Simplicial Manifold Reconstruction via Tangent Space Estimation

Eddie Aamari & Clément Levrard

04/20/2015

Outline

- Building the triangulation
  - Weighted Delaunay complex
  - Restricted complex
  - A stability result for angles
- Statistical model and estimation procedure
  - Tangent space estimation *via* local PCA
  - Convergence rate
- A noisy model
  - Clustering *via* local PCA
  - Interpolation
  - Convergence rate

General Framework

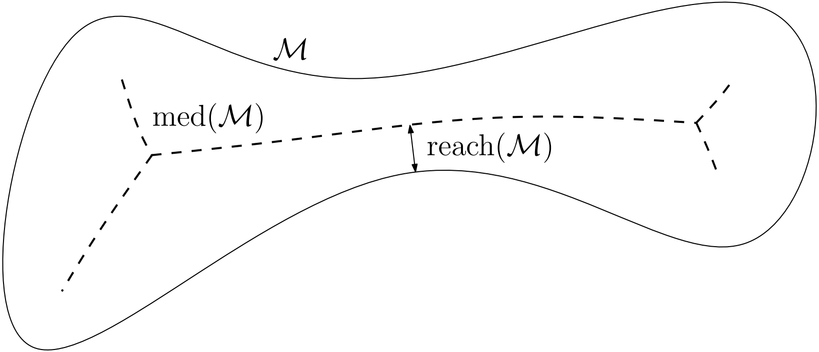
$\mathcal{M} \subset \mathbb{R}^D$ a d-dimensional sub-manifold.

The *reach* of $\mathcal{M}$ is the minimal distance to its medial axis,

$$\mathrm{reach}(X) = \inf_{x \in X} \mathrm{d}(x, \mathrm{med}(X)),$$

where

$\mathrm{med}(X) = \{p \in \mathbb{R}^D, \exists a \neq b \in \mathbb{R}^D, \|p - a\| = \|p - b\| = \mathrm{d}(p, X)\}.$



Reach Condition

Assume $\mathrm{reach}(\mathcal{M}) \geq \rho$ for some fixed $\rho > 0$

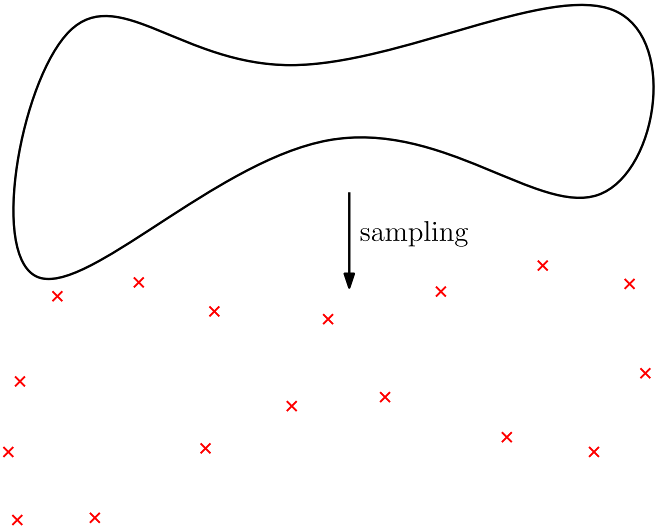


Figure: Reach and sampling

Reach Condition

Assume $\mathrm{reach}(\mathcal{M}) \geq \rho$ for some fixed $\rho > 0$

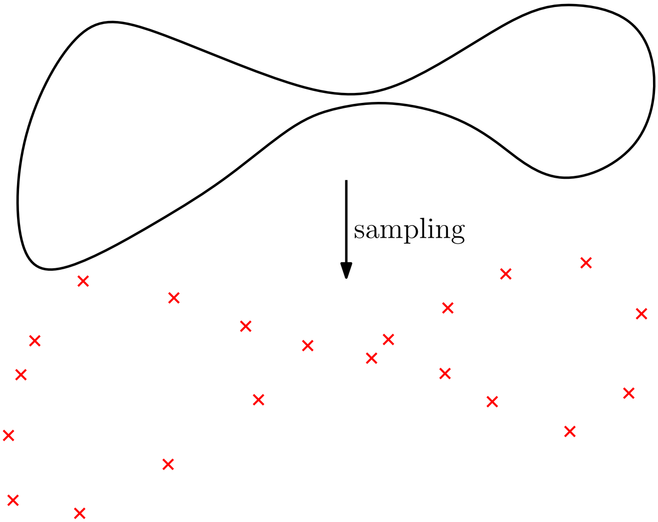


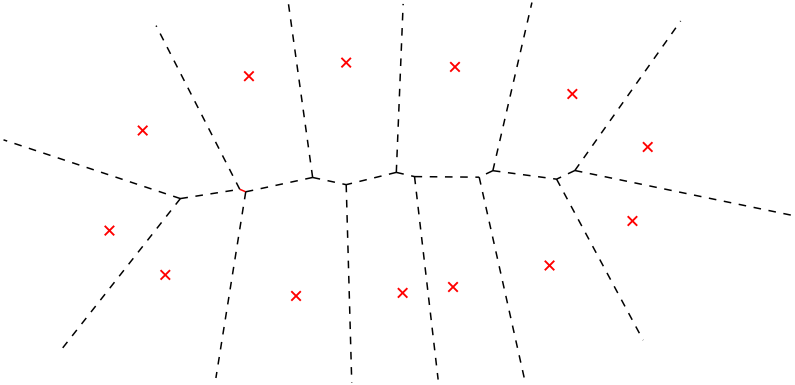
Figure: Reach and sampling

Weighted Delaunay Complex

For $\mathcal{P} = \{x_1, \ldots, x_N\} \subset \mathbb{R}^D$ and a weight $\omega : \mathcal{P} \longrightarrow [0, \infty)$, consider the weighted Voronoï $\mathrm{Vor}^\omega(\mathcal{P})$, where the cell of $p \in \mathcal{P}$ is

$$\mathrm{Vor}^\omega(p) = \{x \in \mathbb{R}^D : \|x - p\|^2 - \omega(p)^2 \leq \|x - q\|^2 - \omega(q)^2, \forall q \in \mathcal{P}\}.$$

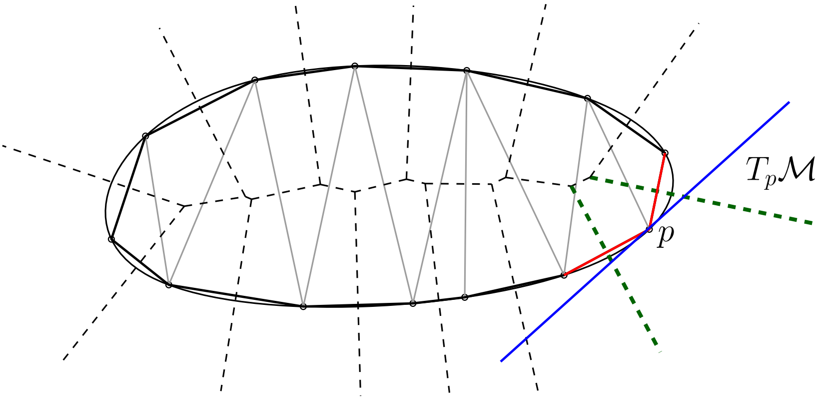
The *weighted Delaunay complex* $\mathrm{Del}^\omega(\mathcal{P})$ is its dual triangulation.



Tangential complex

The tangential complex $\mathrm{Del}^{\omega}(\mathcal{P}, T)$

$$\tau \in \mathrm{Del}^{\omega}(\mathcal{P}, T) \Leftrightarrow \mathrm{Vor}^{\omega}(\tau) \cap \left(\bigcup_{p \in \tau} T_p\mathcal{M} \right) .$$



A Reconstruction Theorem

Theorem (Boissonnat,Ghosh 2014)

There exists ε_0 such that for all $\varepsilon \leq \varepsilon_0$, for all $\mathcal{P} \subset \mathcal{M}$, such that

- $\mathrm{d_H}(\mathcal{P}, \mathcal{M}) \leq 2\varepsilon$,
- $\mathrm{d}(p, \mathcal{P} \setminus \{p\}) \geq \epsilon$ *for all* $p \in \mathcal{P}$,

there exists a weight function ω depending only on $\mathcal{P}$ such that:

- $\mathrm{Del}^{\omega}(\mathcal{P}, T)$ *and* $\mathcal{M}$ *are isotopic;*
- $\mathrm{d_H}\left(\mathrm{Del}^{\omega}(\mathcal{P}, T), \mathcal{M}\right) \leq C\varepsilon^2$, *where* $C = C(d)$.

A Reconstruction Theorem

Theorem (Boissonnat,Ghosh 2014)

There exists ε_0 such that for all $\varepsilon \leq \varepsilon_0$, for all $\mathcal{P} \subset \mathcal{M}$, such that

- $\mathrm{d_H}(\mathcal{P}, \mathcal{M}) \leq 2\varepsilon$,
- $\mathrm{d}(p, \mathcal{P} \setminus \{p\}) \geq \epsilon$ *for all* $p \in \mathcal{P}$,

there exists a weight function ω depending only on $\mathcal{P}$ such that:

- $\mathrm{Del}^\omega(\mathcal{P}, T)$ *and* $\mathcal{M}$ *are isotopic;*
- $\mathrm{d_H}\left(\mathrm{Del}^\omega(\mathcal{P}, T), \mathcal{M}\right) \leq C\varepsilon^2$, *where* $C = C(d)$.

Problem: The $T_p\mathcal{M}$'s are unknown.

$\Rightarrow$ We replace each $T_p\mathcal{M}$ by an estimated version $\tilde{T}_p$.

Stability of the Tangential Complex

Theorem (Stability of the tangential complex)

Let $(\tilde{T}_p)_{p \in \mathcal{M}}$ be a family of affine spaces of dimension d, with $p \in \tilde{T}_p$ and $\angle(\tilde{T}_p, T_p\mathcal{M}) \leq \pi/32$, for all $p \in \mathcal{P}$.
Under the conditions of the reconstruction theorem,

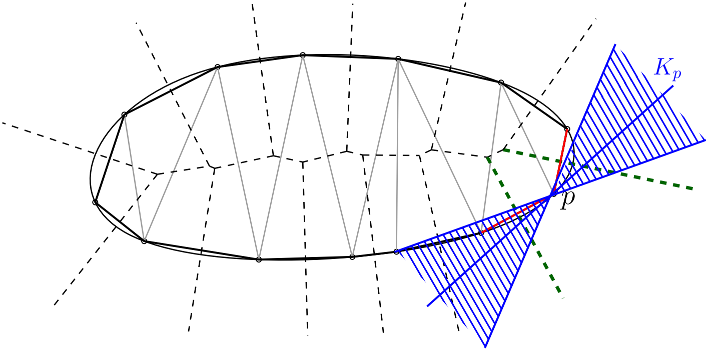
$$\mathrm{Del}^{\omega}(\mathcal{P}, \tilde{T}) = \mathrm{Del}^{\omega}(\mathcal{P}, T).$$

Stability: proof

The cocone complex $\mathrm{Del}^{\omega}(\mathcal{P}, K)$

$$\tau \in \mathrm{Del}^{\omega}(\mathcal{P}, K) \Leftrightarrow \mathrm{Vor}^{\omega}(\tau) \cap \left(\bigcup_{p \in \tau} K_p \right),$$

where $K_p = \{q \in R^D, \angle(pq, T_p\mathcal{M}) \leq \pi/32\}$



$\mathrm{Del}^{\omega}(\mathcal{P}, K)$ has dimension at most d.

Stability: Proof

$\mathrm{Del}^{\omega}(\mathcal{P}, T), \mathrm{Del}^{\omega}(\mathcal{P}, \tilde{T}) \subset \mathrm{Del}^{\omega}(\mathcal{P}, K)$, and if there exists a d-simplex $\tau \in \mathrm{Del}^{\omega}(\mathcal{P}, T)$ but $\tau \notin \mathrm{Del}^{\omega}(\mathcal{P}, \tilde{\tau})$, we get the following configuration:

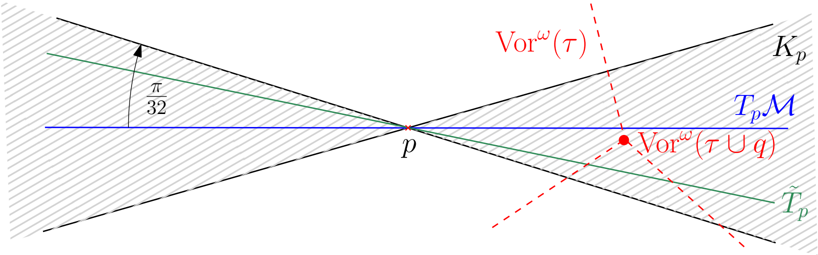


Figure: $\mathrm{Vor}^{\omega}(\tau)$ cannot intersect K_p partially.

Statistical Model

$X_1, \dots, X_n \overset{i.i.d.}{\sim} P$, where $\mathcal{M} = \mathrm{supp}(P) \subset \mathbb{R}^D$ is a connected d-sub-manifold satisfies:

- $\mathcal{M}$ has no boundary,
- $\mathrm{reach}(\mathcal{M}) \geq \rho > 0$,
- P has a twice differentiable density f with respect to μ, the measure induced on $\mathcal{M}$ by the d-dimensional Hausdorff measure over $\mathbb{R}^D$, with:

$$\begin{cases} f_{min} \leq f(x) \leq f_{max} \\ \|\mathrm{d}_x^2 f\| \leq H. \end{cases}$$

Remarks on the Model

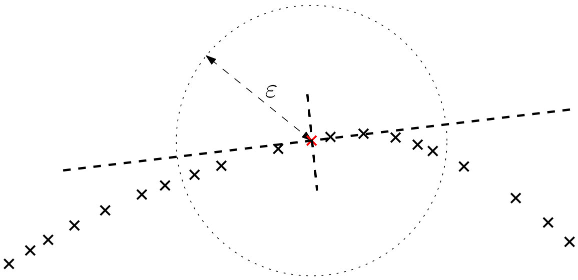
- P is (a, d)-standard: $P(\mathcal{B}(p, r) \geq ar^d$ for small enough r. Therefore,

$$\mathbb{E}\mathrm{d_H}(\mathcal{M}, \{X_1, \ldots, X_n\}) \leq C \left(\frac{\log n}{n}\right)^{1/d} .$$

- The model forces

$$\rho^d f_{min} \leq C_d, \qquad\qquad \mathrm{diam}(\mathcal{M}) \leq \frac{C'_d}{\rho^{d-1} f_{min}}.$$

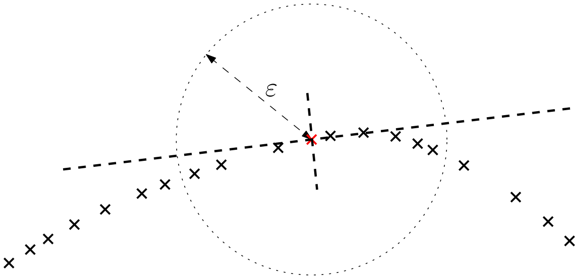
Tangent Space Estimation: Local P.C.A.



Define $\hat{T}_j$ as the span of the d first eigenvectors of

$$\hat{O}_j = \frac{1}{n-1} \sum_{i \neq j} \mathbf{1}_{\|X_i - X_j\| < \varepsilon} (X_i - X_j)(X_i - X_j)^T .$$

Tangent Space Estimation: Local P.C.A.



Theorem

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

$$\max_j \angle(T_{X_j}\mathcal{M}, \hat{T}_j) \leq \frac{\pi}{32}.$$

Estimation Procedure & Convergence Rate

1. Estimate the $T_{X_j}\mathcal{M}$'s with local PCA.
2. Using the *farthest point sampling* algorithm, extract a sparse sample $\mathbb{Y}_n \subset \{X_1, \ldots, X_n\}$.
3. Take as estimator $\hat{\mathcal{M}}$ the weighted Delaunay triangulation of $\mathbb{Y}_n$ restricted to the estimated tangent spaces $\hat{T}_j$'s.

Theorem

$$\lim_{n\to\infty} \mathbb{P}\left(\mathrm{d_H}(\mathcal{M}, \hat{\mathcal{M}}) \leq \left(\frac{\log n}{n}\right)^{2/d} \textit{ and } \mathcal{M} \cong \hat{\mathcal{M}}\right) = 1,$$

where $\cong$ denotes the isotopy equivalence.
Moreover there exists a constant C such that for n large enough,

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

A Noisy Model: Clutter Noise

$$X \sim \beta P + (1-\beta)\mathcal{U},$$

with $0 < \beta < 1$, P as previously and $\mathcal{U} \sim \mathit{Uniform}(\mathcal{B}(0, M))$, for $\mathcal{B}(\mathcal{M}, \rho) \subset \mathcal{B}(0, M)$.

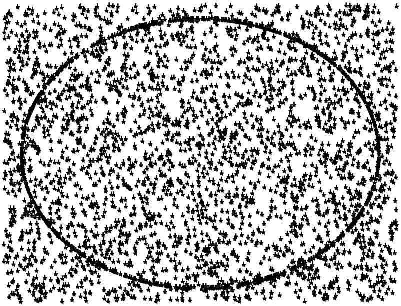
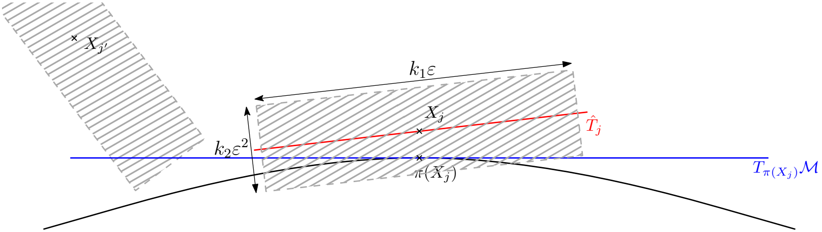


Figure: A realization of the clutter model

Clustering before Estimation

We define boxes S_j centred at each X_j:



To determine if $X_j \in \mathcal{M}$, consider $P_n(S_j) = |S_j \cap \{X_1, \ldots, X_n\}|$.
As $\varepsilon \to 0$,

$$P_n(S_j) \sim \begin{cases} \varepsilon^D & \text{if} \quad X_j \text{ is far from } \mathcal{M} \\ \varepsilon^d \gg \varepsilon^D & \text{if} \quad X_j \in \mathcal{M} \end{cases}$$

Clustering Result

Proposition

There exists constants $k(d,D)$ and $t(d,D,\rho)$ such that, for n large enough, if

$$\varepsilon = k\left(\frac{\log(n)}{\beta n}\right)^{\frac{1}{d+2}},$$

then, with probability larger than $1-\left(\frac{1}{n}\right)^{\frac{2}{d}}-\left(\frac{1}{n}\right)^{2D}$, we have

$$\left(\frac{n}{\log(n)}\right) P_n(S_j) \begin{cases} \leq & t \quad \text{if} \quad d(X_j,\mathcal{M}) \geq \varepsilon^2 \\ > & t \quad \text{if} \quad X_j \in \mathcal{M} \end{cases}$$

Moreover, on the same event, for every X_j such that $d(X_j,\mathcal{M}) \leq \sqrt{\frac{1}{2d+5}}\varepsilon$, we have

$$\angle(\hat{T}_j, T_{\pi(X_j)}\mathcal{M}) \leq \left(\frac{K_1(d,H)}{\rho} + \frac{K_2(d,D)}{\beta}\right)\varepsilon.$$

Reminder: The Reconstruction Theorem

Theorem (Boissonnat,Ghosh 2014)

There exists ε_0 *such that for all* $\varepsilon \leq \varepsilon_0$*, for all* $\mathcal{P} \subset \mathcal{M}$*, such that*

- $\mathrm{d_H}(\mathcal{P}, \mathcal{M}) \leq 2\varepsilon$,
- $\mathrm{d}(p, \mathcal{P} \setminus \{p\}) \geq \epsilon$ *for all* $p \in \mathcal{P}$,

there exists a weight function ω *depending only on* $\mathcal{P}$ *such that:*

- $\mathrm{Del}^{\omega}(\mathcal{P}, T)$ *and* $\mathcal{M}$ *are isotopic;*
- $\mathrm{d_H}\left(\mathrm{Del}^{\omega}(\mathcal{P}, T), \mathcal{M}\right) \leq C\varepsilon^2$, *where* $C = C(d)$.

Interpolation Lemma

Lemma

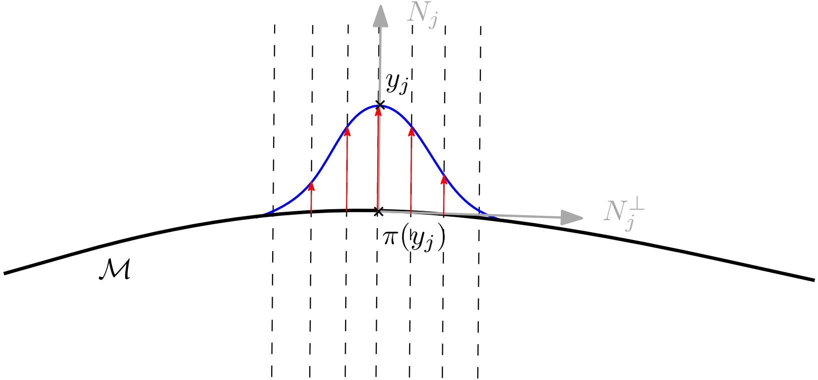
Let $\mathbb{Y} = \{y_1, \ldots, y_q\} \subset \mathbb{R}^D$ *be a sample such that:*

- $\mathbb{Y}$ ε*-approximates* $\mathcal{M}$*:* $\mathrm{d_H}(\mathbb{Y}, \mathcal{M}) < \varepsilon$,
- $\mathbb{Y}$ *is* δ*-sparse:* $\min_{i \neq j} \|y_j - y_i\| \geq \delta > 0$ *for all* j,
- *the* y_j*'s are* η*-close to* $\mathcal{M}$*:* $\max_{1 \leq j \leq q} d(y_j, \mathcal{M}) < \eta$,

Provided that $\eta < \mathrm{reach}(\mathcal{M}) \wedge \delta/25$, *there exists a smooth sub-manifold* $\mathcal{M}' \subset \mathbb{R}^D$ *such that:*

1. $\mathcal{M}'$ *interpolates the points:* $\mathbb{Y} \subset \mathcal{M}'$;
2. $T_{y_j}\mathcal{M}' = (y_j - \pi(y_j)) + T_{\pi(y_j)}\mathcal{M}$ *for all* $1 \leq j \leq q$;
3. $\mathrm{d_H}(\mathbb{Y}, \mathcal{M}') \leq \varepsilon + \eta$;
4. $\mathcal{M}$ *and* $\mathcal{M}'$ *are ambient isotopic;*
5. $reach(\mathcal{M}') \geq \rho \cdot \dfrac{\left(1 - 25\frac{\eta}{\delta}\right)^2}{1 + 1475\frac{\eta}{\delta^2}\rho + 25\frac{\eta}{\delta}}$.

Interpolation Lemma: Proof



Interpolation Lemma

Proposition

Taking

$$\varepsilon \asymp \left(\frac{\log(n)}{n}\right)^{\frac{1}{d+2}}, \qquad \delta = \varepsilon/2, \qquad \eta \asymp \epsilon^2$$

in last lemma, we get for n large enough,

1. $\mathcal{M}'$ *interpolates the points:* $\mathbb{Y} \subset \mathcal{M}'$*;*
2. $T_{y_j}\mathcal{M}' = (y_j - \pi(y_j)) + T_{\pi(y_j)}\mathcal{M}$ *for all* $1 \leq j \leq q$*;*
3. $\mathrm{d_H}(\mathbb{Y}, \mathcal{M}') \leq 2\varepsilon$*;*
4. $\mathcal{M}$ *and* $\mathcal{M}'$ *are ambient isotopic;*
5. *reach*$(\mathcal{M}') \geq C\rho$*.*

Convergence Result

1. Partition the sample into noise/data with slab counting,
2. Using the *farthest point sampling* algorithm, extract a sparse sample $\mathbb{Y}_n \subset \{X_1, \ldots, X_n\}$.
3. Take as estimator $\hat{\mathcal{M}}$ the weighted Delaunay triangulation of $\mathbb{Y}_n$ restricted to the estimated tangent spaces $\hat{T}_j$'s.

Theorem

$$\lim_{n\to\infty} \mathbb{P}\left(\mathrm{d_H}(\mathcal{M}, \hat{\mathcal{M}}) \leq \left(\frac{\log n}{n}\right)^{2/(d+2)} \textit{ and } \mathcal{M} \cong \hat{\mathcal{M}} \right) = 1,$$

where $\cong$ *denotes the isotopy equivalence.*
Moreover there exists a constant C *such that for* n *large enough,*

$$\mathbb{E}\mathrm{d_H}(\mathcal{M}, \hat{\mathcal{M}}) \leq C \left(\frac{\log n}{n}\right)^{2/(d+2)}.$$