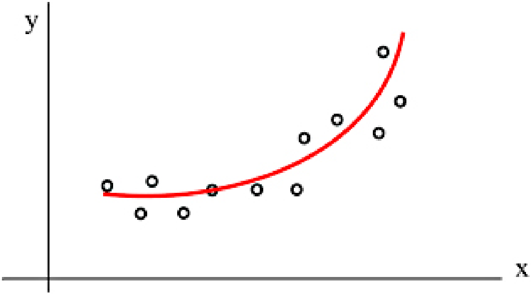


Whitney problem with errors

Let $K \subset \mathbb{R}^n$ be an arbitrary set, $h : K \to \mathbb{R}$, $m \in \mathbb{Z}_+$, and $\varepsilon > 0$. Does there exists a function $F \in C^m(\mathbb{R}^n)$ such that

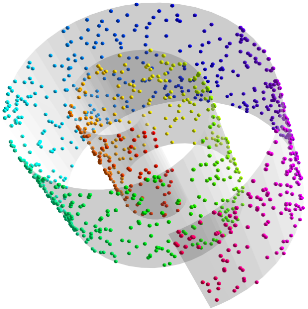
$$\sup_{x \in K} |F(x) - h(x)| \leq \varepsilon \ ?$$

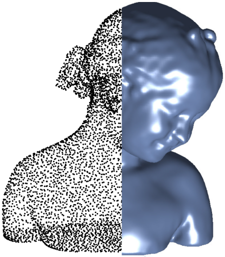
If such extension F exists, what is its optimal C^m-norm?



Problem A: Construction of a surface in $\mathbb{R}^d$ from a point cloud.

Assume that we are given a set $X \subset \mathbb{R}^d$ and $n < d$.
When one can construct a smooth n-dimensional surface $M \subset \mathbb{R}^d$ that approximates X?
How can the surface M can be constructed when X is given?





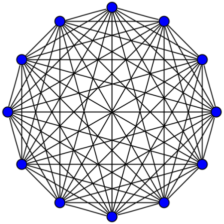
Figures by Matlab and M. Rouhani.

Problem B: Construction of a manifold from a discrete metric space.

Let (X, d_X) be a metric space. We ask when there exists a Riemannian manifold (M, g) such that

- the curvature and injectivity radius of M are bounded, and
- X approximates well M in the Gromov-Hausdorff topology.

How can the manifold (M, g) be constructed when X is given?



Unsolved extension problems

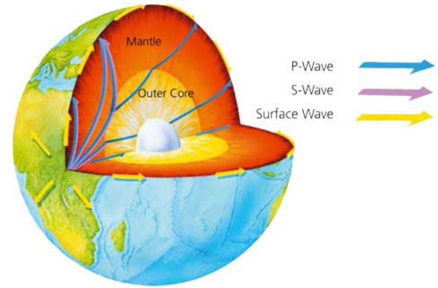
In the above problems a neighbourhood of the data points "covers" the whole manifold M (there are no holes).

The following extension problem for metric space is unsolved:
Let (X, d_X) be a metric space. Is there a Riemannian manifold (M, g) such that X can be embedded isometrically in M?

A special case is the boundary rigidity problem:
Let ∂M be the boundary of a compact manifold and $f : \partial M \times \partial M \to \mathbb{R}$. When we can construct a Riemannian metric g on M such that

$$\text{dist}_{(M,g)}(y_1, y_2) = f(y_1, y_2) \quad \text{for all } y_1, y_2 \in \partial M?$$

Example: Imaging of the interior of the Earth



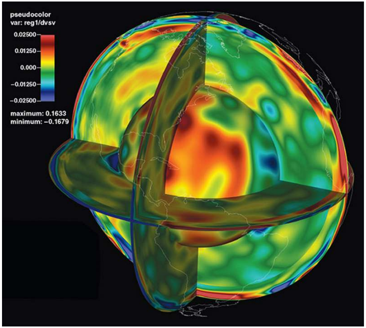


Fig. by Bozdag and Pugmire,

Let $M \subset \mathbb{R}^3$ and

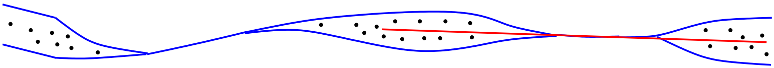
$$d_g(x, y) = \text{ travel time of waves from } x \text{ to } y, \quad x, y \in M.$$

Inverse problem: Can we determine the metric g in M when we know $d_g(z_1, z_2)$ for $z_1, z_2 \in \partial M$, that is, the travel times of the earthquakes between the points on the surface of the Earth?

When $g = c(x)^{-2}\delta_{jk}$ and $c(x)$ is close to 1, these data determine g uniquely (Burago-Ivanov 2010).

Outline:

- Classical and geometric Whitney problems
- Surface interpolation
- Riemannian manifolds in inverse problems and other applications
- Manifold interpolation: Construction of a manifold from distances with small errors
- Learning a manifold from distances with large random noise



Example: Manifold learning from point cloud data

Consider a data set $\mathcal{X} = \{x_j\}_{j=1}^N \subset \mathbb{R}^d$.
The ISOMAP face data set contains $N = 2370$ images of faces with $d = 2914$ pixels.

Question: Define $d_{\mathcal{X}}(x_j, x_k) = |x_j - x_k|_{\mathbb{R}^d}$ using the Euclidean distance. Can we find a submanifold of $\mathbb{R}^d$ that approximates $\mathcal{X}$?

Distance of two subsets

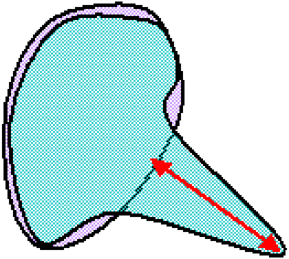
For a metric space Y and $A \subset Y$, the ε-neighborhood $U_\varepsilon(A)$ of A is

$$U_\varepsilon(A) = \{y \in Y;\ d(y, A) < \varepsilon\}, \quad \varepsilon > 0.$$

We say that A is ε-dense in Y if $U_\varepsilon(A) = Y$.

For a metric space Y and sets $A, B \subset Y$, the Hausdorff distance between A and B in Y is

$$d_H(A, B) = \max\left(\sup_{x\in A} d(x, B), \sup_{y \in B} d(y, A)\right).$$

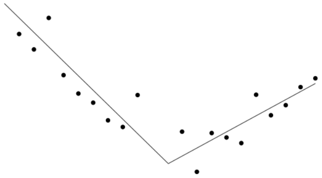


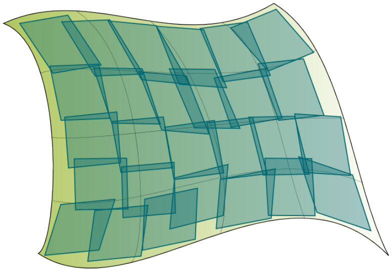
Let $E = \mathbb{R}^d$ and $B^E_r(x)$ be the ball in E with center x and radius r.

Definition

Let $X \subset E$, $n \in \mathbb{Z}_+$, and $r, \delta > 0$.
We say that X is δ-close to n-flats at scale r if for any $x \in X$, there exists an n-dimensional affine space $A_x \subset E$ through x such that

$$d_H\left(X \cap B^E_r(x),\, A_x \cap B^E_r(x)\right) \leq \delta.$$





Note: A bounded smooth n-surface in $\mathbb{R}^d$ is (Cr^2)-close to n-flats in scale r.

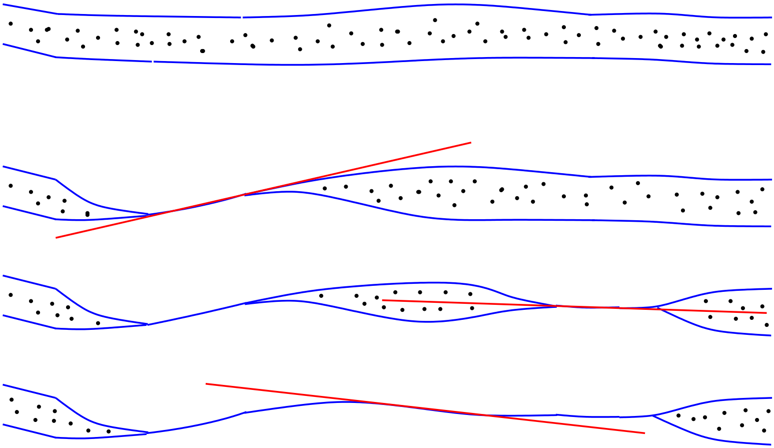
Surface interpolation

Theorem

Let E be a separable Hilbert space, $n \in \mathbb{Z}_+$, $r > 0$, and $\delta < \delta_0(r, n)$.
Suppose that $X \subset E$ is δ-close to n-flats at scale r. Then there exists a closed (or complete) n-dimensional smooth submanifold $M \subset E$ such that:

1. $d_H(X, M) \leq 5\delta$.
2. *The second fundamental form of M at every point is bounded by $C_n \delta r^{-2}$.*
3. *The normal injectivity radius of M is at least $r/3$.*

In particular, if $\delta < Cr^2$, the surface M has bounded curvature.



Algorithm SurfaceInterpolation: We consider the case $r = 1$ and assume that $X \subset E = \mathbb{R}^d$ is finite. We suppose that X is δ-close to n-flats at scale r. We implement the following steps:

1. Construct a maximal $\frac{1}{100}$-separated set $X_0 = \{q_i\}_{i=1}^k \subset X$.
2. For every point $q_i \in X_0$, let $A_i \subset E$ be an affine subspace that approximates $X \cap B_1(q_i)$ near q_i. Let $P_i : E \to E$ be orthogonal projectors onto A_i.
3. Let $\psi \in C_0^\infty([-\frac{1}{2}, \frac{1}{2}])$ be 1 in $[0, \frac{1}{3}]$ and $\varphi_i : E \to E$ be

   $$\varphi_i(x) = \mu_i(x) P_i(x) + (1 - \mu_i(x))x, \quad \mu_i(x) = \psi(|x - q_i|).$$

   Define $f : E \to E$ by

   $$f = \varphi_k \circ \varphi_{k-1} \circ \ldots \circ \varphi_1.$$

4. Construct the image $M = f(U_\delta(X))$.

The output is the n-dimensional surface $M \subset E$.

Outline:

- Classical and geometric Whitney problems
- Surface interpolation
- Riemannian manifolds in inverse problems and other applications
- Manifold interpolation: Construction of a manifold from distances with small errors
- Learning a manifold from distances with large random noise

Some earlier methods for manifold learning

Let $\{x_j\}_{j=1}^J \subset \mathbb{R}^d$ be points on submanifold $M \subset \mathbb{R}^d$, $d > n$.

- ‘Multi Dimensional Scaling’ (MDS) finds an embedding of data points into $\mathbb{R}^m$, $n < m < d$ by minimising a cost function

$$\min_{y_1,\dots,y_J\in\mathbb{R}^m} \sum_{j,k=1}^{J} \Big| \|y_j - y_k\|_{\mathbb{R}^m} - d_{jk} \Big|^2, \quad d_{jk} = \|x_j - x_k\|_{\mathbb{R}^d}$$

- ‘Isomap’ makes a graph of the K nearest neighbours and computes graph distances d_{jk}^G that approximate distances $d_M(x_j, x_k)$ along the surface. Then MDS is applied. Note that if there is $F : M \to \mathbb{R}^m$ such that $|F(x) - F(x')| = d_M(x, x')$, then the curvature of M is zero.

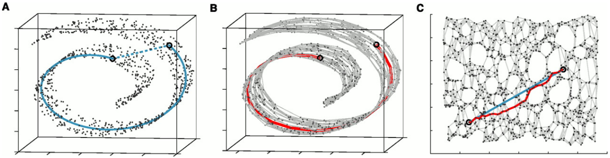


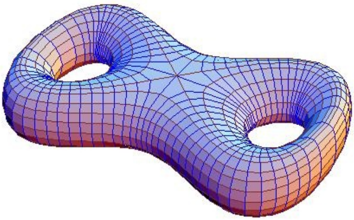
Figure by Tenenbaum et al., Science 2000

Construction of a manifold from discrete data.

Let $(\mathcal{X}, d_{\mathcal{X}})$ be a (discrete) metric space. We want to approximate it by a Riemannian manifold (M^*, g^*) so that

- $(\mathcal{X}, d_{\mathcal{X}})$ and (M^*, d_{g^*}) are almost isometric,
- the curvature and the injectivity radius of M^* are bounded.

Note that $\mathcal{X}$ is an "abstract metric space" and not a set of points in $\mathbb{R}^d$, and we want to learn the intrinsic metric of the manifold.



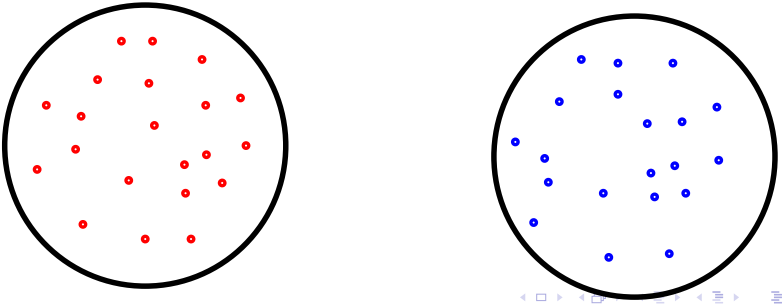
Distance of two metric spaces

Let (X, d_X) and (Y, d_Y) be (compact) metric spaces. Their Gromov-Hausdorff distance is

$$d_{GH}(X, Y) = \inf_Z \{d_H(X, Y);\ (Z, d_Z) \text{ is a metric space, } X \subset Z,\ Y \subset Z\}.$$

More practical definition: $d_{GH}(X, Y)$ is the infimum of all $\varepsilon > 0$ for which there are ε-dense sequences $(x_j)_{j=1}^J \subset X$ and $(y_j)_{j=1}^J \subset Y$ such that

$$|d_X(x_j, x_k) - d_Y(y_j, y_k)| \le \varepsilon, \quad \text{for all } j, k = 1, 2 \ldots, J.$$



Example 1: Non-Euclidean metric in data sets

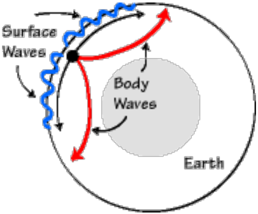
Consider a data set $\mathcal{X} = \{x_j\}_{j=1}^N \subset \mathbb{R}^d$.
The ISOMAP face data set contains $N = 2370$ images of faces with $d = 2914$ pixels.

Question: Define $d_{\mathcal{X}}(x_j, x_k)$ using Wasserstein distance related to optimal transport. Does $(\mathcal{X}, d_{\mathcal{X}})$ approximate a manifold and how this manifold can be constructed?

Example 2: Travel time distances of points

Surface waves produced by earthquakes travel near the boundary of the Earth. The observations of several earthquakes give information on travel times $d_T(x, y)$ between the points $x, y \in \mathbb{S}^2$.

Question: Can one determine the Riemannian metric associated to surface waves from the travel times with measurement errors?



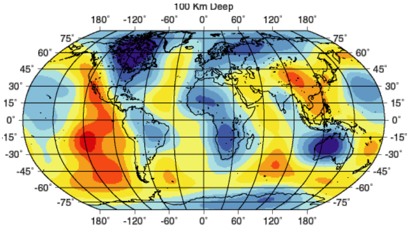
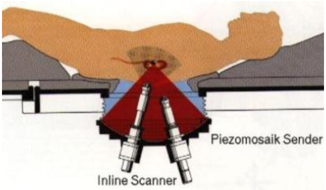
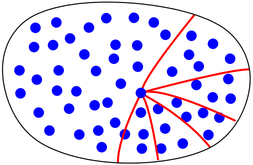


Figure by Su-Woodward-Dziewonski, 1994

Example 3: An inverse problem for a manifold

Consider a physical $D \subset \mathbb{R}^3$ with an unknown wave speed $c(x)$. We can use boundary measurements to construct the distances $d_g(x_j, x_k)$ in a discrete set $X = \{x_j \in M : \; j = 1, 2, \ldots, N\}$ (Belishev-Kurylev 1992, Bingham-Kurylev-L.-Siltanen 2008).





The solution for Problem B gives a construction of a smooth Riemannian manifold from (X, d_X). This Riemannian metric is close to the travel time metric g determined by $c(x)$.

Outline:

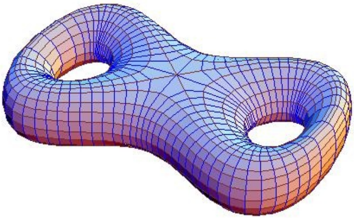
- Classical and geometric Whitney problems
- Surface interpolation
- Riemannian manifolds in inverse problems and other applications
- Manifold interpolation: Construction of a manifold from distances with small errors
- Ideas of the proofs and applications in geometry
- Learning a manifold from distances with large random noise

Construction of a manifold from discrete data.

Let $(\mathcal{X}, d_{\mathcal{X}})$ be a (discrete) metric space. We aim to answer the question if there exists a Riemannian manifold (M^*, g^*) that approximates $\mathcal{X}$ so that

- $d_{GH}((\mathcal{X}, d_{\mathcal{X}}), (M^*, d_{g^*})) < \varepsilon$,
- the curvature and the injectivity radius of M^* are bounded.

Note that $\mathcal{X}$ is an "abstract metric space" and not a set of points in $\mathbb{R}^d$, and we want to learn the intrinsic metric of the manifold.



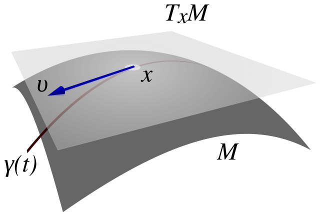
A local condition

Let $B_r^X(x)$ denote the ball of the metric space X and $B_r^{\mathbb{R}^n}(0)$ denote the ball of $\mathbb{R}^n$.

Definition

Let X be a metric space, $r > \delta > 0$, $n \in \mathbb{Z}_+$. We say that X is δ-close to $\mathbb{R}^n$ at scale r if, for any $x \in X$,

$$d_{GH}\left(B_r^X(x)\,,\ B_r^{\mathbb{R}^n}(0)\right) < \delta.$$



Note: A compact smooth n-manifold is (Cr^3)-close $\mathbb{R}^n$ at scale r.

A global condition

Definition

Let $X = (X, d)$ be a metric space and $\delta > 0$. A δ-*chain* in X is a sequence $x_1, x_2, \ldots, x_N \in X$ such that $d(x_j, x_{j+1}) < \delta$ for all j.

A sequence $x_1, x_2, \ldots, x_N \in X$ is said to be δ-*straight* if

$$d(x_i, x_j) + d(x_j, x_k) < d(x_i, x_k) + \delta \quad \text{for all } 1 \leq i < j < k \leq N.$$

We say that X is δ-intrinsic if for every pair of points $x, y \in X$ there is a δ-straight δ-chain $x_1, \ldots, x_N$ with $x_1 = x$ and $x_N = y$.

Manifold learning

Theorem

Let X be a metric space with a bounded diameter, $n \in \mathbb{Z}_+$, $r > 0$, and $0 < \delta < \delta_0(r, n)$. Suppose that X is δ-intrinsic and δ-close to $\mathbb{R}^n$ at scale r. Then there exists a compact n-dimensional Riemannian manifold M such that

1. *X and M satisfy*

$$d_{GH}(X, M) < C\delta r^{-1} diam\,(X).$$

2. *The sectional curvature $\mathrm{Sec}(M)$ of M satisfies*

$$|\,\mathrm{Sec}(M)| \le C\delta r^{-3}.$$

3. *The injectivity radius of M is bounded below by $r/2$.*

In particular, if $\delta < Cr^3$, the constructed manifold M has bounded curvature.

Manifold learning

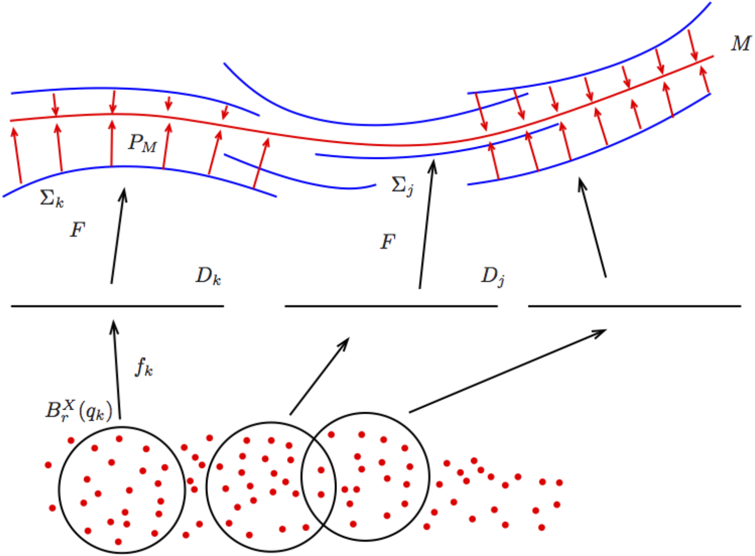
Theorem

Let X be a metric space with a bounded diameter, $n \in \mathbb{Z}_+$, $r > 0$, and $0 < \delta < \delta_0(r,n)$. Suppose that X is δ-intrinsic and δ-close to $\mathbb{R}^n$ at scale r. Then there exists a compact n-dimensional Riemannian manifold M such that

1. *X and M satisfy*
$$d_{GH}(X,M) < C\delta r^{-1} diam\,(X).$$
2. *The sectional curvature $\mathrm{Sec}(M)$ of M satisfies*
$$|\,\mathrm{Sec}(M)| \leq C\delta r^{-3}.$$
3. *The injectivity radius of M is bounded below by $r/2$.*

In particular, if $\delta < Cr^3$, the constructed manifold M has bounded curvature.

Rough idea of the proof of manifold interpolation



Assume that we are given a finite metric space (X, d).
We do the following steps:

1. Select a maximal $\frac{r}{100}$-separated set $X_0 = \{q_i\}_{i=1}^{J} \subset X$.
2. Choose disjoint balls $D_i = B_r(p_i) \subset \mathbb{R}^n$ for $i = 1, 2, \ldots, J$ and construct a δ-isometry $f_i : B_r^X(q_i) \to D_i$.
3. For all $q_i, q_j \in X_0$ such that $d(q_i, q_j) < r$, find affine transition maps $A_{ij} : \mathbb{R}^n \to \mathbb{R}^n$, such that

$$|A_{ij}(f_i(x)) - f_j(x)| < C\delta, \quad \text{for } x \in B_r^X(q_i) \cap B_r^X(q_j).$$

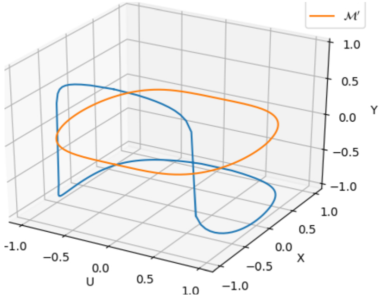
   When $i = j$, we define $A_{jj} = Id$.

4. Let $\Phi \in C_0^\infty(\mathbb{R}^n)$ be 1 near zero, and $\Omega = \bigcup_i D_i$.
Define smooth indicator functions $\psi_{ij}(x) = \Phi(A_{ij}(x) - p_j)$.
Define a map $F_j \colon \Omega \to \mathbb{R}^{n+1}$ as follows: For $x \in D_i = B_r(p_i)$, put

$$F_j(x) = \begin{cases} \big(\ \psi_{ij}(x) \cdot (A_{ij}(x) - p_j)\ ,\ \psi_{ij}(x)\,\big), & \text{if } d(q_i, q_j) < r, \\ 0, & \text{otherwise.} \end{cases}$$

5. Denote $E = \mathbb{R}^m$, $m = (n+1)J$ and define

$$F \colon \Omega \to E, \quad F(x) = (F_j(x))_{j=1}^J.$$



6. Construct the local patches $\Sigma_i = F(D_i) \subset E$.
7. Apply algorithm *SurfaceInterpolation* for the set $\bigcup_i \Sigma_i$ to construct a surface $M \subset E$.
8. Let P_M be the normal projection on M.
9. Construct a metric tensor g on M by pushing forward the Euclidean metric g^e on D_i in the maps $P_M \circ F$ and computing a weighted average of the obtained metric tensors.

The output is the surface $M \subset E$ and the metric g on it.

Next we consider applications of the above theorem in reconstruction of an unknown manifold.

Theorem (Fefferman, Ivanov, Kurylev, L., Narayanan 2015)

Let $0 < \delta < c_1(n, K)$ and M be a compact n-dimensional manifold with $|\operatorname{Sec}(M)| \le K$ and $\operatorname{inj}(M) > 2(\delta/K)^{1/3}$. Let $\mathcal{X} = \{x_j\}_{j=1}^N$ be δ-dense in M and $\widetilde{d} \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+ \cup \{0\}$ satisfy

$$|\widetilde{d}(x,y) - d_M(x,y)| \le \delta, \quad x, y \in \mathcal{X}.$$

Given the values $\widetilde{d}(x_j, x_k)$, $j, k = 1, \dots, N$, one can construct a compact n-dimensional Riemannian manifold (M^*, g^*) such that:

1. *There is a diffeomorphism $F : M^* \to M$ satisfying*

$$\frac{1}{L} \le \frac{d_M(F(x), F(y))}{d_{M^*}(x,y)} \le L, \quad \text{for } x, y \in M^*, \ L = 1 + C_n K^{1/3} \delta^{2/3}.$$

2. $|\operatorname{Sec}(M^*)| \le C_n K$.
3. $\operatorname{inj}(M^*) \ge \min\{(C_n K)^{-1/2}, (1 - C_n K^{1/3} \delta^{2/3}) \operatorname{inj}(M)\}$.

Outline:

- Classical and geometric Whitney problems
- Surface interpolation
- Riemannian manifolds in inverse problems and other applications
- Manifold interpolation: Construction of a manifold from distances with small errors
- Learning a manifold from distances with large random noise

Random sample points and random errors

Manifolds with bounded geometry:

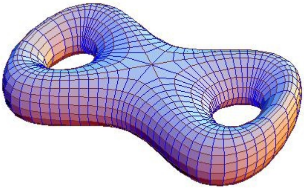
Let $n \geq 2$ be an integer, $K > 0$, $D > 0$, $i_0 > 0$. Let (M, g) be a compact Riemannian manifold of dimension n such that

$$\begin{aligned} &i)\, \|\mathrm{Sec}_M\|_{L^\infty(M)} \leq K, \\ &ii)\, \mathrm{diam}\,(M) \leq D, \\ &iii)\, \mathrm{inj}\,(M) \geq i_0, \end{aligned} \tag{1}$$

We consider measurements in randomly sampled points:

Let X_j, $j = 1, 2, \dots, N$ be independently samples from probability distribution μ on M satisfying

$$0 < c_{min} \leq \frac{d\mu}{d\mathrm{Vol}_g} \leq c_{max}.$$



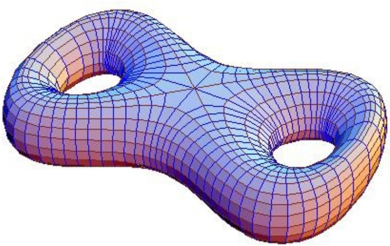
Definition

Let X_j, $j = 1, 2, \ldots, N$ be independent, identically distributed (i.i.d.) random variables having distribution μ.
Let $\sigma > 0$, $\beta > 1$ and η_{jk} be i.i.d. random variables satisfying

$$\mathbb{E}\eta_{jk} = 0, \quad \mathbb{E}(\eta_{jk}^2) = \sigma^2, \quad \mathbb{E}e^{|\eta_{jk}|} = \beta.$$

In particular, Gaussian noise satisfies these conditions.
We assume that all random variables η_{jk} and X_j are independent.
We consider noisy measurements

$$D_{jk} = d_M(X_j, X_k) + \eta_{jk}.$$



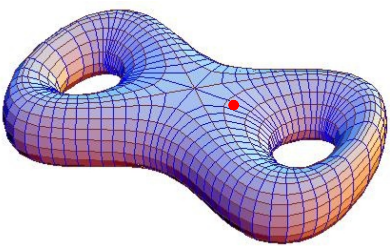
Definition

Let X_j, $j = 1, 2, \ldots, N$ be independent, identically distributed (i.i.d.) random variables having distribution μ.
Let $\sigma > 0$, $\beta > 1$ and η_{jk} be i.i.d. random variables satisfying

$$\mathbb{E}\eta_{jk} = 0, \quad \mathbb{E}(\eta_{jk}^2) = \sigma^2, \quad \mathbb{E}e^{|\eta_{jk}|} = \beta.$$

In particular, Gaussian noise satisfies these conditions.
We assume that all random variables η_{jk} and X_j are independent.
We consider noisy measurements

$$D_{jk} = d_M(X_j, X_k) + \eta_{jk}.$$



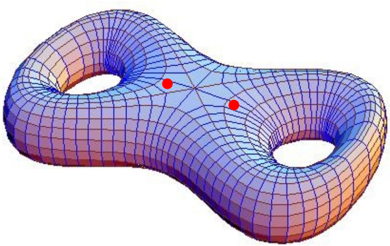
Definition

Let X_j, $j = 1, 2, \dots, N$ be independent, identically distributed (i.i.d.) random variables having distribution μ.
Let $\sigma > 0$, $\beta > 1$ and η_{jk} be i.i.d. random variables satisfying

$$\mathbb{E}\eta_{jk} = 0, \quad \mathbb{E}(\eta_{jk}^2) = \sigma^2, \quad \mathbb{E}e^{|\eta_{jk}|} = \beta.$$

In particular, Gaussian noise satisfies these conditions.
We assume that all random variables η_{jk} and X_j are independent.
We consider noisy measurements

$$D_{jk} = d_M(X_j, X_k) + \eta_{jk}.$$



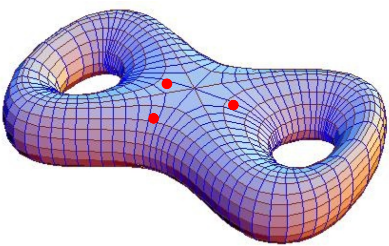
Definition

Let X_j, $j = 1, 2, \ldots, N$ be independent, identically distributed (i.i.d.) random variables having distribution μ.
Let $\sigma > 0$, $\beta > 1$ and η_{jk} be i.i.d. random variables satisfying

$$\mathbb{E}\eta_{jk} = 0, \quad \mathbb{E}(\eta_{jk}^2) = \sigma^2, \quad \mathbb{E}e^{|\eta_{jk}|} = \beta.$$

In particular, Gaussian noise satisfies these conditions.
We assume that all random variables η_{jk} and X_j are independent.
We consider noisy measurements

$$D_{jk} = d_M(X_j, X_k) + \eta_{jk}.$$



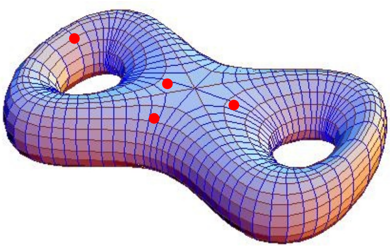
Definition

Let X_j, $j = 1, 2, \ldots, N$ be independent, identically distributed (i.i.d.) random variables having distribution μ.
Let $\sigma > 0$, $\beta > 1$ and η_{jk} be i.i.d. random variables satisfying

$$\mathbb{E}\eta_{jk} = 0, \quad \mathbb{E}(\eta_{jk}^2) = \sigma^2, \quad \mathbb{E}e^{|\eta_{jk}|} = \beta.$$

In particular, Gaussian noise satisfies these conditions.
We assume that all random variables η_{jk} and X_j are independent.
We consider noisy measurements

$$D_{jk} = d_M(X_j, X_k) + \eta_{jk}.$$



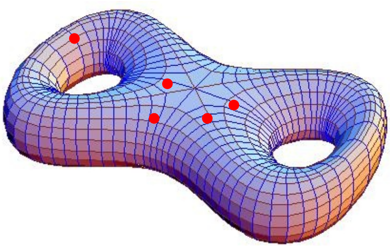
Definition

Let X_j, $j = 1, 2, \dots, N$ be independent, identically distributed (i.i.d.) random variables having distribution μ.
Let $\sigma > 0$, $\beta > 1$ and η_{jk} be i.i.d. random variables satisfying

$$\mathbb{E}\eta_{jk} = 0, \quad \mathbb{E}(\eta_{jk}^2) = \sigma^2, \quad \mathbb{E}e^{|\eta_{jk}|} = \beta.$$

In particular, Gaussian noise satisfies these conditions.
We assume that all random variables η_{jk} and X_j are independent.
We consider noisy measurements

$$D_{jk} = d_M(X_j, X_k) + \eta_{jk}.$$



Theorem (Fefferman, Ivanov, L., Narayanan 2019)

Let $n \geq 3$, $D, K, i_0, c_{min}, c_{max}, \sigma, \beta > 0$ be given. Then there are δ_0, C_0 and C_1 such that the following holds: Let $\delta \in (0, \delta_0)$, $\theta \in (0, \frac{1}{2})$ and (M, g) be a compact manifold satisfying bounds (1). Then with a probability $1 - \theta$, σ^2 and the noisy distances $D_{jk} = d_M(X_j, X_k) + \eta_{jk}$, $j, k \leq N$ of N randomly chosen points, where

$$N \geq C_0 \frac{1}{\delta^{3n}} \left(\log(\frac{1}{\theta}) + \log(\frac{1}{\delta}) \right),$$

determine a Riemannian manifold (M^*, g^*) such that

1. *There is a diffeomorphism $F : M^* \to M$ satisfying*

$$\frac{1}{L} \leq \frac{d_M(F(x), F(y))}{d_{M^*}(x, y)} \leq L, \quad \text{for all } x, y \in M^*,$$

   where $L = 1 + C_1 \delta$.

2. *The sectional curvature Sec_{M^*} of M^* satisfies $|Sec_{M^*}| \leq C_1 K$.*
3. *The injectivity radius $inj(M^*)$ of M^* is close to $inj(M)$.*

Theorem (Fefferman, Ivanov, L., Narayanan 2019)

Let $n \geq 3$, $D, K, i_0, c_{min}, c_{max}, \sigma, \beta > 0$ be given. Then there are δ_0, C_0 and C_1 such that the following holds: Let $\delta \in (0, \delta_0)$, $\theta \in (0, \frac{1}{2})$ and (M, g) be a compact manifold satisfying bounds (1). Then with a probability $1 - \theta$, σ^2 and the noisy distances $D_{jk} = d_M(X_j, X_k) + \eta_{jk}$, $j, k \leq N$ of N randomly chosen points, where

$$N \geq C_0 \frac{1}{\delta^{3n}} \left(\log(\frac{1}{\theta}) + \log(\frac{1}{\delta}) \right),$$

determine a Riemannian manifold (M^*, g^*) such that

1. *There is a diffeomorphism $F : M^* \to M$ satisfying*

   $$\frac{1}{L} \leq \frac{d_M(F(x), F(y))}{d_{M^*}(x, y)} \leq L, \quad \text{for all } x, y \in M^*,$$

   where $L = 1 + C_1 \delta$.
2. *The sectional curvature Sec_{M^*} of M^* satisfies $|Sec_{M^*}| \leq C_1 K$.*
3. *The injectivity radius $inj(M^*)$ of M^* is close to $inj(M)$.*

For $z \in M$, let $r_z : M \to \mathbb{R}$ be the distance function from z,

$$r_z(x) = d_M(z, x), \quad x \in M.$$

For $y, z \in M$, we consider the "rough distance function"

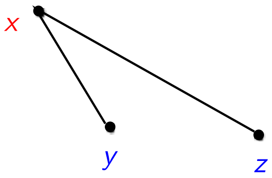
$$\kappa(y, z) = \|r_y - r_z\|^2_{L^2(M)} = \int_M |d_M(y, x) - d_M(z, x)|^2 d\mu(x).$$

Lemma

There is a constant $c_0 \in (0, 1)$ such that

$$c_0^2 d_M(y, z)^2 \leq \|r_y - r_z\|^2_{L^2(M, d\mu)} \leq d_M(y, z)^2, \quad y, z \in M.$$

That is, the map $R : z \mapsto r_z$ is a bi-Lipschitz embedding $R : M \to R(M) \subset L^2(M)$.



Lemma (Hoeffding's inequality)

Let $Z_1, \ldots, Z_N$ *be* N *independent, identically distributed copies of the random variable* Z *whose range is* $[0,1]$*. Then, for* $\varepsilon > 0$*, we have*

$$\mathbb{P}\left[\left|\frac{1}{N}\left(\sum_{j=1}^{N} Z_j\right) - \mathbb{E}Z\right| \leq \varepsilon\right] \geq 1 - 2\exp(-2N\varepsilon^2).$$

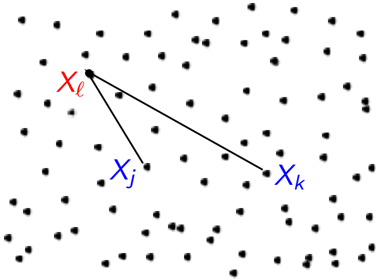
We consider three sets $S_1, S_2, S_3 \subset \{X_j\}$, where $N_i = \#S_i$ satisfy $N_1 > N_2 > N_3$. We call $S_1 = \{X_1, \dots, X_{N_1}\}$ the densest net, S_2 the medium dense net and S_3 the coarse net.

We give an algorithm to construct (M^*, g^*) from noisy data.

Step 1: For $X_j, X_k \in S_2$ are in the "medium dense net", we compute

$$\kappa_{app}(X_j, X_k) = \frac{1}{N_1} \sum_{\ell=1}^{N_1} |D_{j\ell} - D_{k\ell}|^2 - 2\sigma^2,$$

where we take a sum over the "densest net" S_1.



Denote $\kappa(X_j, X_k) = \|r_{X_j} - r_{X_k}\|^2_{L^2(M)}$. A simple calculation shows

$$\mathbb{E}\left(|D_{j\ell} - D_{k\ell}|^2 \,\middle|\, X_j, X_k \right) = \|r_{X_j} - r_{X_k}\|^2_{L^2(M)} + 2\sigma^2.$$

We recall that for $X_j, X_k \in S_2$,

$$\kappa_{app}(X_j, X_k) = \frac{1}{N_1} \sum_{\ell=1}^{N_1} |D_{j\ell} - D_{k\ell}|^2 - 2\sigma^2$$

Thus Hoeffding's inequality yields the following:

Lemma

Let $L > D + 1$ and $\varepsilon > 0$. If $|\eta_{jk}| < L$ almost surely, then

$$\mathbb{P}\left[\left| \kappa_{app}(X_j, X_k) - \kappa(X_j, X_k) \right| \leq \varepsilon \right] \geq 1 - 2\exp(-\frac{1}{8} N_1 L^{-4} \varepsilon^2).$$

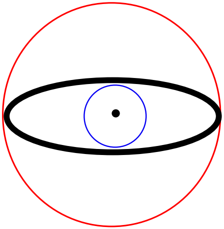
Recall that function $\kappa(y,z)=\|r_y-r_z\|^2_{L^2(M)}\approx\kappa_{app}(y,z)$ is a rough distance function:

$$c_0^2 d_M(y,z)^2\le\kappa(y,z)\le d_M(y,z)^2.$$

Let $W(y,\rho)$ and $W_{app}(y,\rho)$ be the sets

$$\begin{aligned}W(y,\rho)&=\{z\in M:\ \kappa(y,z)<\rho^2\},\\ W_{app}(y,\rho)&=\{z\in M:\ \kappa_{app}(y,z)<\rho^2\}.\end{aligned}$$

We have $B_M(y,\frac{1}{c_0}\rho)\subset W(y,\rho)\subset B_M(y,\rho)$.



For $y_1, y_2 \in M$, we define the averaged distances

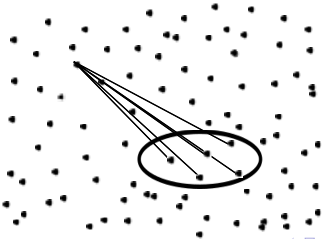
$$d_\rho(y_1, y_2) = \frac{1}{\mu(W(y_1, \rho))} \int_{W(y_1,\rho)} d_M(z, y_2)\, d\mu(z).$$

Step 2: For $X_j, X_{j'} \in S_3$, where S_3 is the coarse net, compute

$$d_\rho^{app}(X_j, X_{j'}) = \frac{1}{\#(S_2 \cap W_{app}(X_j, \rho))} \sum_{X_k \in S_2 \cap W_{app}(X_j,\rho)} D_{kj'}.$$

There is $\delta_1 = \delta_1(\rho, \theta)$ such that

$$\mathbb{P}[\forall X_j, X_{j'} \in S_3 : \ |d_\rho^{app}(X_j, X_{j'}) - d_M(X_j, X_{j'})| < \delta_1] \geq 1 - \theta.$$



Summarizing, for points $S_3 = \{y_1, y_2, \ldots, y_{N_3}\}$ we find $d_\rho^{app}(y_j, y_{j'})$ such that

$$|d_\rho^{app}(y_j, y_{j'}) - d_M(y_j, y_{j'})| < \delta_1$$

with a large probability.

Step 3: Using the deterministic results with small errors we find a smooth manifold (M^*, g^*) using the net S_3 and the approximate distance $d_\rho^{app}(y_1, y_2)$ of $y_1, y_2 \in S_3$.

Generalization with missing data

Recall that $D_{jk} = d_M(X_j, X_k) + \eta_{jk}$.
We can assume that we are given

$$D_{jk}^{(\text{partial data})} = \begin{cases} D_{jk} & \text{if } Y_{jk} = 1, \\ \text{`missing'} & \text{if } Y_{jk} = 0, \end{cases}$$

where $Y_{jk} \in \{0, 1\}$ are independent random variables,

$$\mathbb{P}(Y_{jk} = 1 \,|\, X_j, X_k) = \Phi(X_j, X_k)$$

and $\Phi : M \times M \to \mathbb{R}$ is some (unknown) function such that there is a smooth non-increasing function $h : [0, \infty) \to [0, 1]$ so that

$$c_1\, h(d_M(x, y)) \leq \Phi(x, y) \leq c_2\, h(d_M(x, y)).$$

Thank you for your attention!