# Manifold learning with random errors and inverse problems

Matti Lassas

in collaboration with

Charles Fefferman,
Sergei Ivanov,
Hariharan Narayanan

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

UNIVERSITY OF HELSINKI

Outline:

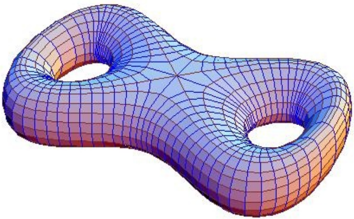
- Manifold learning problems and inverse problems
- Learning a manifold from distances with small noise
- Learning a manifold from distances with large random noise

# Construction of a manifold from discrete data.

Let $(\mathcal{X}, d_{\mathcal{X}})$ be a (discrete) metric space. We 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.



# 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 having measurement errors?

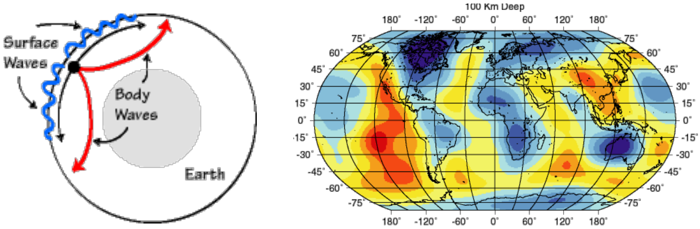



Figure by Su-Woodward-Dziewonski, 1994

# Example 3: An inverse problem for a manifold

Consider the eigenvalues $\lambda_j$ and eigenfunctions $\varphi_j$ satisfying

$$-\Delta_g \varphi_j = \lambda_j \varphi_j \quad \text{on } M.$$

In the inverse interior spectral problem one is given

a ball $B = B_M(p, r) \subset M$,
eigenvalues $\lambda_j, \quad j = 1, 2, 3, \dots,$
restrictions of eigenfunctions, $\varphi_j|_B, \quad j = 1, 2, 3, \dots$

and the goal is to determine the isometry type of $(M, g)$.

**Theorem (Bosi-Kurylev-L. 2017)**

*Let $n \in \mathbb{Z}_+$ and $K, D, i_0, r_0 > 0$.*
*There are $\theta, C_0, \delta_0$ such that for all $\delta < \delta_0$ the following is true:*

*Let $(M, g)$ be a Riemannnian manifold such that*
*$\|Ric(M)\|_{C^3(M)} \leq K$, diam $(M) \leq D$, inj$(M) \geq i_0$.*

*Identify the ball $B_M(p, r_0)$ with $B(r_0) \subset \mathbb{R}^n$ in normal coordinates.*

*Assume that we are given $g^a$, $\varphi_j^a$ and $\lambda_j^a$ such that*
*i) The metric tensor satisfies $\|g^a - g\|_{L^\infty(B(r_0))} < \delta$,*
*ii) $|\lambda_j^a - \lambda_j| < \delta$ and $\|\varphi_j^a - \varphi_j\|_{L^2(B(r_0))} < \delta$ when $\lambda_j < \frac{1}{\delta}$.*

*Then we can construct a metric space $(\mathcal{X}, d_\mathcal{X})$ such that*

$$d_{GH}(M, \mathcal{X}) \leq \frac{C_0}{\left(\ln\left(\ln \frac{1}{\delta}\right)\right)^\theta} = \varepsilon,$$

*that is, there is an $\varepsilon$-dense subset $\{p_j : \ j = 1, \dots, N\} \subset M$ and*
*$\mathcal{X} = \{x_j : \ j = 1, \dots, N\}$ such that $|d_M(p_j, p_k) - d_\mathcal{X}(x_j, x_k)| \leq \varepsilon$.*

# 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 $K$ 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.

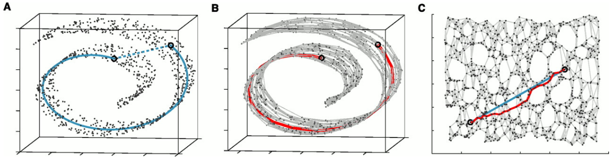



Figure by Tenenbaum et al., Science 2000

Outline:

- Manifold learning problems and inverse problems
- Learning a manifold from distances with small noise
- Learning a manifold from distances with large random noise

**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* $\delta$*-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* $\widetilde{d}(x_j, x_k)$*,* $j, k = 1, \ldots, N$*, 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. *The injectivity radius* $\operatorname{inj}(M^*)$ *of* $M^*$ *satisfies*

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

Outline:

- Manifold learning problems and inverse problems
- Learning a manifold from distances with small noise
- Learning a manifold from distances with large random noise

# Random sample points and random errors

**Manifolds with bounded geometry:**

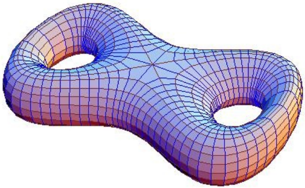
Let $n \geq 2$ be an integer, $K > 0$, $D > 0$, $i_0 > 0$. Let $(M, g)$ be a compact Riemannian manifold of dimension $n$ such that

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

**We consider measurements in randomly sampled points:**

Let $X_j$, $j = 1, 2, \ldots, N$ be independently samples from probability distribution $\mu$ on $M$ such that

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



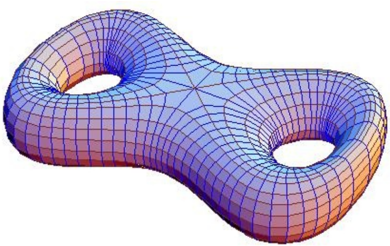
### Definition

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

Let $\sigma > 0$, $\beta > 1$ and $\eta_{jk}$ be i.i.d. random variables satisfying

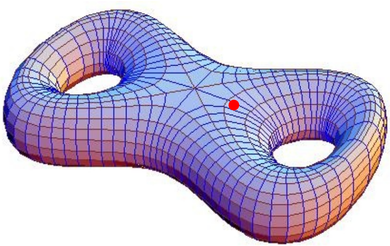
$$\mathbb{E}\eta_{jk} = 0, \quad \mathbb{E}(\eta_{jk}^2) = \sigma^2, \quad \mathbb{E}e^{|\eta_{jk}|} = \beta.$$

In particular, Gaussian noise satisfies these conditions.

We assume that all random variables $\eta_{jk}$ 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 $\mu$.

Let $\sigma > 0$, $\beta > 1$ and $\eta_{jk}$ be i.i.d. random variables satisfying

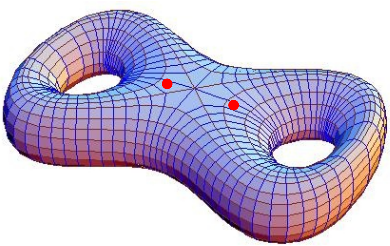
$$\mathbb{E}\eta_{jk} = 0, \quad \mathbb{E}(\eta_{jk}^2) = \sigma^2, \quad \mathbb{E}e^{|\eta_{jk}|} = \beta.$$

In particular, Gaussian noise satisfies these conditions.

We assume that all random variables $\eta_{jk}$ and $X_j$ are independent.

We consider noisy measurements

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



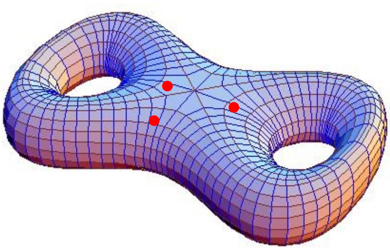
### Definition

Let $X_j$, $j = 1, 2, \ldots, N$ be independent, identically distributed (i.i.d.) random variables having distribution $\mu$.
Let $\sigma > 0$, $\beta > 1$ and $\eta_{jk}$ 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 $\eta_{jk}$ and $X_j$ are independent.
We consider noisy measurements

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



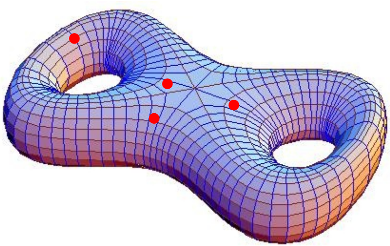
## Definition

Let $X_j$, $j = 1, 2, \ldots, N$ be independent, identically distributed (i.i.d.) random variables having distribution $\mu$.
Let $\sigma > 0$, $\beta > 1$ and $\eta_{jk}$ 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 $\eta_{jk}$ and $X_j$ are independent.
We consider noisy measurements

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



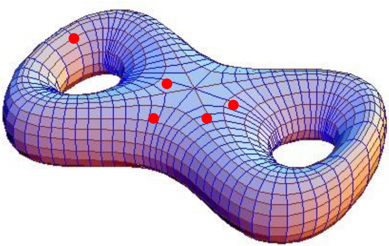
### Definition

Let $X_j$, $j = 1, 2, \dots, N$ be independent, identically distributed (i.i.d.) random variables having distribution $\mu$.
Let $\sigma > 0$, $\beta > 1$ and $\eta_{jk}$ 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 $\eta_{jk}$ 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 $n \geq 2$, $D, K, i_0, c_{min}, c_{max}, \sigma, \beta > 0$ be given. Then there are $\delta_0, C_0$ 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 $1 - \theta$, $\sigma^2$ and the noisy distances $D_{jk} = d_M(X_j, X_k) + \eta_{jk}$, $j, k \leq N$ of $N$ randomly chosen points, where*

$$N \geq C_0 \frac{1}{\delta^{3n}} \left( \log^2(\frac{1}{\theta}) + \log^8(\frac{1}{\delta}) \right),$$

*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 $n \geq 2$, $D, K, i_0, c_{min}, c_{max}, \sigma, \beta > 0$ be given. Then there are $\delta_0$, $C_0$ 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 $1 - \theta$, $\sigma^2$ and the noisy distances $D_{jk} = d_M(X_j, X_k) + \eta_{jk}$, $j, k \leq N$ of $N$ randomly chosen points, where*

$$N \geq C_0 \frac{1}{\delta^{3n}} \left( \log^2(\frac{1}{\theta}) + \log^8(\frac{1}{\delta}) \right),$$

*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 $inj(M)$.*

# 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,

$$\mathbb{P}(Y_{jk} = 1 \,|\, X_j, X_k) = \Phi(X_j, X_k) \tag{2}$$

and there is a smooth non-increasing function $h : [0, \infty) \to [0, 1]$ so that

$$c_1\, h(d_M(x, y)) \leq \Phi(x, y) \leq c_2\, h(d_M(x, y)). \tag{3}$$

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

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

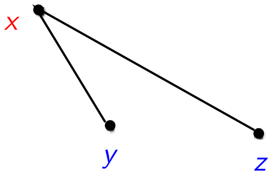
For $y, z \in M$, we consider the "rough distance function"

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

**Lemma**

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

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



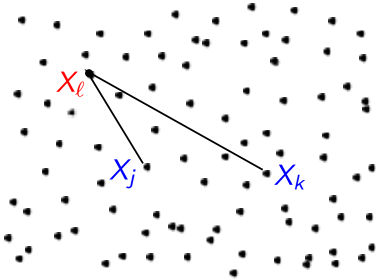
We consider three sets $S_1, S_2, S_3 \subset \{X_j\}$, where $N_i = \#S_i$ satisfy $N_1 > N_2 > N_3$. We call $S_1 = \{X_1, \dots, X_{N_1}\}$ the densest net, $S_2$ the medium dense net and $S_3$ the coarse net.

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" $S_1$.



Let $y, z \in M$, $X$ be a random point on $M$ having the distribution $\mu$, and $\eta, \eta'$ be independent random variables with variance $\sigma^2$. Then

$$\begin{aligned}
&\mathbb{E}\Big((d_M(y,X)+\eta) - (d_M(z,X)+\eta')\Big)^2 \\
&= \mathbb{E}|d_M(y,X) - d_M(z,X)|^2 + \mathbb{E}|\eta - \eta'|^2 \\
&= \int_M |d_M(y,x) - d_M(z,x)|^2 d\mu(x) + 2\sigma^2 \\
&= \|r_y - r_z\|^2_{L^2(M)} + 2\sigma^2.
\end{aligned}$$

This yields for $r_y(x) = d_M(y,x)$ and $D_{j\ell} = d_M(X_j, X_\ell) + \eta_{j\ell}$ the following:

**Lemma**

*Under the condition that $X_j$ and $X_k$ are known, we have*

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

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

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

and

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

Hoeffding's inequality yields the following:

**Lemma**

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

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

## Lemma (Hoeffding's inequality)

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

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

This is a generalization of tail estimates for Gaussian variables:
For independent Gaussian random variables $Y_j \sim N(0,1)$,

$$S = \frac{1}{N}\sum_{j=1}^{N} Y_j, \quad \text{satisfies } \mathbb{E}S^2 = \frac{1}{N}.$$

For $N > \varepsilon^{-2}$, $\mathbb{P}(S < \varepsilon) = \mathbb{P}(Y < N^{1/2}\varepsilon) \ge 1 - e^{-N\varepsilon^2/2}$ as

$$\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt \le \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}\frac{t}{x}dt = \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2}, \quad x > 1.$$

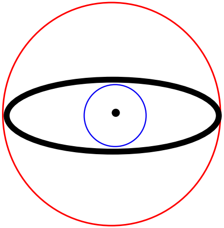
Recall that function $\kappa(y,z)$ is a rough distance function:

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

Let $W(y,\rho)$ be the set

$$W(y,\rho) = \{z \in M : \ \kappa(y,z) < \rho^2\}.$$

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 avaraged distances

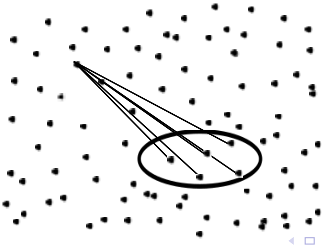
$$d_\rho(y_1, y_2) = \frac{1}{\mu(W(y_1,\rho))} \int_{W(y_1,\rho)} d_M(z, y_2)\, d\mu(z).$$

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

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

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 $S_3 = \{y_1, y_2, \dots, y_{N_3}\}$ 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: We find a smooth manifold $(M^*, g^*)$ using the net $S_3$ and the approximate distance $d_\rho^{app}(y_1, y_2)$ of $y_1, y_2 \in S_3$.

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

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

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

*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:*

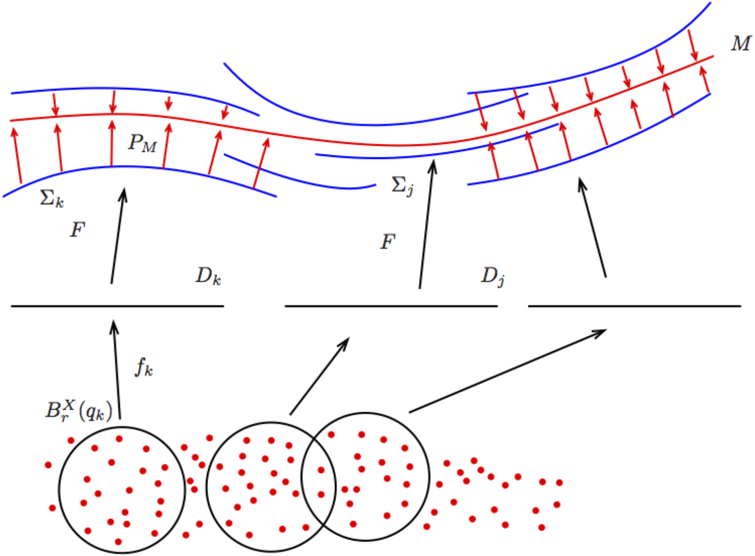
1. *There is a diffeomorphism $F : M^* \to M$ satisfying*

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

2. $|\mathrm{Sec}(M^*)| \le C_n K$.
3. *The injectivity radius $\mathrm{inj}(M^*)$ of $M^*$ satisfies*

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

# Rough idea of the proof of manifold interpolation



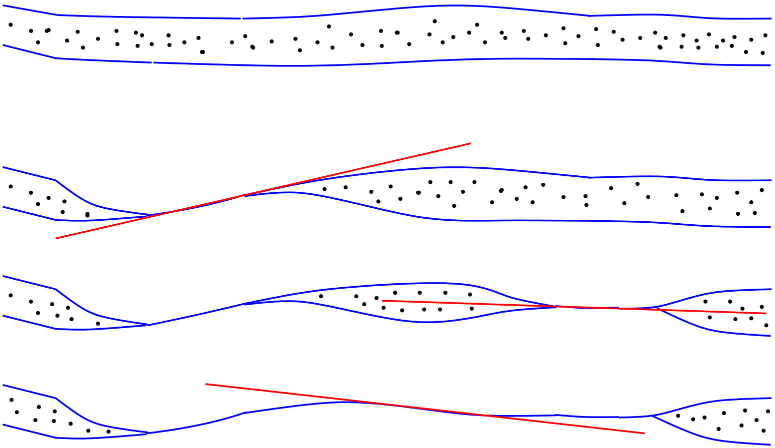
Assume that we are given a finite metric space $(X, d)$.
Let $r = (\delta/K)^{1/3}$ and do following steps:

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 $i = 1, 2, \dots, J$ and construct a $\delta$-isometry $f_i : B_1^X(q_i) \to D_i$.
3. For all $q_i, q_j \in X_0$ such that $d(q_i, q_j) < 1$, find affine transition maps $A_{ij} : \mathbb{R}^n \to \mathbb{R}^n$, such that $A_{ij}(p_i + y) = p_j + L_{ij}y$ and
$$|A_{ij}(f_i(x)) - f_j(x)| < C\delta, \quad \text{for } x \in B_1^X(q_i) \cap B_1^X(q_j).$$
4. Let $\Phi \in C_0^\infty(\mathbb{R}^n)$ be 1 near zero, and $\Omega = \bigcup_i D_i$.
Define a map $F_j : \Omega \to \mathbb{R}^{n+1}$ as follows: For $x \in D_i$, put
$$F_j(y) = \begin{cases} \big(\ \varphi_{ij}(y) \cdot L_{ij}(y)\ ,\ \varphi_{ij}(y)\,\big), & \text{if } d(q_i, q_j) < 1, \\ 0, & \text{otherwise}, \end{cases}$$
where $\varphi_{ij}(y) = \Phi(L_{ij}(y))$.
5. Denote $E = \mathbb{R}^m$, $m = (n+1)J$ and define an embedding
$$F : \Omega \to E, \quad F(y) = (F_j(y))_{j=1}^J.$$

6. Construct the local patches $\Sigma_i = F(D_i)$.
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 metric tensor on $M$ by pushing forward the Euclidean metric $g^e$ on $D_i$ in the maps $P_M \circ F$ and compute a weighted average of the obtained metric tensors.

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

# Interpolation of surface in $\mathbb{R}^m$ from data points



# Surface interpolation

**Theorem**

*Let $E$ be a separable Hilbert space, $n \in \mathbb{Z}_+$, $\delta < \delta_0(n)$, and $r = K\delta^{1/2}$*
*Suppose that $X \subset E$ and for all $x \in X$, there is an n-dimensional affine plane $A_x$ such that*

$$dist_H(X \cap B^E(x,r), A_x \cap B^E(x,r)) < \delta.$$

*Then there exists a closed 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 $C_n K$.*
3. *The normal injectivity radius of $M$ is at least $r/3$.*

Algorithm SurfaceInterpolation: Let $X \subset E = \mathbb{R}^d$ is finite and $r = K\delta^{1/2}$. We implement the following steps:

1. Construct a maximal $\frac{r}{100}$-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 $X \cap B_r(q_i)$ near $q_i$. Let $P_i \colon E \to E$ be orthogonal projectors onto $A_i$.
3. Let $\psi \in C_0^\infty([-\frac{r}{2}, \frac{r}{2}])$ be 1 in $[0, \frac{r}{3}]$ 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 $f \colon E \to E$ 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 $M \subset E$.

Thank you for your attention!