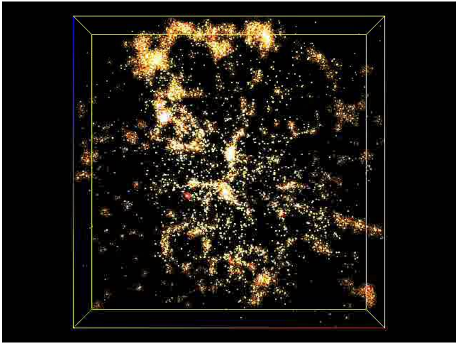
Submanifold Reconstruction

Jean-Daniel Boissonnat
DataShape, INRIA
http://www-sop.inria.fr/datashape

Geometric data analysis

Images, text, speech, neural signals, GPS traces,...



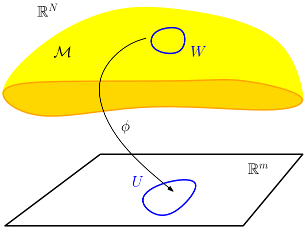
Geometrisation : Data = points + distances between points

Hypothesis : Data lie close to a structure of "small" intrinsic dimension

Problem : Infer the structure from the data

Submanifolds of $\mathbb{R}^d$

A compact subset $\mathbb{M} \subset \mathbb{R}^d$ is a submanifold without boundary of (intrinsic) dimension $k < d$, if any $p \in \mathbb{M}$ has an open (topological) k-ball as a neighborhood in $\mathbb{M}$



Intuitively, a submanifold of dimension k is a subset of $\mathbb{R}^d$ that looks locally like an open set of an affine space of dimension k

A curve a 1-dimensional submanifold
A surface is a 2-dimensional submanifold

Triangulation of a submanifold

We call triangulation of a submanifold $\mathbb{M} \subset \mathbb{R}^d$ a simplicial complex $\hat{\mathbb{M}}$ such that

- $\hat{\mathbb{M}}$ is embedded in $\mathbb{R}^d$
- its vertices are on $\mathbb{M}$
- it is homeomorphic to $\mathbb{M}$

Submanifold reconstruction

The problem is to construct a triangulation $\hat{\mathbb{M}}$ of some unknown submanifold $\mathbb{M}$ given a finite set of points $P \subset \mathbb{M}$

Triangulation of a submanifold

We call triangulation of a submanifold $\mathbb{M} \subset \mathbb{R}^d$ a simplicial complex $\hat{\mathbb{M}}$ such that

- $\hat{\mathbb{M}}$ is embedded in $\mathbb{R}^d$
- its vertices are on $\mathbb{M}$
- it is homeomorphic to $\mathbb{M}$

Submanifold reconstruction

The problem is to construct a triangulation $\hat{\mathbb{M}}$ of some unknown submanifold $\mathbb{M}$ given a finite set of points $P \subset \mathbb{M}$

Issues in high-dimensional geometry

- Dimensionality severely restricts our intuition and ability to visualize data
  $\Rightarrow$ need for automated and provably correct methods methods

- Complexity of data structures and algorithms rapidly grow as the dimensionality increases

  $\Rightarrow$ no subdivision of the ambient space is affordable

  $\Rightarrow$ data structures and algorithms should be sensitive to the intrinsic dimension (usually unknown) of the data

- Inherent defects : sparsity, noise, outliers

Issues in high-dimensional geometry

- Dimensionality severely restricts our intuition and ability to visualize data
  $\Rightarrow$ need for automated and provably correct methods methods

- Complexity of data structures and algorithms rapidly grow as the dimensionality increases

  $\Rightarrow$ no subdivision of the ambient space is affordable

  $\Rightarrow$ data structures and algorithms should be sensitive to the intrinsic dimension (usually unknown) of the data

- Inherent defects : sparsity, noise, outliers

Issues in high-dimensional geometry

- Dimensionality severely restricts our intuition and ability to visualize data
  $\Rightarrow$ need for automated and provably correct methods methods

- Complexity of data structures and algorithms rapidly grow as the dimensionality increases

  $\Rightarrow$ no subdivision of the ambient space is affordable

  $\Rightarrow$ data structures and algorithms should be sensitive to the intrinsic dimension (usually unknown) of the data

- Inherent defects : sparsity, noise, outliers

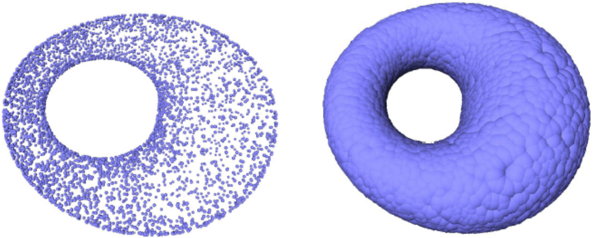
Looking for small and faithful simplicial complexes

Need to compromise

- Size of the complex
  - can we have $\dim \hat{\mathbb{M}} = \dim \mathbb{M}$?
- Efficiency of the construction algorithms and of the representations
  - can we avoid the exponential dependence on d ?
  - can we minimize the number of simplices ?
- Quality of the approximation
  - Homotopy type & homology (Cech and α complexes, persistence)
  - Homeomorphism (Delaunay-type complexes)

Sampling and distance functions [Niyogi et al.], [Chazal et al.]

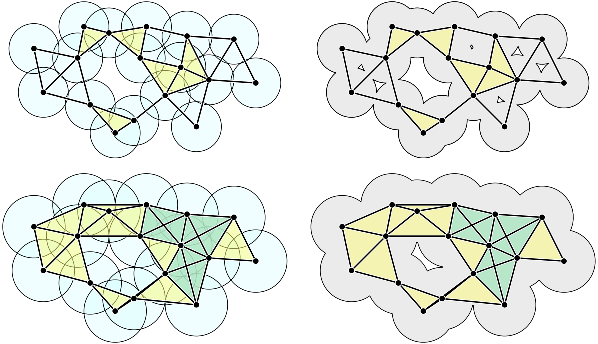
Distance to a compact K : $d_K : x \to \inf_{p \in K} \|x - p\|$



Stability

If the data points C are close (Hausdorff) to the geometric structure K, the topology and the geometry of the offsets $K_r = d^{-1}([0,r])$ and $C_r = d^{-1}([0,r])$ are close

Distance functions and triangulations



Nerve theorem (Leray)

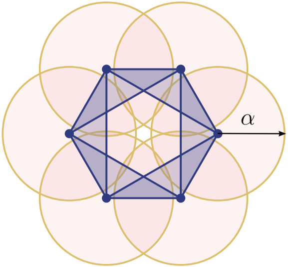
The nerve of the balls (Cech complex) and the union of balls have the same homotopy type (same result for the α-complex)

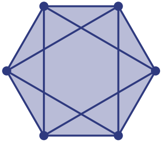
Questions

+ The homotopy type of a compact set X can be computed from the Čech complex of a sample of X
+ The same is true for the α-complex
- The Čech and the α-complexes are huge ($O(n^d)$ and $O(n^{\lceil d/2 \rceil})$) and difficult to compute in high dimensions
- Both complexes are not in general homeomorphic to X (i.e. not a triangulation of X)
- The Čech complex cannot be realized in general in the same space as X

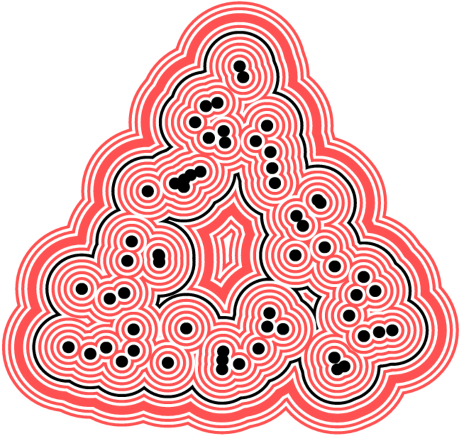
Čech and Rips complexes

The Rips complex is easier to compute but still very big, and less precise in approximating the topology





An example where no offset has the right topology !

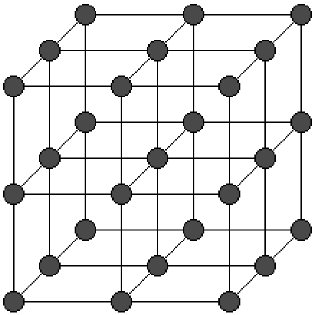


1. Manifold + small noise assumption
2. Call persistent homology at rescue !

The curses of Delaunay triangulations in higher dimensions

- Complexity depends exponentially on the ambient dimension. Robustness issues become very tricky
- Higher dimensional Delaunay triangulations are not thick even if the vertices are well-spaced
- The restricted Delaunay triangulation is no longer a good approximation of the manifold even under strong sampling conditions (for $d > 2$)

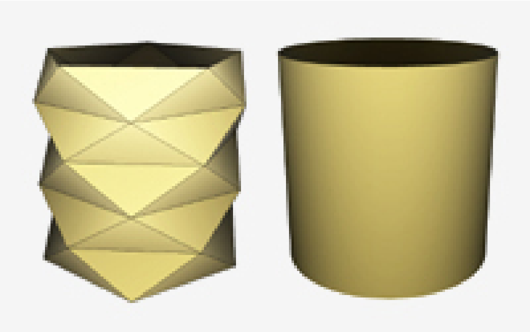
3D Delaunay Triangulations are not thick even if the vertices are well-spaced



- Each square face can be circumscribed by an empty sphere
- This remains true if the grid points are slightly perturbed therefore creating thin simplices

Badly-shaped simplices

Badly-shaped simplices lead to bad geometric approximations



which in turn may lead to topological defects in $\mathrm{Del}_{|\mathcal{M}}(\mathsf{P})$ [Oudot]

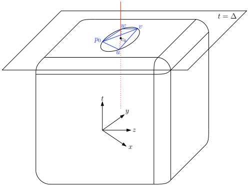
see also [Cairns], [Whitehead], [Munkres], [Whitney]

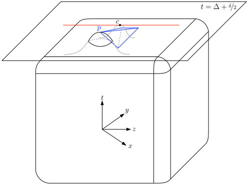
Tangent space approximation

$\mathbb{M}$ is a smooth k-dimensional manifold ($k > 2$) embedded in $\mathbb{R}^d$

Bad news [Oudot 2005]

The Delaunay triangulation restricted to $\mathbb{M}$ may be a bad approximation of the manifold even if the sample is dense





Thickness and tangent space approximation

Lemma [Whitney 1957]

If σ is a j-simplex whose vertices all lie within a distance η from a hyperplane $H \subset \mathbb{R}^d$, then

$$\sin \angle(\text{aff}\,(\sigma), H) \leq \frac{2j\,\eta}{D(\sigma)}$$

Corollary

If σ is a j-simplex, $j \leq k$, vert $(\sigma) \subset \mathbb{M}$, $\Delta(\sigma) \leq \delta\,\text{rch}(\mathbb{M})$

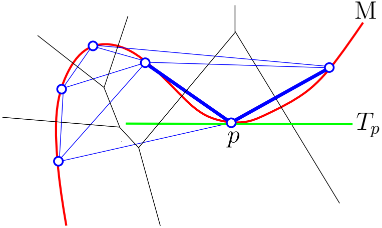
$$\forall p \in \sigma, \quad \sin \angle(\text{aff}(\sigma), T_p) \leq \frac{\delta}{\Theta(\sigma)}$$

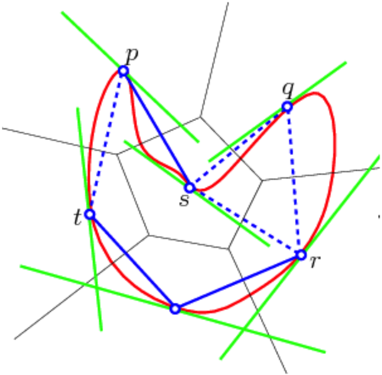
($\eta \leq \frac{\Delta(\sigma)^2}{2\,\text{rch}(\mathbb{M})}$ by the Chord Lemma)

The assumptions

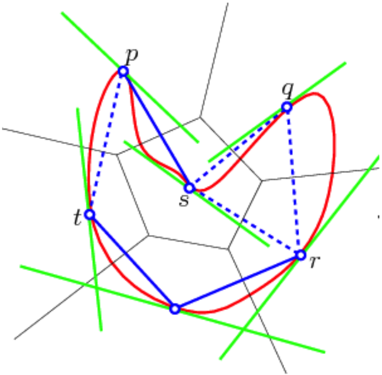
- $\mathbb{M}$ is a differentiable submanifold of positive reach of $\mathbb{R}^d$
- The dimension k of $\mathbb{M}$ is small
- P is an ε-net of $\mathbb{M}$, i.e.
  - $\forall x \in \mathbb{M}, \ \exists \ p \in \mathsf{P}, \ \ \|x-p\| \leq \varepsilon \, \mathrm{rch}(\mathbb{M})$
  - $\forall p, q \in \mathsf{P}, \ \ \|p-q\| \geq \bar{\eta} \, \varepsilon$
- ε is small enough

The tangential Delaunay complex [B. & Ghosh 2010]





1. Construct the star of $p \in \mathsf{P}$ in the Delaunay triangulation $\mathrm{Del}_{T_p}(\mathsf{P})$ of P restricted to T_p
2. $\mathrm{Del}_{T\mathbb{M}}(\mathsf{P}) = \bigcup_{p \in \mathsf{P}} \mathrm{star}(p)$



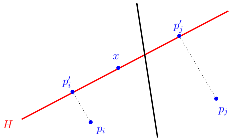
+ $\mathrm{Del}_{T\mathbb{M}}(\mathsf{P}) \subset \mathrm{Del}(\mathsf{P})$

+ $\mathrm{star}(p)$, $\mathrm{Del}_{T_p}(\mathsf{P})$ and therefore $\mathrm{Del}_{T\mathbb{M}}(\mathsf{P})$ can be computed without computing $\mathrm{Del}(\mathsf{P})$

– $\mathrm{Del}_{T\mathbb{M}}(\mathsf{P})$ is not necessarily a triangulated manifold

Construction of $\mathrm{Del}_{T_p}(\mathsf{P})$

Given a d-flat $H \subset \mathbb{R}$, $\mathrm{Vor}(\mathsf{P}) \cap H$ is a weighted Voronoi diagram in H



$$\|x - p_i\|^2 \leq \|x - p_j\|^2$$
$$\Leftrightarrow \quad \|x - p'_i\|^2 + \|p_i - p'_i\|^2 \leq \|x - p'_j\|^2 + \|p_j - p'_j\|^2$$

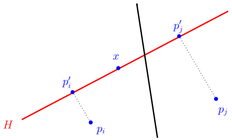
Corollary: construction of Del_{T_p}

$$\Psi_p(p_i) = (p'_i, -\|p_i - p'_i\|^2) \qquad \text{(weighted point)}$$

1. project P onto T_p which requires $O(Dn)$ time
2. construct $\mathrm{star}(\Psi_p(p_i))$ in $\mathrm{Del}(\Psi_p(p_i)) \subset T_{p_i}$
3. $\mathrm{star}(p_i) \approx \mathrm{star}(\Psi_p(p_i))$ (isomorphic)

Construction of $\mathrm{Del}_{T_p}(\mathrm{P})$

Given a d-flat $H \subset \mathbb{R}$, $\mathrm{Vor}(\mathrm{P}) \cap H$ is a weighted Voronoi diagram in H



$$\|x - p_i\|^2 \leq \|x - p_j\|^2$$
$$\Leftrightarrow \quad \|x - p'_i\|^2 + \|p_i - p'_i\|^2 \leq \|x - p'_j\|^2 + \|p_j - p'_j\|^2$$

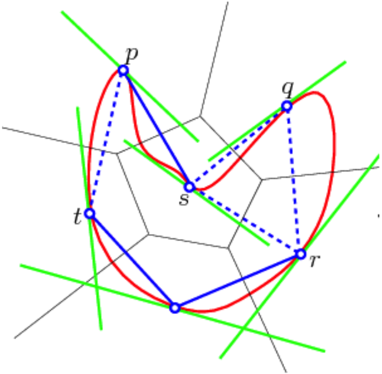
Corollary: construction of Del_{T_p}

$$\Psi_p(p_i) = (p'_i, -\|p_i - p'_i\|^2) \qquad \text{(weighted point)}$$

1. project P onto T_p which requires $O(Dn)$ time
2. construct $\mathrm{star}(\Psi_p(p_i))$ in $\mathrm{Del}(\Psi_p(p_i)) \subset T_{p_i}$
3. $\mathrm{star}(p_i) \approx \mathrm{star}(\Psi_p(p_i))$ (isomorphic)

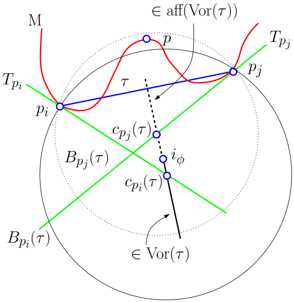
Inconsistencies in the tangential complex

A simplex is not in the star of all its vertices



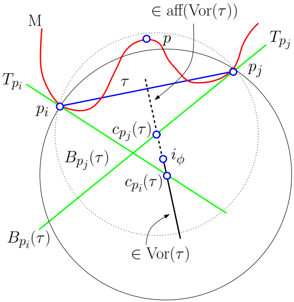
- $\tau \in \mathrm{star}(p_i) \quad \Leftrightarrow \quad T_{p_i} \cap \mathrm{Vor}(\tau) \neq \emptyset \quad \Leftrightarrow \quad B(c_{p_i}(\tau) \cap \mathsf{P} = \emptyset$
- $\tau \notin \mathrm{star}(p_j) \quad \Leftrightarrow \quad T_{p_j} \cap \mathrm{Vor}(\tau) = \emptyset \quad \Leftrightarrow \quad B(c_{p_j}(\tau) \cap \mathsf{P} \ni p$

Inconsistency $(k+1)$-trigger



- $B_{p_i}(\tau)$: ball circumscribing τ centered on T_{p_i}, c_{p_i} its center
- Inconsistency : $B_{p_i}(\tau) \cap P = \emptyset$ and $B_{p_j}(\tau) \cap P \neq \emptyset$
- $p_l \in B_{ij}$, first point hit by $(1-\lambda)B_{p_i} + \lambda B_{p_j}$, $\lambda : 0 \to 1$
- Trigger $\tau^\dagger$: $(k+1)$-simplex $\tau \star p_l \in \text{Del}_((P))$

Inconsistency $(k+1)$-triggers are flakes



If τ is small and thick, then

- $T_{p_i} \approx T_{p_j} \approx \mathrm{aff}(\tau)$ (sample density)
- $\|c_{p_i} - c_{p_j}\|$ small $\Rightarrow$ $B_{ij} := B_{p_i}(\tau) \setminus B_{p_j}(\tau) \neq \emptyset$ is small (τ thick)
- the trigger $\tau^\dagger = \tau \star p_l$ is not thick

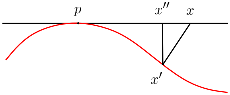
Bound on the diameter of the simplices of $\mathrm{Del}_{T\mathbb{M}}(P)$

(i) $\mathrm{Vor}(p) \cap T_p \subseteq B(p, \alpha_0 \,\mathrm{rch}(\mathbb{M}))$ where $\alpha_0 \approx \varepsilon$

(ii) $\forall \sigma \in \mathrm{star}(p)$, $R_p(\sigma) \leq \alpha \,\mathrm{rch}(\mathbb{M})$

(iii) $\forall \sigma \in \mathrm{Del}_{T\mathbb{M}}(\mathsf{P})$, $\Delta(\sigma) \leq 2\alpha \,\mathrm{rch}(\mathbb{M})$.

Proof of (i)



$x \in \mathrm{Vor}(p) \cap T_p$, $\|p - x\| = \alpha\, \mathrm{rch}(\mathbb{M})$

x' be the point of $\mathbb{M}$ closest to x and $x'' = \Pi_p(x')$.

$$\|p - x'\| \leq \|p - x\| + \|x - x'\| \leq 2\|p - x\|$$
$$\Rightarrow \quad \|x' - x''\| \leq \frac{\|p-x'\|^2}{2\mathrm{rch}(\mathbb{M})} \leq 2\alpha^2\, \mathrm{rch}(\mathbb{M})$$ (Chord Lemma)

$\|x' - x''\| = \|x - x'\| \cos\phi$, where $\phi = \angle(T_{x'}, T_p)$ and $\cos\phi \geq 1 - 8\alpha^2$

$$\Rightarrow \quad \|x - x'\| \leq \frac{2\alpha^2\, \mathrm{rch}(\mathbb{M})}{1 - 8\alpha^2} \ \text{ assuming } \alpha \leq \frac{\sqrt{2}}{4}$$

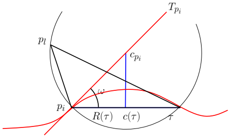
P is an ε-dense sample : $\exists q \in$ P, $\|x' - q\| \leq \varepsilon\, \mathrm{rch}(\mathbb{M})$

$$\|x-p\| = \alpha\, \mathrm{rch}(\mathbb{M}) \leq \|x-q\| \leq \|x-x'\|+\|x'-q\| \leq \left(\frac{2\alpha^2}{1-8\alpha^2} + \varepsilon\right) \mathrm{rch}(\mathbb{M})$$

Bound on the diameter of inconsistency triggers

τ an inconsistent k-simplex, $\tau^\dagger$ a trigger, $\theta = \max_{p\in\tau} \angle(\mathrm{aff}(\tau), T_p)$

Lemma $\quad \sin\theta \leq \frac{\Delta(\tau)}{\Theta(\tau))\,\mathrm{rch}(\mathbb{M})}$ and $R(\tau^\dagger) \leq \frac{R(\tau)}{\cos\theta}$



Proof

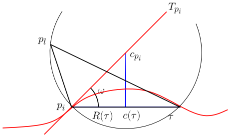
$$d(p_l, T_p) \leq \frac{\Delta^2(\tau)}{2\mathrm{rch}(\mathbb{M})}$$ (Chord Lemma)

$$\sin\angle(\mathrm{aff}(\tau), T_p) \leq \frac{2\,\frac{\Delta^2(\tau)}{2\mathrm{rch}(\mathbb{M})}}{\Theta(\tau)\,\Delta(\tau)} = \frac{\Delta(\tau)}{\Theta(\tau)\,\mathrm{rch}(M)}$$ (Whitney's angle bound)

$$R_{p_i}(\tau) = \|p_i - c_{p_i}\| \leq \frac{R(\tau)}{\cos\theta} \quad \text{and} \quad R(\tau^\dagger) \leq \|i(\tau^\dagger) - p_i\| \leq \frac{R(\tau)}{\cos\theta}$$

Bound on the thickness of inconsistency triggers

Lemma $\Theta(\tau^\dagger) \leq \frac{\Delta(\tau^\dagger)}{2(k+1)\mathrm{rch}(\mathbb{M})}\left(1+\frac{2}{\Theta(\tau)}\right)$



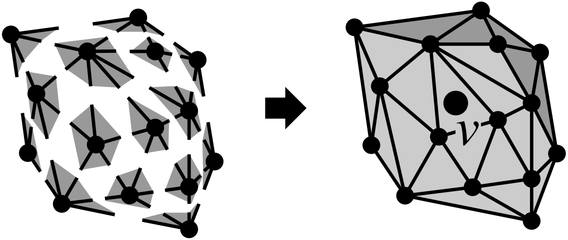
Proof Let $q \in \tau$.

$$\begin{aligned} D(p_l,\tau^\dagger) &= \|p_l - q\| \ \sin\angle(p_l - q, \mathrm{aff}(\tau)) \\ &\leq \Delta(\tau^\dagger)\ (\sin\angle(p_l - q, T_q) + \sin\angle(T_q, \mathrm{aff}(\tau))) \\ &\leq \Delta(\tau^\dagger)\left(\frac{\Delta(\tau^\dagger)}{2\mathrm{rch}(\mathbb{M})} + \frac{\Delta(\tau)}{\Theta(\tau)\,\mathrm{rch}(\mathbb{M})}\right) \quad \text{(Chord + previous Lemmas)} \\ &\leq \frac{\Delta^2(\tau^\dagger)}{2\mathrm{rch}(\mathbb{M})}\left(1+\frac{2}{\Theta(\tau)}\right) \end{aligned}$$

Hence, if τ is thick, $\tau^\dagger$ cannot be so : we say that $\tau^\dagger$ is a flake

Reconstruction of smooth submanifolds

1. For each vertex v, compute the star $\mathrm{star}(p)$ of p in $\mathrm{Del}_p(\mathsf{P})$
2. Remove inconsistencies among the stars by weighting the points
3. Stitch the stars to obtain a triangulation of P



Algorithm hypotheses

Known quantities in red

- $\mathbb{M}$ = a differentiable submanifold of positive reach of dim. $k \subset \mathbb{R}^d$
- P = an (ε, δ)-sample of $\mathbb{M}$
- $\varepsilon \leq \varepsilon_0$
- $\varepsilon/\delta \leq \eta_0$
- we can estimate the tangent space T_p at any $p \in$ P

Removing inconsistencies by removing flakes

Another application of Moser-Tardos algorithmic LLL

Input: P, $\{T_p, p \in P\}$, $\tilde{w}_0$, Θ_0

Initialize all weights to 0 and compute $\mathrm{Del}_{T\mathbb{M}}(\hat{\mathsf{P}})$

while there are Θ_0-flakes or inconsistencies in $\mathrm{Del}_{T\mathbb{M}}(\hat{\mathsf{P}})$ **do**

  while there is a Θ_0-flake σ in $\mathrm{Del}_{T\mathbb{M}}(\hat{\mathsf{P}})$ **do**
    resample σ, i.e. reweight the vertices of σ
    update $\mathrm{Del}_{T\mathbb{M}}(\hat{\mathsf{P}})$

  if there is an inconsistent simplex σ in $\mathrm{Del}_{T\mathbb{M}}(\hat{\mathsf{P}})$ **then**
    compute a trigger simplex $\sigma^\dagger$ associated to σ
    resample the flake $\sigma \subset \sigma^\dagger$
    update $\mathrm{Del}_{T\mathbb{M}}(\hat{\mathsf{P}})$

Output: A weighting scheme on P and $\mathrm{Del}_{T\mathbb{M}}(\hat{\mathsf{P}})$
$\mathrm{Del}_{T\mathbb{M}}(\hat{\mathsf{P}})$ is Θ_0-thick and has no inconsistency

Summary

- Termination
  - If $\frac{\bar{\eta}}{2} \geq \bar{\rho} \geq f(\Theta_0)$, the algorithm terminates and returns a complex $\hat{\mathbb{M}}$ that has no inconsistent configurations
- Complexity
  - No d-dimensional data structure $\Rightarrow$ linear in d
  - exponential in k
- Approximation
  - $\hat{\mathbb{M}}$ is a PL simplicial k-manifold
  - $\hat{\mathbb{M}} \subset \text{tub}(\mathbb{M}, \varepsilon)$
  - $\hat{\mathbb{M}}$ is homeomorphic to $\mathbb{M}$

$\hat{\mathbb{M}}$ is a PL simplicial k-manifold

Lemma Let P be an ε-sample of a manifold $\mathbb{M}$ and let $p \in$ P. The link of any vertex p in $\hat{\mathbb{M}}$ is a topological $(k-1)$-sphere

Proof :

1. Since $\hat{\mathbb{M}}$ contains no inconsistencies, $\forall p \in \mathrm{vert}(\hat{\mathbb{M}}), \quad \mathrm{star}(p, \hat{\mathbb{M}}) = \mathrm{star}(p, \mathrm{Del}_p(\mathsf{P}))$
2. $\mathrm{Del}_p(\mathsf{P}) \subset \mathbb{R}^d \;\approx\; \mathrm{Del}(\Psi_p(\mathsf{P})) \subset T_p \;\Rightarrow\; \mathrm{star}(p) \approx \mathrm{star}_p(p)$
3. $\mathrm{star}_p(p)$ is a k-dimensional triangulated topological ball (general position)
4. p cannot belong to the boundary of $\mathrm{star}_p(p)$

(the Voronoi cell of $p = \Psi_p(p)$ in $\mathrm{Vor}(\Psi_p(\mathsf{P}))$ is bounded)

$\hat{\mathbb{M}}$ is a triangulation of $\mathbb{M}$

Theorem

$\mathbb{M}$: a connected compact k-submanifold of $\mathbb{R}^d$ without boundary
$\hat{\mathbb{M}}$: combinatorial k-manifold without boundary, embedded in $\mathbb{R}^d$ s.t.

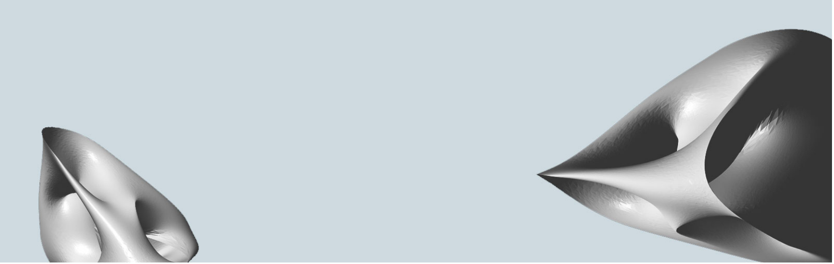
1. $\mathrm{P} = \mathrm{vert}(\hat{\mathbb{M}}) \subset \mathbb{M}$
2. $\forall \sigma \in \hat{\mathbb{M}}$, $\quad \Delta(\sigma) \leq \delta_0 \,\mathrm{rch}(\mathbb{M})$ where $\delta_0 \leq \frac{1}{5}$ (longest edge)
   $L(\sigma) \geq \lambda_0 \,\mathrm{rch}(\mathbb{M})$ (shortest edge)
   $\Theta(\sigma) \geq 8.13\ \delta_0^2/\lambda_0$ (thickness)
3. $\exists p_i \in \hat{\mathbb{M}}$ s.t. $\Pi^{-1}(p_i) = \{p_i\}$

Then

1. $\hat{\mathbb{M}}$ is a triangulation of $\mathbb{M}$
2. The Hausdorff distance between $\hat{\mathbb{M}}$ and $\mathbb{M}$ is at most $2\delta_0^2 \,\mathrm{rch}(\mathbb{M})$
3. If σ is a k-simplex of $\hat{\mathbb{M}}$ and p one of its vertices, we have

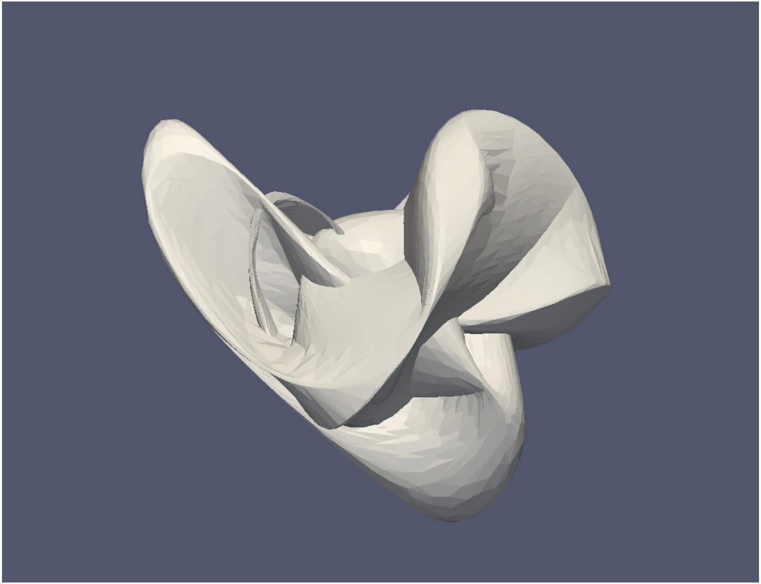
$$\sin \angle(\mathrm{aff}(\sigma), T_p) \leq \frac{\delta_0}{\Theta_0}$$

Triangulating a Riemann surface



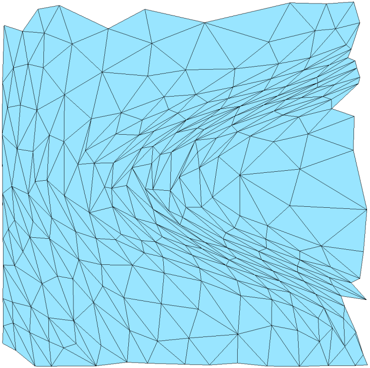
a complex curve in the projective plane (parameterized as a real surface of $\mathbb{R}^8$)

Triangulating the conformational space of C_8H_{16}



Applications and extensions

- Discrete metric sets (see the previous lecture on the witness complex)
- Anisotropic mesh generation



- Non euclidean embedding space (e.g. statistical manifolds)