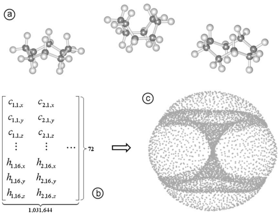
Rates of Convergence for Geometric Inference

Eddie Aamari

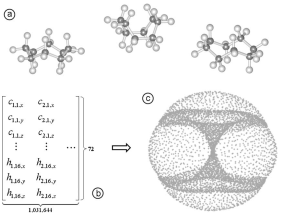
September 1^{st}, 2017

Data with a Global Geometric Structure

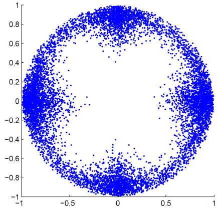


Cyclo-octane conformations
[Martin *et al.* 2010]

Data with a Global Geometric Structure

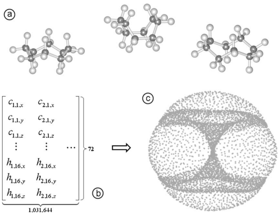


Cyclo-octane conformations
[Martin *et al.* 2010]

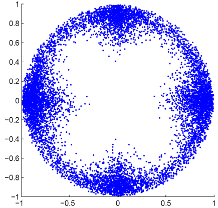


7×7 pixels patch space
[Xia 2016]

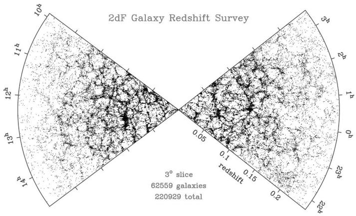
Data with a Global Geometric Structure



Cyclo-octane conformations
[Martin *et al.* 2010]

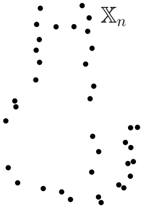


7×7 pixels patch space
[Xia 2016]

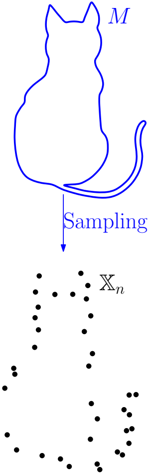


Large Scale Galaxy Structures
[2dF Galaxy Redshift Survey]

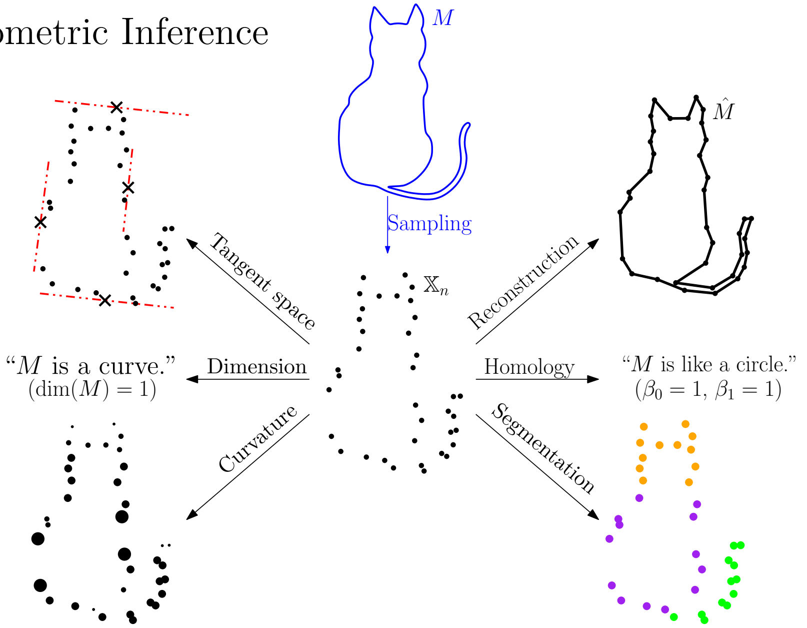
Geometric Inference



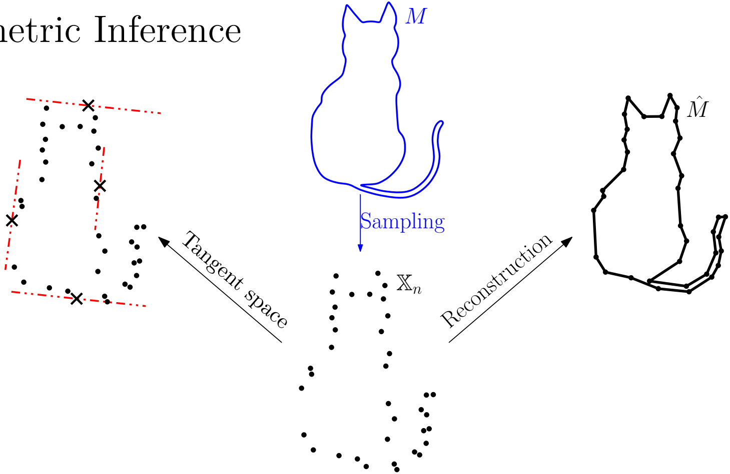
Geometric Inference



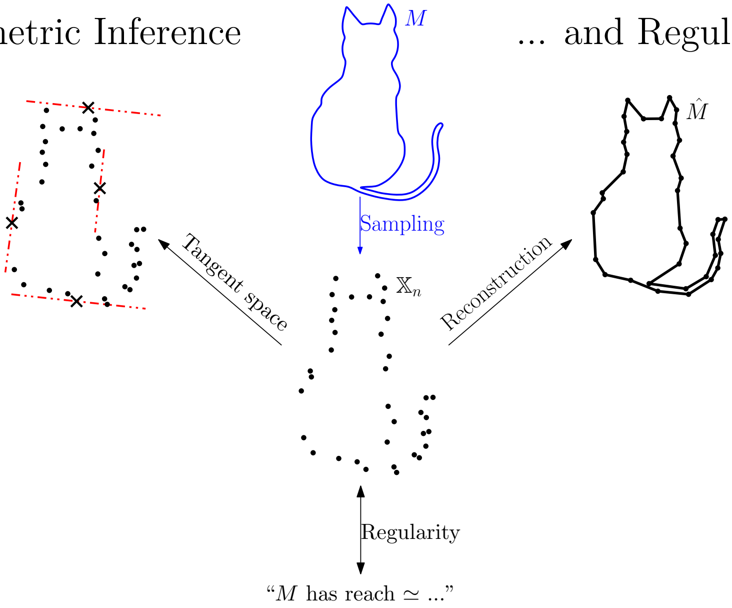
Geometric Inference



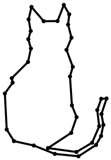
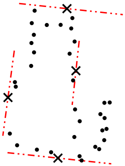
Geometric Inference



Geometric Inference ... and Regularity

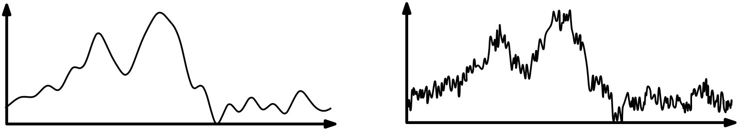


Outline

Part I	**Part II**	**Part III**
		"M has reach $\simeq$ ..."
Reconstruction	Tangent space	Regularity

Given a n-sample $\mathbb{X}_n = \{X_1, \ldots, X_n\}$, what precision can we expect?

Regularity in Function Spaces

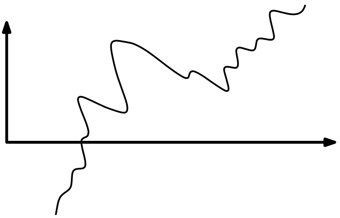


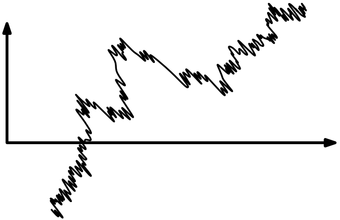
Usual regularity classes (Hölder, Sobolev, Besov) control increments

$$\|f(x) - f(y)\| \leq L \|x - y\|^{\beta} .$$

(L, β) drives the difficulty of the statistical problem.

Regularity in Function Spaces





Usual regularity classes (Hölder, Sobolev, Besov) control increments

$$\|f(x) - f(y)\| \leq L \|x - y\|^{\beta} .$$

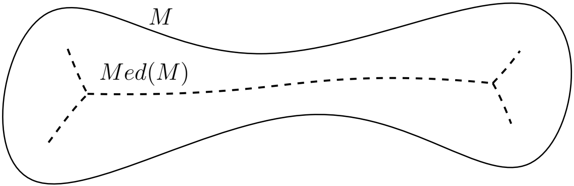
(L, β) drives the difficulty of the statistical problem.

$\rightarrow$ Without natural coordinates, "$\|f(x) - f(y)\|$" = ?

Medial Axis

The **medial axis** of $M \subset \mathbb{R}^D$ is

$$Med(M) = \{z \in \mathbb{R}^D, z \text{ has several nearest neighbors on } M\}.$$



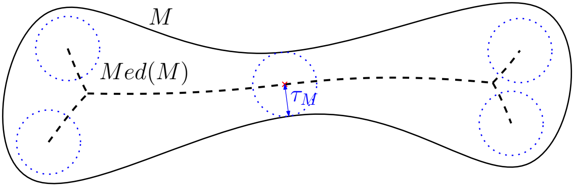
Medial axis of a curve $M \subset \mathbb{R}^2$.

Reach

For a closed subset $M \subset \mathbb{R}^D$, the **reach** τ_M of M is the least distance to its medial axis.

$$\tau_M = \inf_{x \in M} d\left(x, Med(M)\right),$$

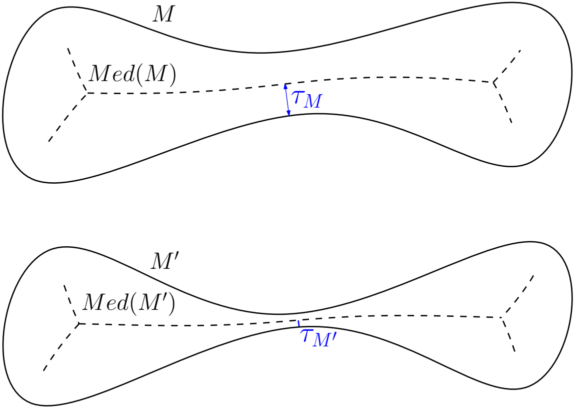
where $d(x, A) = \inf_{a \in A} \|x - a\|$ for all $x \in \mathbb{R}^D$.



One can also flip the formula:

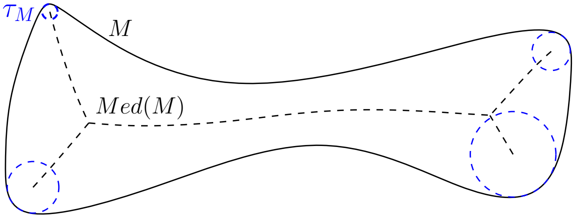
$$\tau_M = \inf_{z \in Med(M)} d\left(z, M\right).$$

Global Regularity



The smaller τ_M, the tighter a bottleneck structure is possible.

Local Regularity



High curvature $\equiv$ Small radius of curvature $\Rightarrow \tau_M \ll 1$.

Proposition (Federer 1959, Niyogi *et al.* 2006)

The sectional curvatures κ of M satisfy

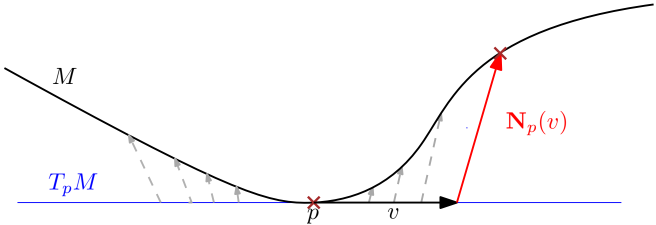
$$-2/\tau_M^2 \leq \kappa \leq 1/\tau_M^2.$$

Local Regularity

If $\tau_M \geq \tau_{\min} > 0$, M has local parametrizations of the form

$$\begin{aligned} \Psi_p : T_pM &\longrightarrow M \subset \mathbb{R}^D \\ v &\longmapsto p + v + \mathbf{N}_p(v) \end{aligned}$$

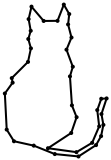
where $\mathbf{N}_p(0) = 0$ and $\left\| d_v \mathbf{N}_p \right\|_{op} \leq \|v\| / (2\tau_{\min})$.



Definition (Regularity class $\mathcal{C}^2_{\tau_{\min}}$)

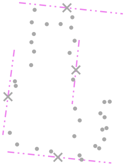
Let $\mathcal{C}^2_{\tau_{\min}}$ be the set of d-dimensional compact connected submanifolds $M \subset \mathbb{R}^D$ with $\tau_M \geq \tau_{\min} > 0$.

Part I



Reconstruction

Part II



Tangent space

Part III

"M has reach $\simeq$ ..."

Regularity

Tangential Delaunay Complex

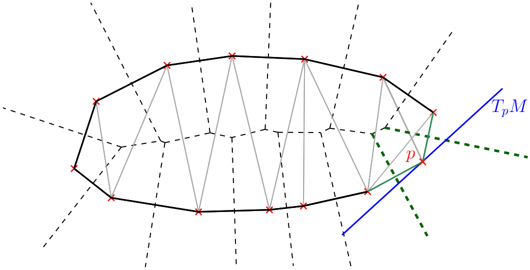
Theorem (Boissonnat,Ghosh 2014)

If $M \in \mathcal{C}^2_{\tau_{\min}}$, *for all* $\varepsilon \le c_d \tau_{\min}$, *if a point cloud* $\mathcal{X} \subset M$ *is*

- 2ε*-dense:* $d_H(M, \mathcal{X}) \le 2\varepsilon$,
- ε*-sparse:* $d(p, \mathcal{X} \setminus \{p\}) \ge \varepsilon$ *for all* $p \in \mathcal{X}$,

there exists a triangulation $\hat{M}_{\mathtt{TDC}}(\mathcal{X}, T_{\mathcal{X}}M)$ *of* $\mathcal{X}$ *such that:*

- M *and* $\hat{M}_{\mathtt{TDC}}$ *are isotopic*
- $d_H\left(M, \hat{M}_{\mathtt{TDC}}\right) \le C_d \varepsilon^2 / \tau_{\min}$.



Tangential Delaunay Complex

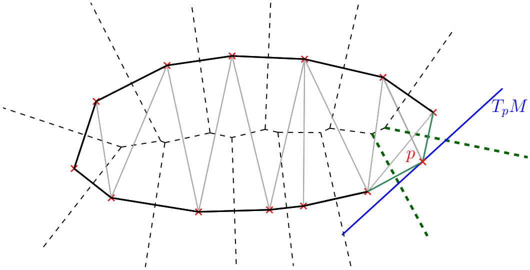
Theorem (Boissonnat,Ghosh 2014)

If $M \in \mathcal{C}^2_{\tau_{\min}}$, for all $\varepsilon \leq c_d \tau_{\min}$, if a point cloud $\mathcal{X} \subset M$ is

- *2ε-dense:* $d_H(M, \mathcal{X}) \leq 2\varepsilon$,
- *ε-sparse:* $d(p, \mathcal{X} \setminus \{p\}) \geq \varepsilon$ *for all* $p \in \mathcal{X}$,

there exists a triangulation $\hat{M}_{\mathtt{TDC}}(\mathcal{X}, T_{\mathcal{X}}M)$ of $\mathcal{X}$ such that:

- *M and $\hat{M}_{\mathtt{TDC}}$ are isotopic*
- $d_H\big(M, \hat{M}_{\mathtt{TDC}}\big) \leq C_d \varepsilon^2 / \tau_{\min}$.

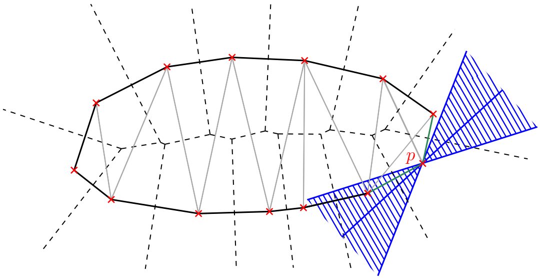


Stability

Theorem (A.,Levrard, 2016)

The result still holds with:

- ***approximate tangent spaces:*** *For all $p \in \mathcal{X}$, we use $\hat{T}_p$ instead of T_pM, with $\angle(T_pM, \hat{T}_p) \lesssim \varepsilon$.*

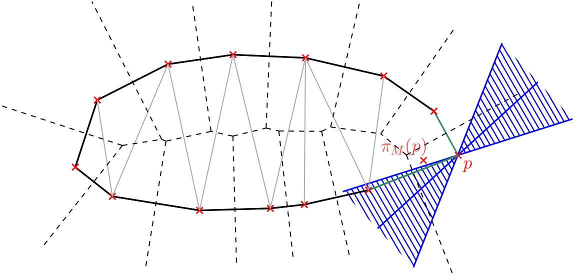


Stability

Theorem (A.,Levrard, 2016)

The result still holds with:

- ***approximate tangent spaces:*** *For all* $p \in \mathcal{X}$*, we use* $\hat{T}_p$ *instead of* T_pM*, with* $\angle(T_pM, \hat{T}_p) \lesssim \varepsilon$.
- ***small noise:*** *For all* $p \in \mathcal{X}, d(p, M) \lesssim \varepsilon^2$.



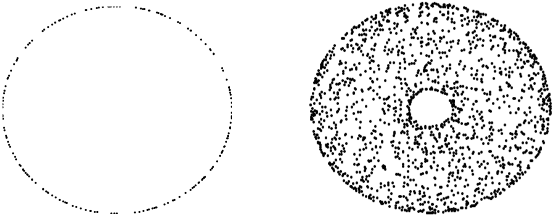
Statistical Model $\mathcal{P}^2_{\tau_{\min}}$

We let $\mathcal{P}^2_{\tau_{\min}}$ denote the set of distributions P such that:

- $M = Supp(P) \in \mathcal{C}^2_{\tau_{\min}}$,
- P has a density f with respect to the uniform measure on M, with

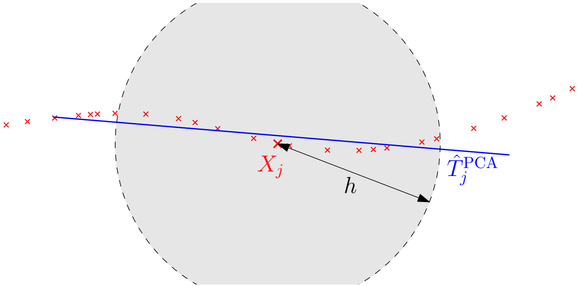
$$0 < f_{\min} \leq f(x) \leq f_{\max} < \infty.$$

We observe an i.i.d. n-sample $\mathbb{X}_n = \{X_1, \ldots, X_n\}$ of some $P \in \mathcal{P}^2_{\tau_{\min}}$.



Same model studied in [Genovese *et al.* 2012].

Tangent Space Estimation: Local PCA



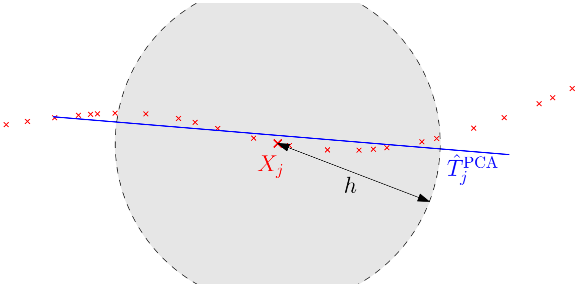
Writing $P_n^{(j)}$ for the integration with respect to $\frac{1}{n}\sum_{\ell \neq j} \delta_{X_\ell - X_j}$, define

$$\hat{T}_j^{\texttt{PCA}} \in \underset{T \in \mathbb{G}^{D,d}}{\operatorname{argmin}} P_n^{(j)} \left[\| x - \pi_T(x) \|^2 \mathbb{1}_{B(0,h)}(x) \right].$$

$\mathbb{G}^{D,d}$: space of d-dimensional linear subspaces of $\mathbb{R}^D$;

π_T : orthogonal projection onto T.

Tangent Space Estimation: Local PCA



Proposition (A., Levrard 2016)

Taking $h \asymp \left(\frac{\log n}{n}\right)^{1/d}$*, for* n *large enough, with high probability,*

$$\max_{1 \leq j \leq n} \angle(T_{X_j} M, \hat{T}_j^{\mathtt{PCA}}) \leq ch \quad \text{and} \quad d_H\left(M, \mathbb{X}_n\right) \leq Ch.$$

Convergence Rate

Theorem (A., Levrard 2016)

For a sparsified subsample $\mathbb{Y}_n \subset \mathbb{X}_n$, $\hat{M}_{\texttt{TDC}}\left(\mathbb{Y}_n, \hat{T}^{\texttt{PCA}}\right)$ *satisfies*

$$\lim_{n\to\infty} \mathbb{P}\left(d_H(M, \hat{M}_{\texttt{TDC}}) \leq c\left(\frac{\log n}{n}\right)^{2/d} \text{ and } M \cong \hat{M}_{\texttt{TDC}} \right) = 1,$$

where $\cong$ *denotes the isotopy equivalence.*
Moreover, for n *large enough,*

$$\mathbb{E}_{P^n} d_H(M, \hat{M}_{\texttt{TDC}}) \leq C\left(\frac{\log n}{n}\right)^{2/d}.$$

c, C *depend only on* $d, \tau_{\min}, f_{\min}, f_{\max}$ *(not on* D*).*

Minimax Optimality of the TDC

- We have

$$\sup_{P \in \mathcal{P}^2_{\tau_{\min}}} \mathbb{E}_{P^n} d_H(M, \hat{M}_{\mathtt{TDC}}) \leq C \left(\frac{\log n}{n} \right)^{2/d},$$

  for $C = C_{d, \tau_{\min}, f_{\min}, f_{\max}}$.
- From [Kim & Zou 2015],

$$\inf_{\hat{M}} \sup_{P \in \mathcal{P}^2_{\tau_{\min}}} \mathbb{E}_{P^n} d_H(M, \hat{M}) \geq c \left(\frac{\log n}{n} \right)^{2/d},$$

  where the infimum ranges over all the estimators $\hat{M} = \hat{M}(\mathbb{X}_n)$.

Minimax Optimality of the TDC

- We have

$$\sup_{P \in \mathcal{P}^2_{\tau_{\min}}} \mathbb{E}_{P^n} d_H(M, \hat{M}_{\texttt{TDC}}) \leq C \left(\frac{\log n}{n} \right)^{2/d},$$

  for $C = C_{d,\tau_{\min},f_{\min},f_{\max}}$.
- From [Kim & Zou 2015],

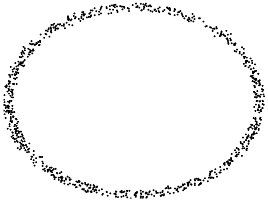
$$\inf_{\hat{M}} \sup_{P \in \mathcal{P}^2_{\tau_{\min}}} \mathbb{E}_{P^n} d_H(M, \hat{M}) \geq c \left(\frac{\log n}{n} \right)^{2/d},$$

  where the infimum ranges over all the estimators $\hat{M} = \hat{M}(\mathbb{X}_n)$.

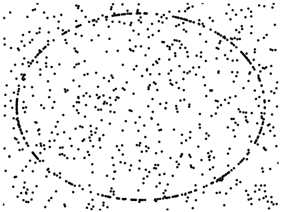
→ $\hat{M}_{\texttt{TDC}}$ is minimax optimal.

→ The minimax rate is achieved by computable triangulations.

Extensions to Noisy Data



Circle with noise amplitude $\sigma > 0$.



Circle with portion $1 - \beta > 0$ outliers.

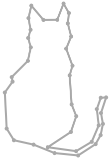
Tangent spaces:
Local PCA.

Noise:
negigible if $\sigma \ll 1$.

Tangent spaces:
Local PCA.

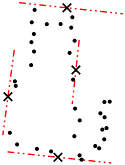
Noise:
Remove outliers far away from M.

Part I



Reconstruction

Part II



Tangent space

Part III

"M has reach $\simeq$ ..."

Regularity

Optimality of Tangent Space Estimation?

In the process, we have shown that

$$\sup_{P \in \mathcal{P}^2_{\tau_{\min}}} \mathbb{E}_{P^n} \max_{1 \le j \le n} \angle \left(T_{X_j} M, \hat{T}_j^{\texttt{PCA}} \right) \le C \left(\frac{\log n}{n} \right)^{1/d},$$

for $C = C_{d,\tau_{\min},f_{\min},f_{\max}}$.

Optimality of Tangent Space Estimation?

In the process, we have shown that

$$\sup_{P \in \mathcal{P}^2_{\tau_{\min}}} \mathbb{E}_{P^n} \max_{1 \leq j \leq n} \angle \left(T_{X_j} M, \hat{T}_j^{\texttt{PCA}} \right) \leq C \left(\frac{\log n}{n} \right)^{1/d},$$

for $C = C_{d,\tau_{\min},f_{\min},f_{\max}}$.

→ Can we do better if $M = Supp(P)$ is more regular than $\mathcal{C}^2_{\tau_{\min}}$?

→ Is this rate optimal?

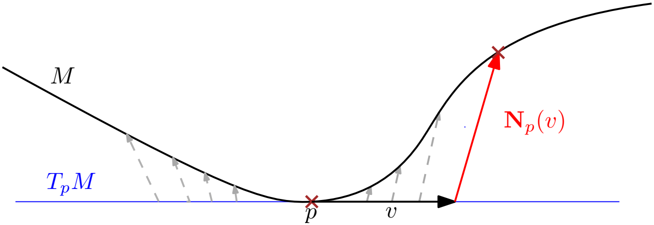
More Regularity

Reminder (Regularity Class $\mathcal{C}^2_{\tau_{\min}}$)

Submanifolds $M \in \mathcal{C}^2_{\tau_{\min}}$ have local parametrizations

$$\begin{aligned} \Psi_p : T_pM &\longrightarrow M \subset \mathbb{R}^D \\ v &\longmapsto p + v + \mathbf{N}_p(v) \end{aligned}$$

where $\mathbf{N}_p(0) = 0$, $d_0\mathbf{N}_p = 0$ *and* $\|d_v\mathbf{N}_p\|_{op} \leq \|v\| / (2\tau_{\min})$.



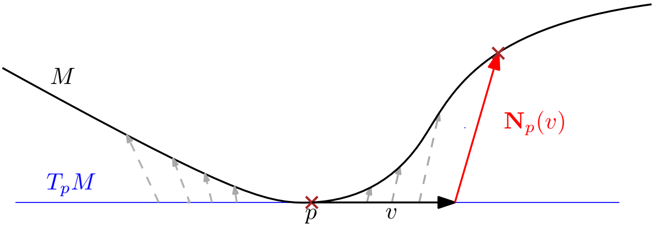
More Regularity

Definition (Model for $\mathcal{C}^k$ submanifolds, $k \geq 3$)

Let $\mathbf{L} = (L_2, L_3, \ldots, L_k)$*, and define* $\mathcal{C}^k_{\tau_{\min},\mathbf{L}}$ *to be the subset of elements* $M \in \mathcal{C}^2_{\tau_{\min}}$ *that have local parametrizations*

$$\begin{aligned} \Psi_p : T_pM &\longrightarrow M \subset \mathbb{R}^D \\ v &\longmapsto p + v + \mathbf{N}_p(v) \end{aligned}$$

where $\mathbf{N}_p(0) = 0$*,* $d_0\mathbf{N}_p = 0$ *and* $\left\| d^i_v \mathbf{N}_p \right\|_{op} \leq L_i$ *for* $2 \leq i \leq k$*.*



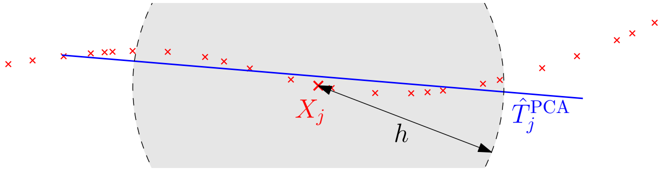
Local PCA

Recall that $P_n^{(j)} = \frac{1}{n} \sum_{\ell \neq j} \delta_{X_\ell - X_j}$, and

$$\hat{T}_j^{\mathtt{PCA}} \in \operatorname*{argmin}_{T \in \mathbb{G}^{D,d}} P_n^{(j)} \left[\left\| x - \pi_T(x) \right\|^2 \mathbb{1}_{B(0,h)}(x) \right].$$

$\mathbb{G}^{D,d}$: space of d-dimensional linear subspaces of $\mathbb{R}^D$;

π_T : orthogonal projection onto T.



Local Polynomials

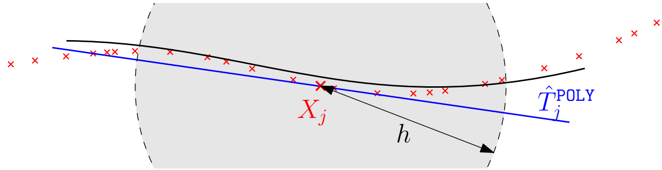
We define $\left(\hat{T}_j^{\texttt{POLY}}, \hat{T}_{2,j}, \ldots, \hat{T}_{k-1,j}\right)$ to be a minimizer of

$$P_n^{(j)}\left[\left\|x - \pi_T(x) - \sum_{i=2}^{k-1} T^{(i)}\left(\pi_T(x)^{\otimes i}\right)\right\|^2 \mathbb{1}_{B(0,h)}(x)\right],$$

where

T: ranges in $\mathbb{G}^{D,d}$;

$T^{(i)}$: ranges in the set of i-linear maps $(2 \leq i \leq k-1)$.



Similar method in [Cazals 2005] and [Cheng & Chiu 2016].

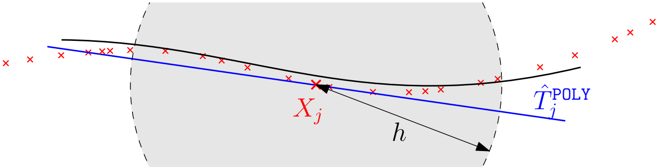
Convergence of Local Polynomials

Theorem (A., Levrard 2017)

If $h = C\left(\frac{\log n}{n}\right)^{1/d}$*, for all* $P \in \mathcal{P}^k_{\tau_{\min},\mathbf{L}}$*,*

$$\mathbb{E}_{P^n} \max_{1 \leq j \leq n} \angle\left(T_{X_j}M, \hat{T}_j^{\mathtt{POLY}}\right) \leq C\left(\frac{\log n}{n}\right)^{\frac{k-1}{d}},$$

where $C = C_{k,d,\tau_{\min},\mathbf{L},f_{\min},f_{\max}}$.



Minimax Risk for Tangent Space Estimation

The minimax risk over the model $\mathcal{P}^k_{\tau_{\min},\mathbf{L}}$ is

$$R_n^{(k)} \triangleq \inf_{\hat{T}} \sup_{P \in \mathcal{P}^k_{\tau_{\min},\mathbf{L}}} \mathbb{E}_{P^n} \angle(T_{X_1}M, \hat{T}),$$

where the infimum ranges over all the estimators $\hat{T} = \hat{T}(\mathbb{X}_n)$.

Theorem (A., Levrard 2017)

For n large enough,

$$R_n^{(k)} \leq C \left(\frac{\log n}{n} \right)^{\frac{k-1}{d}},$$

where $C = C_{d,\tau_{\min},\mathbf{L},f_{\min},f_{\max}}$.

Minimax Risk for Tangent Space Estimation

The minimax risk over the model $\mathcal{P}^k_{\tau_{\min},\mathbf{L}}$ is

$$R_n^{(k)} \triangleq \inf_{\hat{T}} \sup_{P \in \mathcal{P}^k_{\tau_{\min},\mathbf{L}}} \mathbb{E}_{P^n} \angle(T_{X_1}M, \hat{T}),$$

where the infimum ranges over all the estimators $\hat{T} = \hat{T}(\mathbb{X}_n)$.

Theorem (A., Levrard 2017)

For n large enough, (+ technical assumptions)

$$c\left(\frac{1}{n}\right)^{\frac{k-1}{d}} \leq R_n^{(k)} \leq C\left(\frac{\log n}{n}\right)^{\frac{k-1}{d}},$$

where $C = C_{d,\tau_{\min},\mathbf{L},f_{\min},f_{\max}}$ and $c = c_{d,\tau_{\min},\mathbf{L},f_{\min},f_{\max}}$.

Pointwise Lower Bound

For $\mathbf{x} \in \mathbb{R}^D$, write $\mathcal{P}^k_{\tau_{\min},\mathbf{L}}(\mathbf{x}) = \{P \in \mathcal{P}^k_{\tau_{\min},\mathbf{L}} | x \in Supp(P)\}$.

$$R_n^{(k)}(\mathbf{x}) = \inf_{\hat{T}} \sup_{P \in \mathcal{P}^k_{\tau_{\min},\mathbf{L}}(x)} \mathbb{E}_{P^n} \angle(T_{\mathbf{x}} M, \hat{T})$$

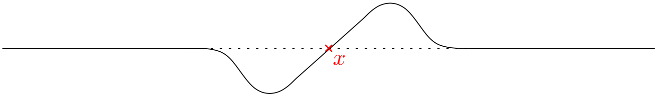
Pointwise Lower Bound

For $\mathbf{x} \in \mathbb{R}^D$, write $\mathcal{P}^k_{\tau_{\min},\mathbf{L}}(\mathbf{x}) = \left\{ P \in \mathcal{P}^k_{\tau_{\min},\mathbf{L}} | x \in Supp(P) \right\}$.

$$R_n^{(k)}(\mathbf{x}) = \inf_{\hat{T}} \sup_{P \in \mathcal{P}^k_{\tau_{\min},\mathbf{L}}(x)} \mathbb{E}_{P^n} \angle(T_{\mathbf{x}} M, \hat{T}) \geq c \left(\frac{1}{n} \right)^{\frac{k-1}{d}}.$$

Outline of the proof:

- Le Cam's lemma;
- bumped hypotheses.



Random Point Lower Bound

$$R_n^{(k)} = \inf_{\hat{T}} \sup_{P \in \mathcal{P}^k_{\tau_{\min}, \mathbf{L}}} \mathbb{E}_{P^n} \angle(T_{\mathbf{X_1}} M, \hat{T})$$

$$= \inf_{\hat{T}} \sup_{P \in \mathcal{P}^k_{\tau_{\min}, \mathbf{L}}} \mathbb{E}_{P^{n-1}} \left[\left\| \angle(T_{\mathbf{x_1}} M, \hat{T}) \right\|_{L^1(P(\mathbf{dx_1}))} \right]$$

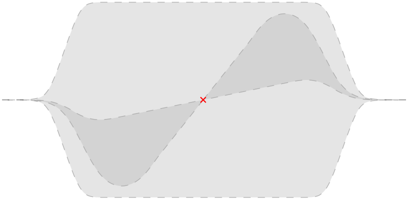
Random Point Lower Bound

$$R_n^{(k)} = \inf_{\hat{T}} \sup_{P \in \mathcal{P}^k_{\tau_{\min},\mathbf{L}}} \mathbb{E}_{P^n} \angle(T_{\mathbf{X_1}} M, \hat{T})$$

$$= \inf_{\hat{T}} \sup_{P \in \mathcal{P}^k_{\tau_{\min},\mathbf{L}}} \mathbb{E}_{P^{n-1}} \left[\left\| \angle(T_{\mathbf{x_1}} M, \hat{T}) \right\|_{L^1(P(\mathbf{dx_1}))} \right] \geq c \left(\frac{1}{n} \right)^{\frac{k-1}{d}} .$$

Outline of the proof:

- conditional Assouad's lemma (condition on X_1);
- thick (mixture) bumped hypotheses.



Further Use of Local Polynomial

Curvature: letting II_p^M be the 2[nd] fundamental form of M at p,

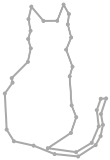
$$\mathbb{E}_{P^n} \max_{1 \leq j \leq n} \left\| II_{X_j}^M \circ \pi_{T_{X_j} M} - \hat{T}_{2,j} \circ \pi_{\hat{T}_j^{\texttt{POLY}}} \right\|_{op} \leq C \left(\frac{\log n}{n} \right)^{\frac{k-2}{d}} .$$

Support: using all the polynomials $\hat{T}_j^{\texttt{POLY}}, \hat{T}_{2,j}, \ldots, \hat{T}_{k-1,j}$, we get

$$\mathbb{E}_{P^n} d_H\big(M, \hat{M}_{\texttt{POLY}}\big) \leq C' \left(\frac{\log n}{n} \right)^{\frac{k}{d}} .$$

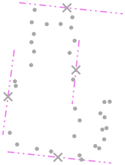
$\longrightarrow$ These rates are minimax optimal (up to a $\log n$ factor).

Part I



Reconstruction

Part II



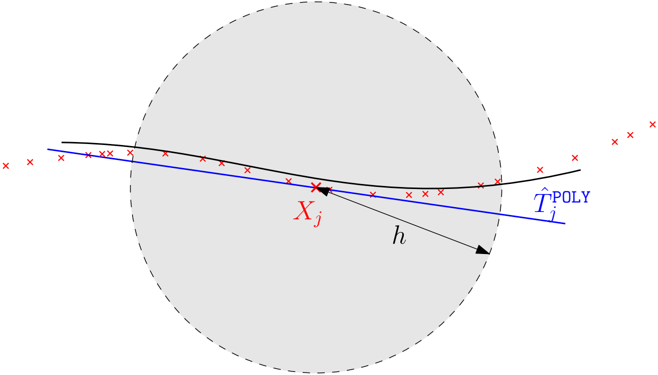
Tangent space

Part III

“M has reach $\simeq$ ...”

Regularity

Back to the Reach



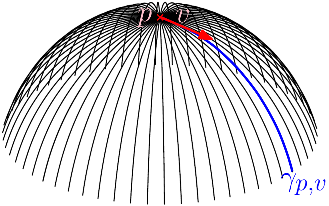
Goal: Estimate the reach τ_M.

Geodesic Regularity for Reach Estimation

Definition (Regularity Class $\mathcal{C}^{(3)}_{\tau_{min},L}$)

For $L > 0$, we let $\mathcal{C}^{(3)}_{\tau_{min},L}$ denote the subset of elements $M \in \mathcal{C}^2_{\tau_{min}}$ for which all unit speed geodesics $\gamma_{p,v}$ satisfy

$$\left\|\gamma'''_{p,v}(t)\right\| \leq L.$$



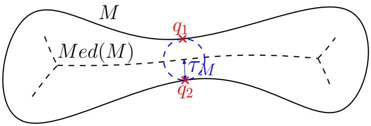
Remark: We have $\mathcal{C}^3_{\tau_{min},\mathbf{L}'} \subset \mathcal{C}^{(3)}_{\tau_{min},L} \subset \mathcal{C}^2_{\tau_{min}}$.

Global and Local Reach

Theorem (A., Kim *et al.* 2017)

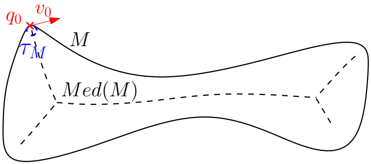
If M is a closed $\mathcal{C}^3$-submanifold, either

- ***(Global case)*** *M has a bottleneck*



or

- ***(Local case)*** *M has an arc-length parametrized geodesic such that $\left\|\gamma''_{q_0,v_0}(0)\right\| = \frac{1}{\tau_M}$.*

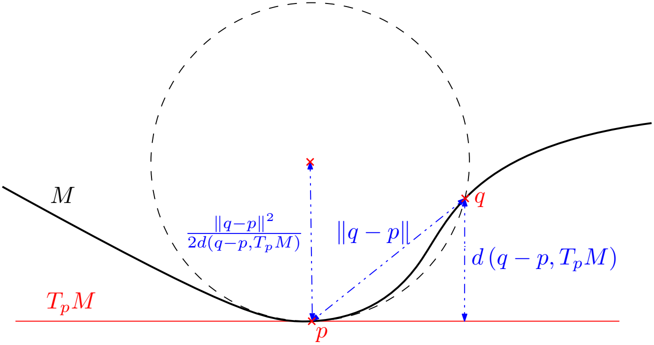


A (Crucial) Local Formulation

Proposition (Federer 1959)

For all closed submanifold $M \subset \mathbb{R}^D$,

$$\tau_M = \inf_{p \neq q \in M} \frac{\|q-p\|^2}{2d\left(q-p, T_pM\right)}.$$



A (Crucial) Local Formulation

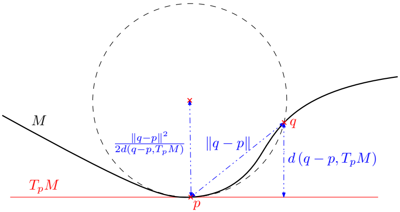
Proposition (Federer 1959)

For all closed submanifold $M \subset \mathbb{R}^D$,

$$\tau_M = \inf_{p \neq q \in M} \frac{\|q-p\|^2}{2d\left(q-p, T_pM\right)}.$$

Plugin estimator: for any point cloud $\mathcal{X} \subset M$,

$$\hat{\tau}\left(\mathcal{X}\right) = \min_{x \neq y \in \mathcal{X}} \frac{\|y-x\|^2}{2d(y-x, T_xM)}.$$

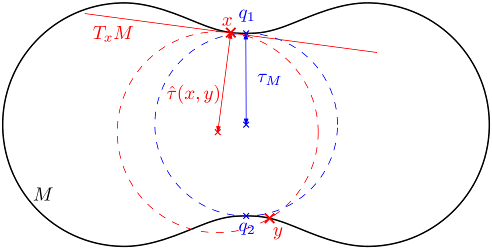


Global Case

Proposition (A., Kim *et al.* 2017)

Assume that $M \in \mathcal{C}^{(3)}_{\tau_{\min},L}$ *has a bottleneck* $q_1, q_2 \in M$.
If $x, y \in \mathcal{X}$ *are close to* q_1, q_2,

$$|\tau_M - \hat{\tau}(\mathcal{X})| \lesssim \max \{\|x - q_1\|, \|y - q_2\|\}.$$



Global Case (Upper Bound)

Corollary (A., Kim *et al.* 2017)

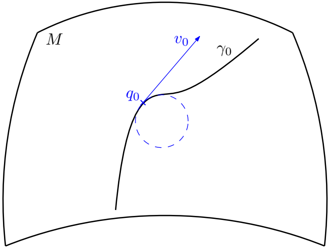
Let $P \in \mathcal{P}^{(3)}_{\tau_{min},L}$ *and* $M = Supp(P)$*. If* M *has a bottleneck, then*

$$\mathbb{E}_{P^n} |\tau_M - \hat{\tau}(\mathbb{X}_n)| \leq C \left(\frac{1}{n}\right)^{\frac{1}{d}},$$

where $C = C_{d,\tau_{\min},f_{\min}}$.

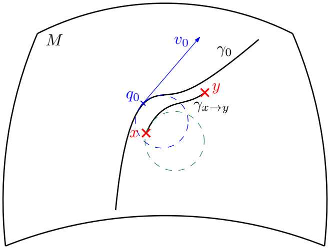
Local Case

Assume there exist $q_0 \in M$ and $v_0 \in T_{q_0}M$ with $\left\|\gamma''_{q_0,v_0}(0)\right\| = 1/\tau_M$.



Local Case

Assume there exist $q_0 \in M$ and $v_0 \in T_{q_0}M$ with $\left\|\gamma''_{q_0,v_0}(0)\right\| = 1/\tau_M$.

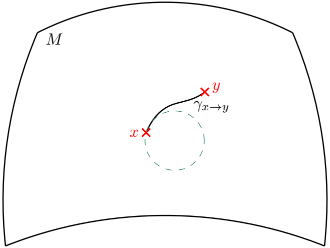


- (Directional Curvature Stability) If x, y are close to q_0 and $\theta = \angle(y - x, v_0)$ is small,

$$\left\|\gamma''_{x\to y}(0)\right\| \simeq \left\|\gamma''_{q_0,v_0}(0)\right\| = 1/\tau_M.$$

Local Case

Assume there exist $q_0 \in M$ and $v_0 \in T_{q_0}M$ with $\left\|\gamma''_{q_0,v_0}(0)\right\| = 1/\tau_M$.

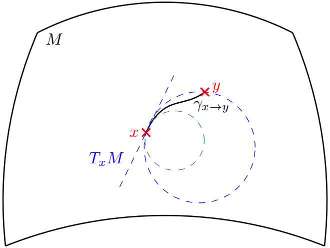


- (Directional Curvature Stability) If x, y are close to q_0 and $\theta = \angle(y - x, v_0)$ is small,

$$\left\|\gamma''_{x\to y}(0)\right\| \simeq \left\|\gamma''_{q_0,v_0}(0)\right\| = 1/\tau_M.$$

Local Case

Assume there exist $q_0 \in M$ and $v_0 \in T_{q_0}M$ with $\left\|\gamma''_{q_0,v_0}(0)\right\| = 1/\tau_M$.



- (Directional Curvature Stability) If x, y are close to q_0 and $\theta = \angle(y - x, v_0)$ is small,

$$\left\|\gamma''_{x\to y}(0)\right\| \simeq \left\|\gamma''_{q_0,v_0}(0)\right\| = 1/\tau_M.$$

- (Directional Curvature Estimation) If $\|y - x\|$ is small,

$$\frac{\left\|y - x\right\|^2}{2d\left(y - x, T_xM\right)} \simeq \left\|\gamma''_{x\to y}(0)\right\|.$$

Local Case

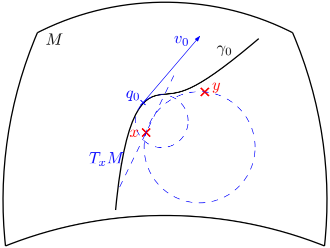
Proposition (A., Kim *et al.* 2017)

Assume that $M \in \mathcal{C}^{(3)}_{\tau_{\min},L}$ *has* $\left\|\gamma''_{q_0,v_0}(0)\right\| = \frac{1}{\tau_M}$, *and write*

$$\theta_z = \angle(v_0, \gamma'_{q_0 \to z}(0)).$$

If $x, y \in \mathcal{X}$ *are close to* q_0, *and* $|\theta_x - \theta_y| \geq \pi/2$,

$$|\tau_M - \hat{\tau}(\mathcal{X})| \lesssim \max\left\{\|x - q_0\|, \|y - q_0\|\right\} + \sin^2(|\theta_x - \theta_y|).$$



Local Case (Upper Bound)

Corollary (A., Kim *et al.* 2017)

Let $P \in \mathcal{P}^{(3)}_{\tau_{min},L}$ *and* $M = Supp(P)$*. Assume that* M *has a geodesic such that* $\|\gamma_0''(0)\| = \frac{1}{\tau_M}$*. Then,*

$$\mathbb{E}_{P^n} |\tau_M - \hat{\tau}(\mathbb{X}_n)| \leq C \left(\frac{1}{n}\right)^{\frac{2}{3d-1}},$$

where $C = C_{d,\tau_{\min},L,f_{\min}}$.

Overall Minimax Estimates

The minimax risk over $\bar{\mathcal{P}}^{(3)}_{\tau_{min},L}$ for reach estimation is,

$$\inf_{\hat{\tau}_n} \sup_{\bar{P} \in \bar{\mathcal{P}}^{(3)}_{\tau_{min},L}} \mathbb{E}_{\bar{P}^n} \left| \tau_{\bar{P}} - \hat{\tau}_n \right|,$$

where the infimum ranges over all the estimators $\hat{\tau}_n = \hat{\tau}_n \left(\mathbb{X}_n, T_{\mathbb{X}_n} M \right)$.

Proposition (A., Kim *et al.* 2017)

For n large enough,

$$\inf_{\hat{\tau}_n} \sup_{\bar{P} \in \bar{\mathcal{P}}^{(3)}_{\tau_{min},L}} \mathbb{E}_{\bar{P}^n} \left| \tau_{\bar{P}} - \hat{\tau}_n \right| \leq C \left(\frac{1}{n} \right)^{\frac{2}{3d-1}},$$

where $C = C_{d,\tau_{\min},L,f_{\min}}$.

Overall Minimax Estimates

The minimax risk over $\bar{\mathcal{P}}^{(3)}_{\tau_{min},L}$ for reach estimation is,

$$\inf_{\hat{\tau}_n} \sup_{\bar{P} \in \bar{\mathcal{P}}^{(3)}_{\tau_{min},L}} \mathbb{E}_{\bar{P}^n} \left| \tau_{\bar{P}} - \hat{\tau}_n \right|,$$

where the infimum ranges over all the estimators $\hat{\tau}_n = \hat{\tau}_n \left(\mathbb{X}_n, T_{\mathbb{X}_n} M \right)$.

Proposition (A., Kim *et al.* 2017)

For n large enough, (+ technical assumptions)

$$c \left(\frac{1}{n} \right)^{\frac{1}{d}} \leq \inf_{\hat{\tau}_n} \sup_{\bar{P} \in \bar{\mathcal{P}}^{(3)}_{\tau_{min},L}} \mathbb{E}_{\bar{P}^n} \left| \tau_{\bar{P}} - \hat{\tau}_n \right| \leq C \left(\frac{1}{n} \right)^{\frac{2}{3d-1}},$$

where $C = C_{d,\tau_{\min},L,f_{\min}}$ and $c = c_{\tau_{min}}$.

Conclusion

In this PhD thesis, we examined:

- Minimax estimates for estimation of $M, T_X M, II_X^M$ and τ_M.
- Stability of manifold reconstruction algorithms.
- Regularity classes for submanifolds.
- Lower bound techniques for $T_X M$ with random X.

Conclusion

In this PhD thesis, we examined:

- Minimax estimates for estimation of $M, T_X M, II_X^M$ and τ_M.
- Stability of manifold reconstruction algorithms.
- Regularity classes for submanifolds.
- Lower bound techniques for $T_X M$ with random X.

Which raised new questions for:

- Noise handling
- Broader regularity classes
    - including boundary;
    - based on other parameters λ-reach, μ-reach, local feature size.
- Effective implementation and calibration.

Summing Up

Minimax Risk Estimates.

Goal \ Model	$\mathcal{C}^2_{\tau_{\min}}$	$\mathcal{C}^k_{\tau_{\min},\mathbf{L}}$	$\mathcal{C}^{(3)}_{\tau_{\min},L}$ $T_X M$ known
M	$\left(\frac{\log n}{n}\right)^{\frac{2}{d}}$	$\left(\frac{[\log n]}{n}\right)^{\frac{k}{d}}$	
$T_X M$	$\left(\frac{[\log n]}{n}\right)^{\frac{1}{d}}$	$\left(\frac{[\log n]}{n}\right)^{\frac{k-1}{d}}$	0
II^M_X	$c > 0$	$\left(\frac{[\log n]}{n}\right)^{\frac{k-2}{d}}$	
τ_M	$c > 0$		$\left(\frac{1}{n}\right)^{\frac{1}{d}-[\alpha]}$ $0 \leq \alpha \leq \frac{d-1}{d(3d-1)}$

[ ] indicates a gap between upper and lower bounds.
No constant depend on D.

Summing Up

Minimax Risk Estimates.

Goal \ Model	$\mathcal{C}^2_{\tau_{\min}}$	$\mathcal{C}^k_{\tau_{\min},\mathbf{L}}$	$\mathcal{C}^{(3)}_{\tau_{\min},L}$ $T_X M$ known
M	$\left(\frac{\log n}{n}\right)^{\frac{2}{d}}$	$\left(\frac{[\log n]}{n}\right)^{\frac{k}{d}}$	$\left(\frac{[\log n]}{n}\right)^{\frac{3}{d}-[\alpha]}$ $0 \leq \alpha \leq \frac{1}{d}$
$T_X M$	$\left(\frac{[\log n]}{n}\right)^{\frac{1}{d}}$	$\left(\frac{[\log n]}{n}\right)^{\frac{k-1}{d}}$	0
II^M_X	$c > 0$	$\left(\frac{[\log n]}{n}\right)^{\frac{k-2}{d}}$	$\left(\frac{[\log n]}{n}\right)^{\frac{1}{d}-[\alpha]}$ $0 \leq \alpha \leq \frac{1}{d}$
τ_M	$c > 0$	?	$\left(\frac{1}{n}\right)^{\frac{1}{d}-[\alpha]}$ $0 \leq \alpha \leq \frac{d-1}{d(3d-1)}$

[ ] indicates a gap between upper and lower bounds.
No constant depend on D.