Learning about Manifolds from Samples

Uli Wagner

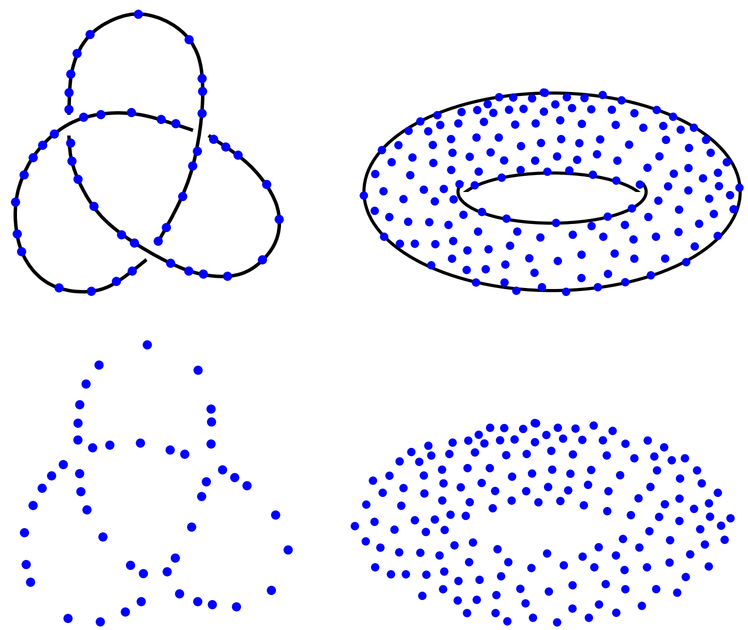
MSRI

Joint work with

Joachim Giesen

ETH Zürich

Mystery Manifolds



Set-up

smooth compact manifold M,
embedded in $\boldsymbol{R}^d$, *unknown to us*

Input

finite set $S \subset M$ of *sample points*,
$|S| =: n$

Goals

Infer properties (topology, geometry)
of M by inspecting S

- with theoretical correctness guarantees
- under assumptions on S that are as weak as possible

Ramifications

$\dim M = 2, d = 3$: Surface Reconstruction

(E.g., data from scan of surface of 3-dimensional object)

Want piecewise linear surface that interpolates the sample points and is homeomorphic and geometrically close (in Hausdorff distance, normal vectors, etc.) to M.

General Intrinsic and Ambient Dimension

Numerous applications (e.g. in speech recognition), often $d \gg \dim M$.

Ideally, would like simplicial complex on S that is homeomorphic and geometrically close to M (*reconstruction*).

Weaker goals include determining *weaker topological invariants*, *dimension*, or t *approximating geodesics* ($\to$ interpolation, low-distortion embeddings)

Focus of this talk

1. Quick Review of Delaunay-based Methods

2. Difficulties in High Dimensions

3. A Few Positive Results in High Dimensions

Ignore Many Important Issues

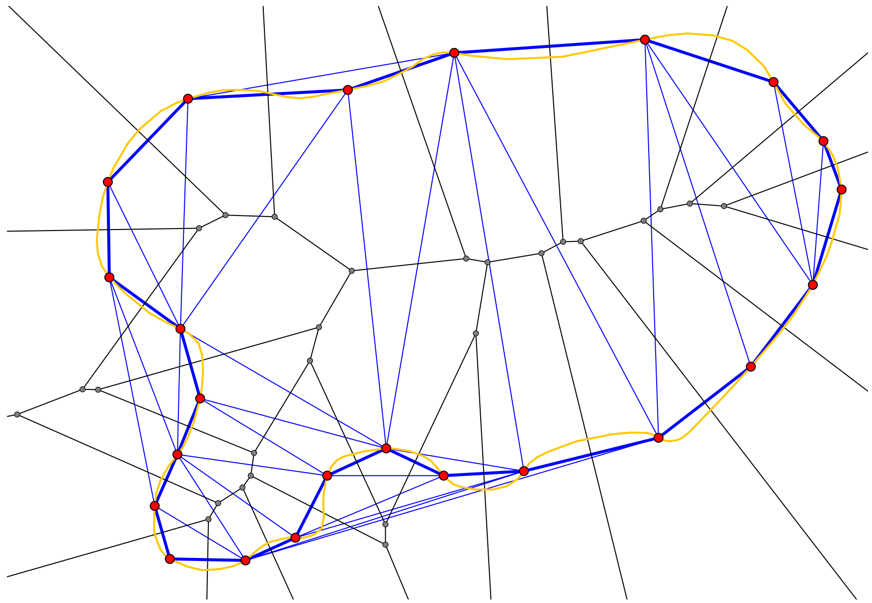
- non-smoothness and boundaries; more general “shapes”
- noise and undersampling
- alternative approaches such as non-linear interpolation or approximation

Restricted Delaunay Triangulation

The Delaunay triangulation of S restricted to M consists of all Delaunay simplices σ whose dual Voronoi object V_σ intersects M.

Closed Ball Property?

For every k-dimensional Delaunay simplex σ of S, $k \leq \dim M$, the intersection $V_\sigma \cap M$ is either empty or a closed ball of dimension $\dim M - k$.



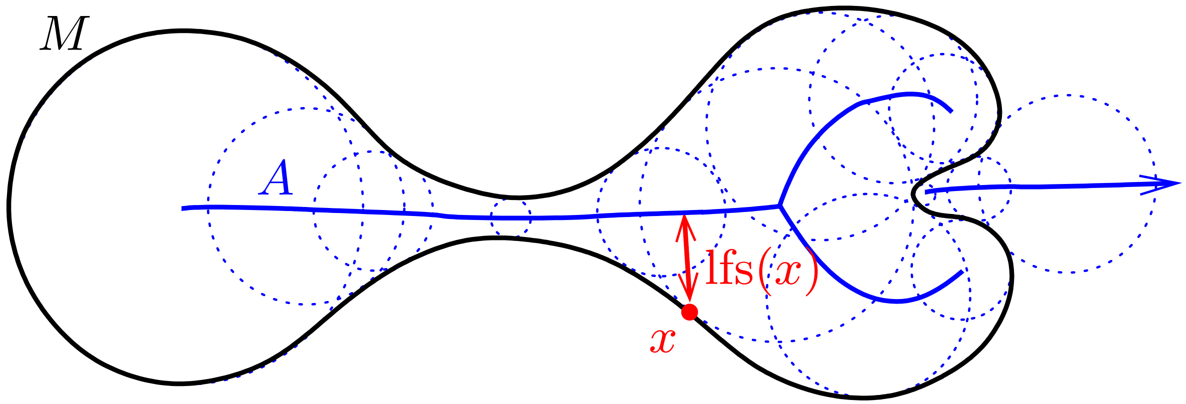
Theorem [Edelsbrunner, Shah '94]

If S has the closed ball property w.r.t. M, then the restricted Delaunay triangulation is homeomorphic to M.

Medial Axis and Local Feature Size

Medial Axis [Blum '67]

$$A = A(M) := \{x \in \boldsymbol{R}^d : \exists \geq 2 \text{ closest points for } x \text{ on } M\}$$



Local Feature Size [Amenta, Bern, Eppstein '98]

$$\mathrm{lfs}(x) := \mathrm{dist}(x, A), \quad x \in M$$

Lipschitz Continuity Lemma

$$\mathrm{lfs}(y) \leq \mathrm{lfs}(x) + \|x - y\|, \quad x, y \in M$$

Sampling Conditions

ε-Sample [ABE '98] Fix $0 < \varepsilon < 1$.

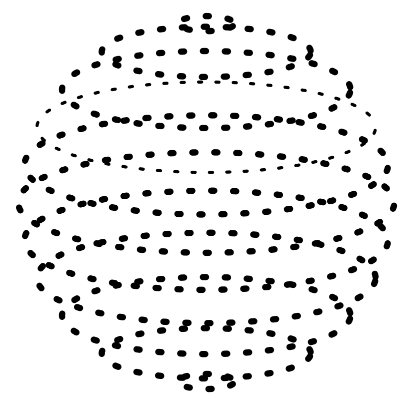
$$S \text{ is an } \varepsilon\text{-sample} :\Leftrightarrow \forall x \in M\ \exists p \in S:\ \|x - p\| \leq \varepsilon \cdot \mathrm{lfs}(x)$$

If $\dim M$ is unknown, we need stronger assumptions to avoid sampling artifacts.

Tight ε-Sample [DGGZ '02]

Fix another constant $0 < \delta < \varepsilon$.
An ε-sample S is an (ε, δ)-sample if

$$\forall p \neq q \in S : \|p - q\| \geq \delta \cdot \mathrm{lfs}(p)$$



Delaunay-Based Surface Reconstruction, 1

Crust [Amenta, Bern '98],
Cocone [Amenta, Choi, Dey, Leekha '00], [Funke, Ramos '02]

Step 1. For every sample point p, compute a vector n_p that well approximates the surface normal at p.

Step 2. Find a collection $\mathcal{T}$ of *candidate* Delaunay triangles such that

1. each abc in $\mathcal{T}$ is almost orthogonal to each of n_a, n_b, and n_c and hence locally almost parallel to M.
2. the circumcircle of each triangle in $\mathcal{T}$ is small, i.e. of size $O(\varepsilon)$ times the local feature size at its corners
3. $\mathcal{T}$ contains the restricted Delaunay triangulation.

Delaunay-Based Surface Reconstruction, 2

Step 3. Clean-Up (delete triangles with “sharp” edges, restricted Delaunay triangles will survive).

Lemma [AB ’98]

If S is an ε-sample from a surface $M \subset \boldsymbol{R}^3$, then the Closed Ball Property is satisfied, and hence the restricted Delaunay triangulation is homeomorphic to M.

Step 4. Manifold Extraction (uses that Delaunay triangles form a geometric simplicial complex, that the candidate triangles can be oriented consistently, and that the restricted Delaunay triangulation is still present)

Higher dimensions

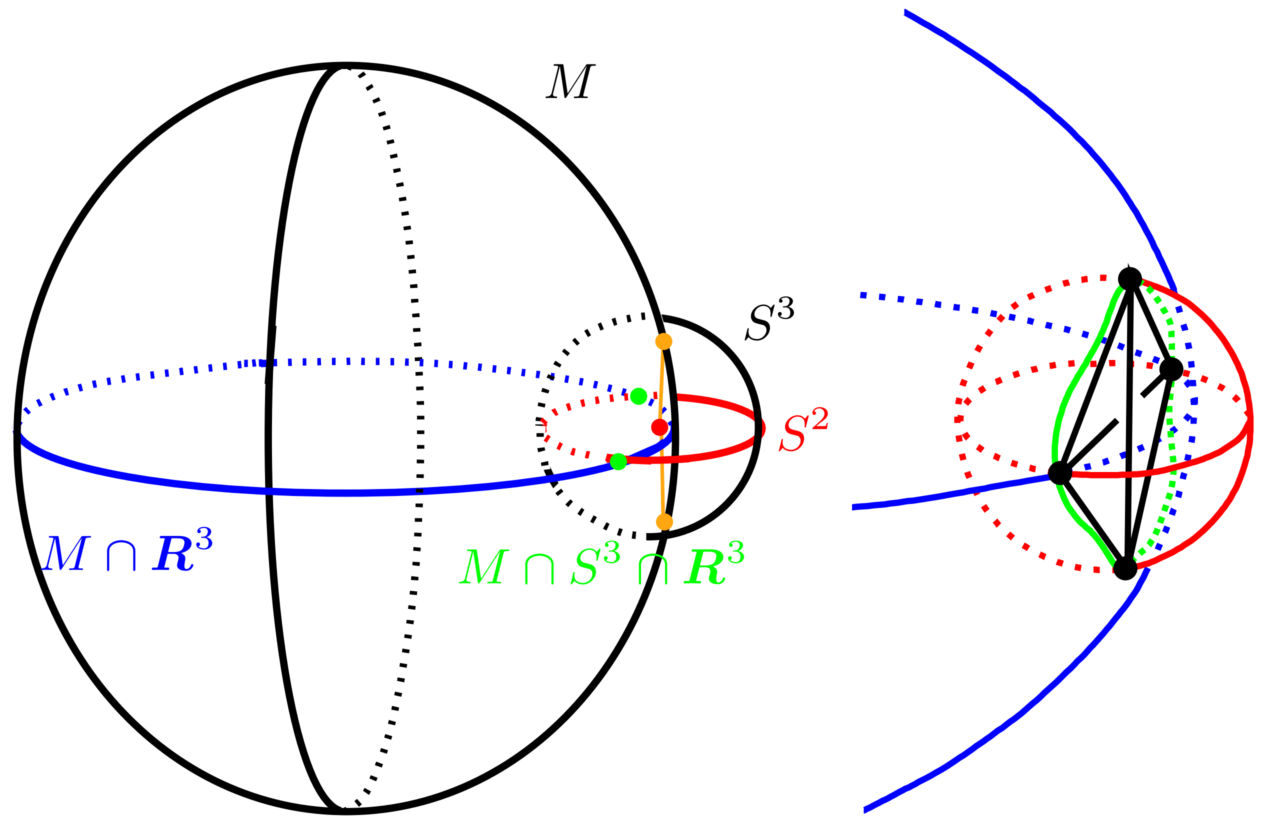
Delaunay-based dimension detection (extension of *Cocone*) [Dey, Giesen, Goswami, Zhao '02]

Drawback: Complexity of Delaunay Triangulation

- A priori, complexity $\Theta(n^{\lceil d/2 \rceil})$.
- There are examples of very uniform samples from the surface of a cylinder in $\boldsymbol{R}^3$ for which the complexity of the Delaunay triangulation is $\Omega(n^{3/2})$ [Erickson '03]
- Even for very uniform samples, the complexity grows exponentially with the codimension.

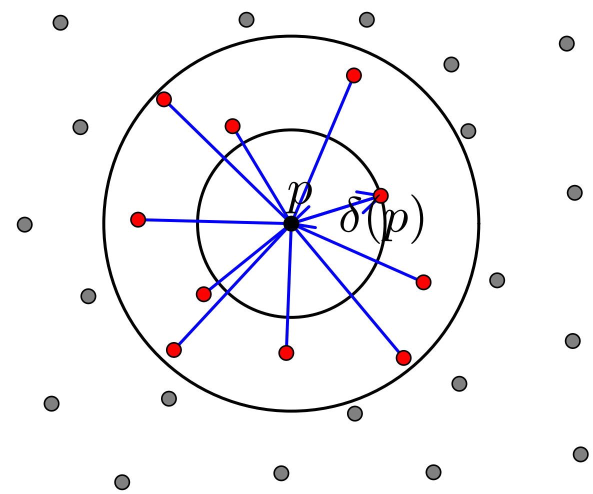
Nasty Restricted Delaunay Slivers

For $\dim M \geq 3$, restricted Delaunay simplices can be *transversal* to M, even in the case of uniform samples.



$\Rightarrow$ Closed Ball Property violated!

Adaptive Neighborhood Graph



c-Neighborhoods

Fix a constant $c > 1$. For $p \in S$, let

$$\delta(p) := \min\{\|p - q\|_2 : q \in S \setminus p\}$$

and

$$N_c(p) := \{q \in S \setminus p : \|p - q\|_2 \le c \cdot \delta(p)\}$$

Adaptive Neighborhood Graph

$$G_c(S) \quad \begin{cases} \text{vertices} \ldots \text{points in } S \\ \quad \text{edges} \ldots \text{line segments } pq \text{ s.t. } q \in N_c(p) \text{ or } p \in N_c(q) \end{cases}$$

Related Approaches

K nearest neigbors; all neighbors within a given radius r

Locally Uniform Samples

$$U_c(p) := \{x \in M : \|p - x\| \leq c \cdot \delta(p)\}, \qquad p \in S$$

Fix a "uniformity parameter" $1 < \rho < c/2$. S is *locally uniform* if

$$\forall p \in S \forall x \in U_c(p) \exists q \in S : \|q - x\| \leq \rho \cdot \delta(p)$$

From now on, assume that S is a locally uniform ε-sample, for suitable (universal) choices of the constants and sufficiently small ε.

For instance, $c = 5$, $\rho = 2$, and $\varepsilon < 1/10$ will work.

Remark

For locally uniform samples, the complexity of the adaptive neighborhood graph is $2^{O(\dim M)} n$.

Connected Components

S a locally uniform ε-sample from M

Theorem

$p, q \in S$ lie in the same connected component of M

$$\Leftrightarrow$$

they are connected by a path in adaptive neighborhood graph G

Proof of "$\Rightarrow$" uses the fact that G contains all restricted Delaunay edges and that these span the connected components of M.

Dimension Detection

S locally uniform ε-sample from M

Small Angle Lemma

$$\operatorname{dist}(q, T_p M) \leq O(\varepsilon^2), \qquad \forall p \in S, q \in N(p)$$

Large Angle Lemma

$$\max_{q \in N(p)} \operatorname{dist}(q, L) \geq \Omega(1), \qquad \forall p \in S, \text{flat L} \ni p, \dim L < \dim M$$

$$\Rightarrow \text{threshold } \beta : \min_{l\text{-dim flat } L \ni p} \max_{q \in N(p)} \operatorname{dist}(q, L) \begin{cases} \leq \beta, \ l \geq \dim M, \\ > 2\beta, \ l < \dim M. \end{cases}$$

Determine $\dim M$ by computing 2-approximation of l-dimensional flat through p that best fits $N(p)$, $l = 1, 2, 3, \ldots$

Can be done in time $d2^{O(k^7 \log k)}n$ [Har-Peled, Varadarajan '02]

Geodesic Distances

S locally uniform ε-sample from M, $p, q \in S$

$\mathrm{dist}_M(p,q)$ = length of a shortest geodesic connecting p and q

$\mathrm{dist}_G(p,q)$ = shortest-path distance in adaptive neighborhood graph, each edge weighted with its euclidean length.

Theorem 1

$$\mathrm{dist}_M(p,q) \leq (1 + O(\varepsilon^2)) \cdot \mathrm{dist}_G(p,q)$$

Theorem 2

$$\mathrm{dist}_G(p,q) \leq (1 + O(1/c)) \cdot d_M(p,q)$$

Note

Denser sample $\Rightarrow$ better approximation to dist_M from below.

However, approximation quality to dist_M from above does not increase with sampling density.