Geometric Methods in Data Analysis

Manifold Learning

Jean-Daniel Boissonnat
INRIA & 3IA

Fall 2021

Reference

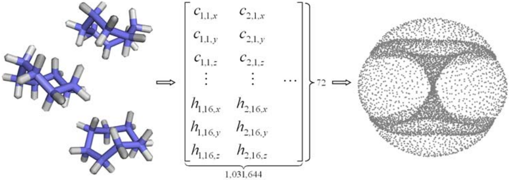
J-D. Boissonnat, F. Chazal, M. Yvinec

Geometric and Topological Inference

Cambridge University Press. 2018

1. The shape of data
2. The manifold model
3. Quality criteria
4. Discrete models
5. Reconstruction of submanifolds
   - Distance functions and homotopic reconstruction
   - Delaunay triangulation of submanifolds of $\mathbb{R}^d$

The shape of data



Geometrisation : Data = points + distances between points

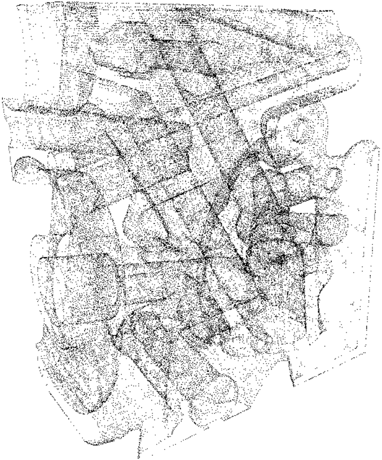
Hypothesis : Data lie close to a structure of “small” intrinsic dimension

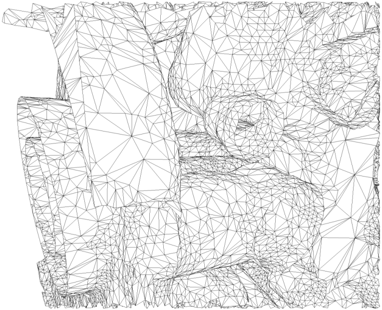
Problem : Infer the structure from the data

Geometric models and approaches

- Dimensionality reduction either linear (PCA, MDS) or non-linear (ISOMAP, LLE)
- Computation of topological and/or geometric invariants
  - Clustering
  - Persistent homology
- Shape reconstruction

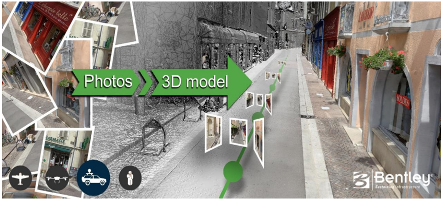
Reconstructing surfaces from point clouds

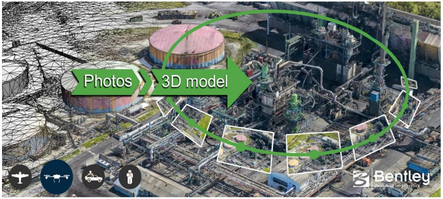




One can reconstruct a surface from 10^6 points within 1mn [CGAL]

Reconstructing surfaces from images

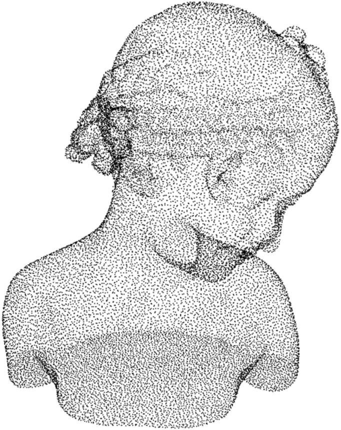


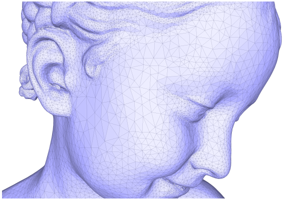


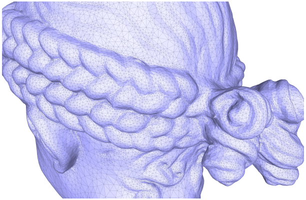
Acute3D, Bentley Systems

Geometric shape reconstruction

Towards a sampling theory for geometric objects

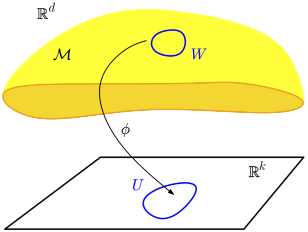


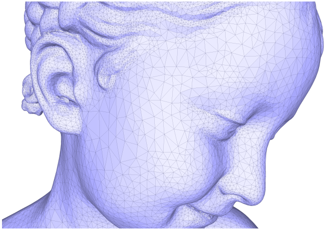




- What spaces ?
- Quality criteria
- Discrete models
- Curse of dimensionality

The manifold model

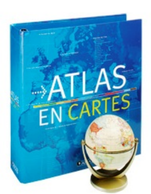


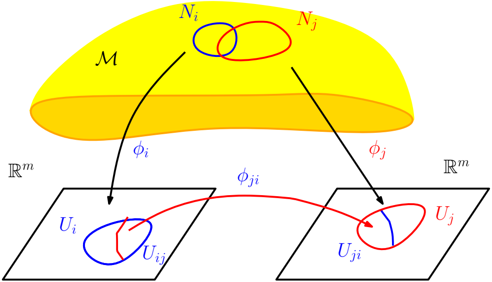


A manifold of dimension m is a topological space that locally looks like (is homeomorphic to) $\mathbb{R}^m$

Examples : curves and surfaces, isomanifolds (zero set of $f : \mathbb{R}^d \rightarrow \mathbb{R}^{d-m}$), configuration spaces of mechanisms, time series

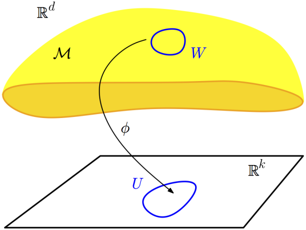
Charts and atlases

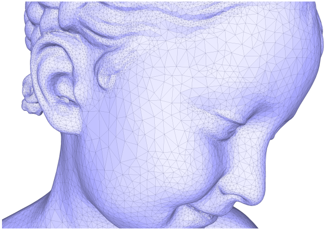




ϕ_{ij} : transition functions between charts

A submanifold of dimension m is a subset of $\mathbb{R}^d$ that locally looks like (is homeomorphic to) $\mathbb{R}^m$

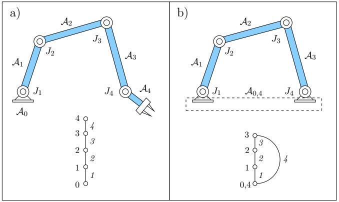


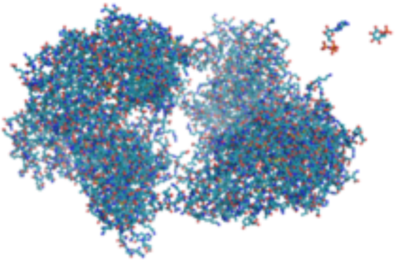


The manifold hypothesis : $m \ll d$

Configuration spaces

Associate a point (a vector of parameters) to each configuration





Definition : $f : X \to Y$ embeds X in Y iff X and $f(X) \subset Y$ are topologically equivalent

Example : The usual sphere is a manifold of intrinsic dimension 2 that cannot be embedded in $\mathbb{R}^2$

Whitney's embedding theorem

Any manifold $\mathbb{M}$ of intrinsic dimension m can be embedded as a submanifold of $\mathbb{R}^{2m+1}$

Dimensionality reduction

Variant of the Johnson Lindenstrauss lemma for manifolds

[Baraniuk & Wakin 2007] [Clarkson 2007]

Let $\mathbb{M}$ be a m-submanifold of $\mathbb{R}^d$ of bounded geometry and $\epsilon \in (0, 1)$. If one projects and scale $\mathbb{M}$ on a random flat H of dimension

$$k = \Omega(\frac{m}{\varepsilon^2} \log \frac{1}{\varepsilon} + \frac{1}{\varepsilon^2} \log \frac{1}{\delta}),$$

then, with probability $1 - \delta$,

$$\forall p, q \in \mathbb{M}, \quad (1-\epsilon) \|p-q\|^2 \leq \|f(p) - f(q)\|^2 \leq (1+\epsilon) \|p-q\|^2$$

where $f(p) = \sqrt{\frac{d}{k}}\, \pi_H(p)$

Time series

Delay Embedding Theorem (Takens 1981)

Dynamical system : $\phi : \mathbb{R} \times X \rightarrow X$ $\quad X \in \mathbb{R}^d$ is the state space

We look for stable manifolds $M \subset X$: $\phi(\mathbb{R}, M) = M$

Typically $\quad m = \dim(M) \ll d = \dim(X)$.

In practice, we don't have access to M but we observe the system at regular time intervals using an observation function $\alpha : M \rightarrow \mathbb{R}$:

$$\psi(x) = (\alpha(x), \alpha(\phi(x)), ..., \alpha(\phi^{k-1}(x)) \in \mathbb{R}^k$$

Takens's theorem

if $k \geq 2m + 1$, then ψ is (generically) an embedding of M in $\mathbb{R}^k$

Dynamical system : $\phi : \mathbb{R} \times X \rightarrow X$ $\qquad X \in \mathbb{R}^d$ is the state space

We look for stable manifolds $M \subset X$: $\phi(\mathbb{R}, M) = M$

Typically $\quad m = \dim(M) \ll d = \dim(X)$.

In practice, we don't have access to M but we observe the system at regular time intervals using an observation function $\alpha : M \rightarrow \mathbb{R}$:

$$\psi(x) = (\alpha(x), \alpha(\phi(x)), ..., \alpha(\phi^{k-1}(x)) \in \mathbb{R}^k$$

Takens's theorem

if $k \geq 2m + 1$, then ψ is (generically) an embedding of M in $\mathbb{R}^k$

Dynamical system : $\phi : \mathbb{R} \times X \rightarrow X$ $X \in \mathbb{R}^d$ is the state space

We look for stable manifolds $M \subset X$: $\phi(\mathbb{R}, M) = M$

Typically $m = \dim(M) \ll d = \dim(X)$.

In practice, we don't have access to M but we observe the system at regular time intervals using an observation function $\alpha : M \rightarrow \mathbb{R}$:

$$\psi(x) = (\alpha(x), \alpha(\phi(x)), ..., \alpha(\phi^{k-1}(x)) \in \mathbb{R}^k$$

Takens's theorem

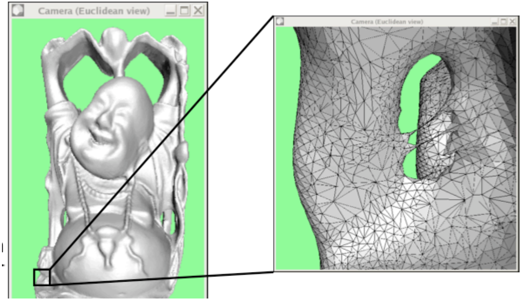
if $k \geq 2m + 1$, then ψ is (generically) an embedding of M in $\mathbb{R}^k$

Topological quality criteria for approximations

Homeomorphism (topological equivalence)

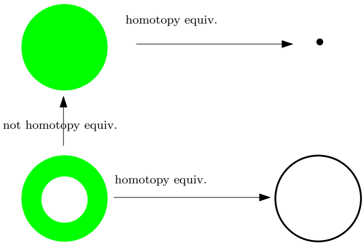
$f : X \to Y$ is a continuous bijective mapping whose inverse is continuous

$$X \approx Y$$



Homotopy equivalence

A weaker notion of topological equivalence



Intuitively, two spaces X and Y are homotopy equivalent if they can be transformed into one another

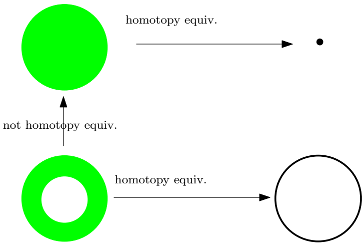
by bending, shrinking and expanding operations

but not by cutting or tearing

X is contractible if it is homotopy equivalent to a point

Homotopy equivalence

A weaker notion of topological equivalence



Intuitively, two spaces X and Y are homotopy equivalent if they can be transformed into one another

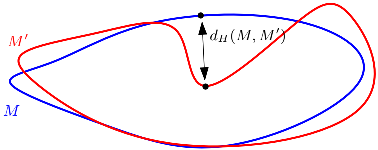
by bending, shrinking and expanding operations

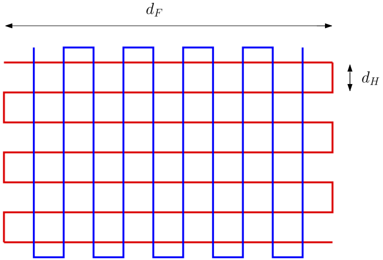
but not by cutting or tearing

X is contractible if it is homotopy equivalent to a point

Geometric quality criteria

1. Hausdorff distance / Fréchet distance





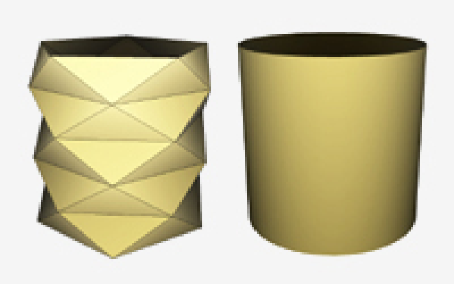
$$d_H(M, M') = \max\left(\sup_{x\in M} \inf_{x'\in M'} \|x - x'\|, \sup_{x\in M} \inf_{x'\in M'} \|x - x'\| \right)$$

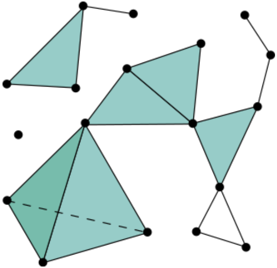
$$= \min\{r : M \subset M'^{+r} \text{ et } M' \subset M^{+r}\}$$

$$d_F(M, M') = \inf_{f\in\mathcal{H}(M,M')} \sup_{x\in M} \|x - f(x)\|$$

Geometric quality criteria

2. Tangent spaces approximation





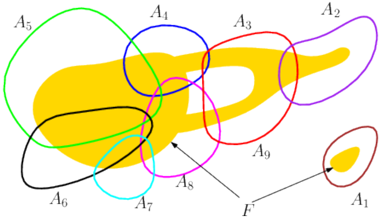
H. Poincaré (1854-1912)

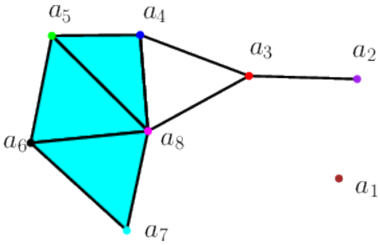
Let V be a finite set. A simplicial complex (abstract) on V is a finite set of subsets of V called the simplexes or faces of K that satisfy :

1. The elements of V belong to K (vertices)
2. If $\tau \in K$ and $\sigma \subseteq \tau$, then $\sigma \in K$

Nerve of a good cover

A simplicial complex to represent the topology of an object





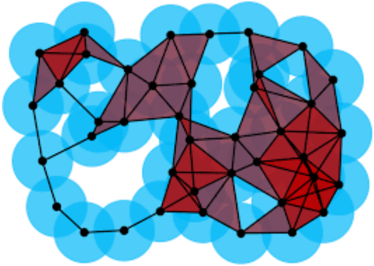
Nerve theorem (J. Leray, 1945)

If the intersection of any subset of elements in the cover is contractible, then the nerve and the union of the elements of the cover have the same homotopy type.

Čech complex

Nerve of a set of balls

A finite set of points $\mathcal{P} \in \mathbb{R}^d$



J. Leray

(1906-1998)

Corollary of the nerve theorem (J. Leray, 1945)

The Čech complex has the same homotopy type as the union of balls

1. The shape of data

2. The manifold model

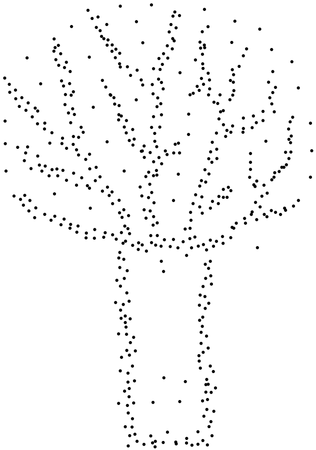
3. Quality criteria

4. Discrete models

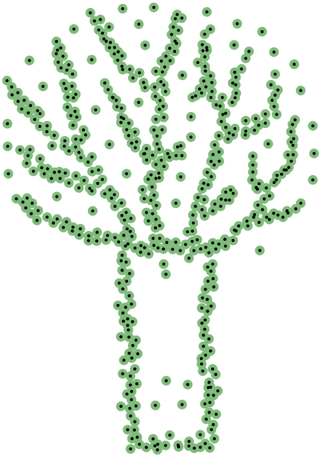
5. Reconstruction of submanifolds
   - Distance functions and homotopic reconstruction
   - Delaunay triangulation of submanifolds of $\mathbb{R}^d$

Reconstruction of geometric shapes

Union of balls and distance functions



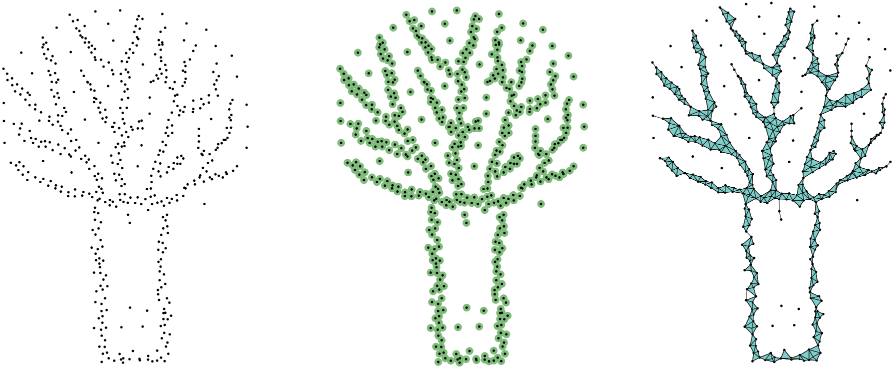
Sample P



Union of balls $P^{+\alpha}$

Shape reconstruction

Discrete approximation of continuous spaces

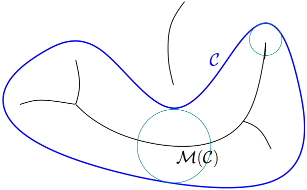


$\boxed{\text{Object}} \rightarrow \boxed{\text{Sample}} \quad \rightarrow \quad \boxed{\text{Union of balls}} \quad \xrightarrow{\text{nerve}} \quad \boxed{\text{Simplicial complex}}$

Reconstruction theorems

Niyogi, Smale, Weinberger [2008]

If $\mathbb{M}$ is a submanifold of positive reach τ ,
P an ε-dense sample of $\mathbb{M}$,
then, for all $\alpha \in [\sim \varepsilon, \sim \tau]$, $P^{+\alpha} \simeq \mathbb{M}$

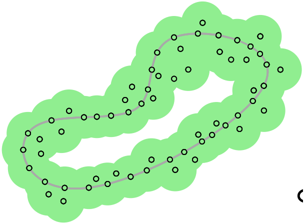


Chazal, Cohen-Steiner, Lieutier [2009]

Extension to general compact sets

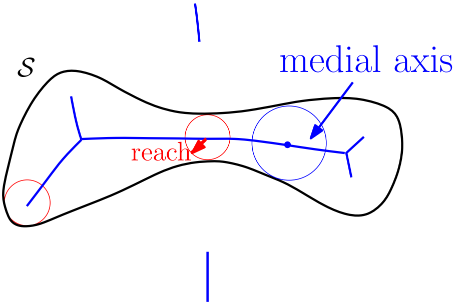
Chazal, Cohen-Steiner, Mérigot [2011]

Extension to points sets with outliers



Regularity condition : bounded reach

[Federer 1958], [Amenta & Bern 1998]

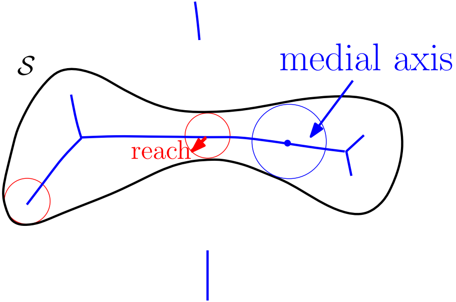


Local feature size

$\forall x \in \mathcal{S},\ \ \mathrm{lfs}(x) = d(x, \mathrm{axis}(\mathcal{S}))$ (1-Lipschitz : $|f(x) - f(y)| \leq \|x - y\|$)

$\mathrm{rch}(\mathcal{S}) = \inf_{x \in \mathcal{S}} \mathrm{lfs}(x)$

Sampling condition : ε-nets



$(\epsilon, \bar{\eta})$-net of $\mathcal{S}$

1. Covering : $\mathcal{P} \subset \mathcal{S}, \forall x \in \mathcal{S}, \quad d(x, \mathcal{P}) \ \leq \ \epsilon \ \mathrm{lfs}(x)$
2. Packing : $\forall p, q \in \mathcal{P}, \quad \|p - q\| \ \geq \ \bar{\eta}\varepsilon \ \min(\mathrm{lfs}(p), \mathrm{lfs}(q))$

Two questions

Complexity

The Čech complex is big ($\Theta(n^d)$) and difficult to compute
(reduces to computing smallest enclosing balls)

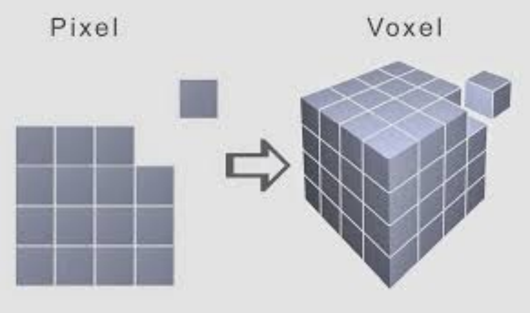
- Can/when we avoid the exponential dependence on d?

Quality of approximation

The Čech complex is not (in general) homeomorphic to X and cannot be embedded in the same space as X

The curse of dimensionallity

Subdividing the ambient space is impossible in high dimensions



Resolution $= 1/N$

$\Downarrow$

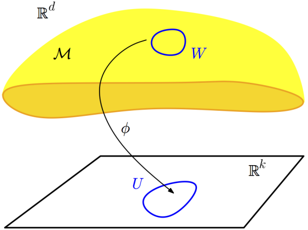
Number of cells $= N^d$

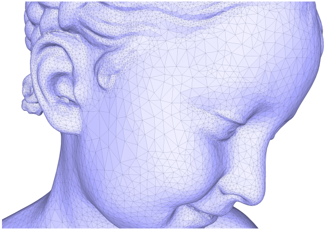
$N = 1000$

$N^2 = 1$ million $\quad N^3 = 1$ billion $\quad N^6 = 1.000.000.000.000.000.000$

Shape model

Smooth manifolds of low intrinsic dimensions



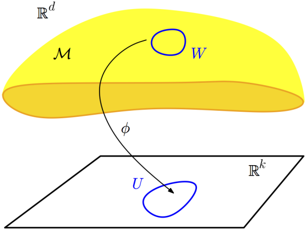


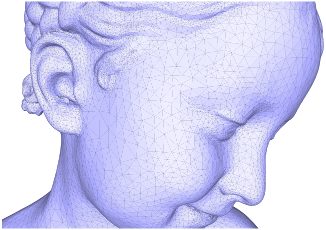
Manifold Hyp. : $\boxed{m \ll d}$

- Can we bound the combinatorial complexity as a function of the intrinsic dimension ?
- Can we reconstruct a simplicial complex homeomorphic to the manifold, i.e. a triangulation of the manifold ?

Shape model

Smooth manifolds of low intrinsic dimensions





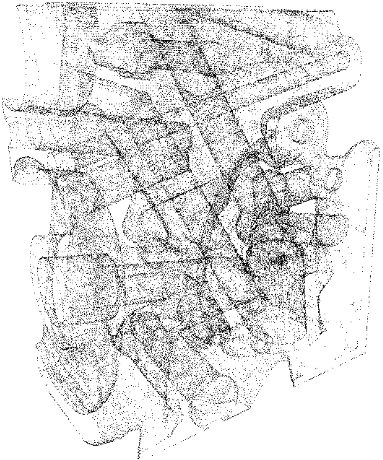
Manifold Hyp. : $\boxed{m \ll d}$

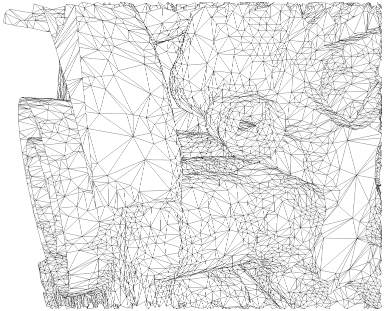
- Can we bound the combinatorial complexity as a function of the intrinsic dimension ?
- Can we reconstruct a simplicial complex homeomorphic to the manifold, i.e. a triangulation of the manifold ?

1. The shape of data
2. The manifold model
3. Quality criteria
4. Discrete models
5. Reconstruction of submanifolds
    - Distance functions and homotopic reconstruction
    - Delaunay triangulation of submanifolds of $\mathbb{R}^d$

Delaunay triangulation of manifolds

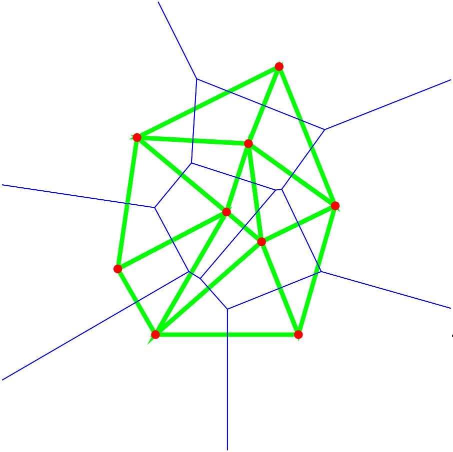
Can we extend the techniques developed for surfaces ?

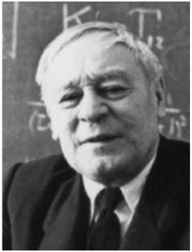




Delaunay Triangulations

Sur la sphère vide (On the empty sphere), Boris Delaunay (1934)





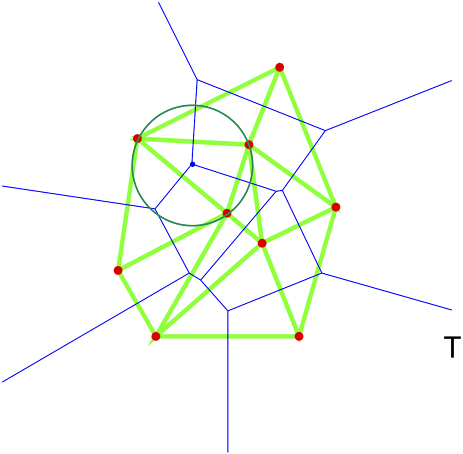
The Delaunay complex $\mathrm{Del}(\mathcal{P})$ is the nerve of $\mathrm{Vor}(\mathcal{P})$

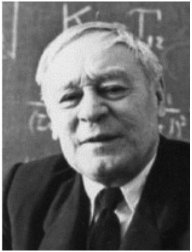
Theorem

If $\mathcal{P}$ contains no subset of $d+2$ points on a same hypersphere, then $\mathrm{Del}(\mathcal{P})$ can be realized as a triangulation of $\mathcal{P}$ in $\mathbb{R}^d$

Delaunay Triangulations

Sur la sphère vide (On the empty sphere), Boris Delaunay (1934)





The Delaunay complex $\mathrm{Del}(\mathcal{P})$ is the nerve of $\mathrm{Vor}(\mathcal{P})$

Theorem

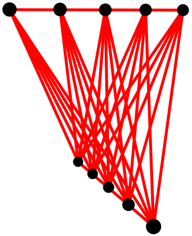
If $\mathcal{P}$ contains no subset of $d+2$ points on a same hypersphere, then $\mathrm{Del}(\mathcal{P})$ can be realized as a triangulation of $\mathcal{P}$ in $\mathbb{R}^d$

Combinatorial and algorithmic complexity

The complexity of the Delaunay triangulation of n points of $\mathbb{R}^d$ is exponential

$\Theta(n^{\lceil \frac{d}{2} \rceil})$ [Mc Mullen 1970]

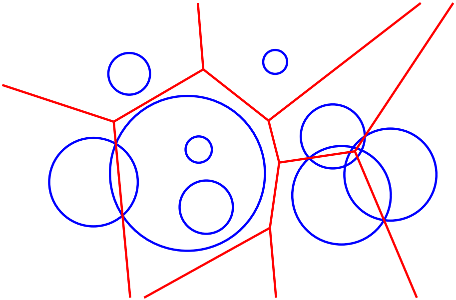
Worst-case : points on the moment curve $\Gamma(t) = \{t, t^2, ..., t^d\} \subset \mathbb{R}^d$



Construction : $\Theta(n \log n + n^{\lceil \frac{d}{2} \rceil})$ [Clarkson & Shor 1989] [Chazelle 1993]

Laguerre (power, weighted) diagrams

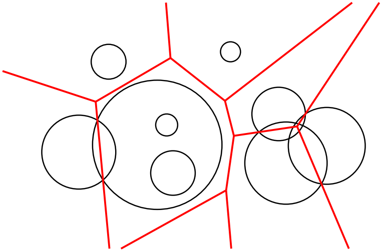
Weighted points : $\mathcal{B} = \{b_1, ..., b_n\}$ $D(x, b_i) = (x - p_i)^2 - r_i^2$



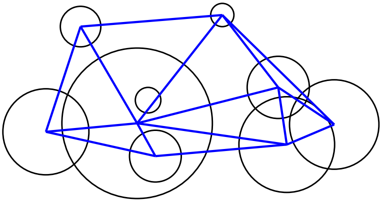
Voronoi cell : $\mathrm{Vor}(b_i) = \{x : \ \ D(x, b_i) \leq D(x, b_j) \forall j\}$

Voronoi diagram of $\mathcal{B}$: $=$ { set of cells $\mathrm{Vor}(b_i), \ \ b_i \in \mathcal{B}\}$

Delaunay triangulations of weighted points



$\mathrm{Vor}(\mathcal{B})$



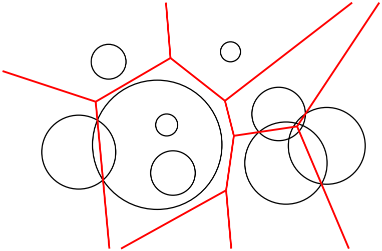
$\mathrm{Del}(\mathcal{B})$ is the nerve of $\mathrm{Vor}(\mathcal{B})$

Theorem

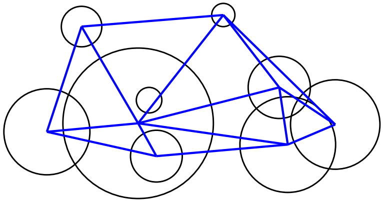
If $\mathcal{B}$ is in general position, then $\mathrm{Del}(\mathcal{B})$ is a triangulation of a subset $\mathcal{P}' \subseteq \mathcal{P}$ of the points

General position : no point of $\mathbb{R}^d$ is at the same distance D to $d+2$ balls

Delaunay triangulations of weighted points



$\mathrm{Vor}(\mathcal{B})$



$\mathrm{Del}(\mathcal{B})$ is the nerve of $\mathrm{Vor}(\mathcal{B})$

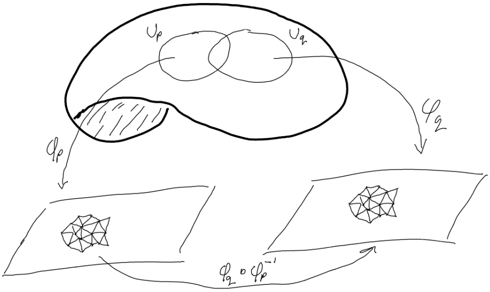
Theorem

If $\mathcal{B}$ is in general position, then $\mathrm{Del}(\mathcal{B})$ is a triangulation of a subset $\mathcal{P}' \subseteq \mathcal{P}$ of the points

General position : no point of $\mathbb{R}^d$ is at the same distance D to $d+2$ balls

Delaunay triangulation of submanifolds of $\mathbb{R}^d$

Euclidean charts and local triangulations

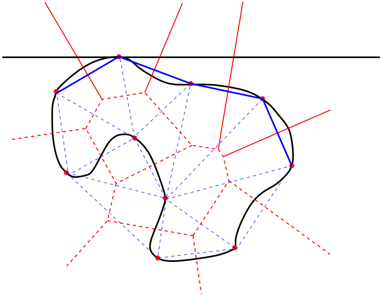


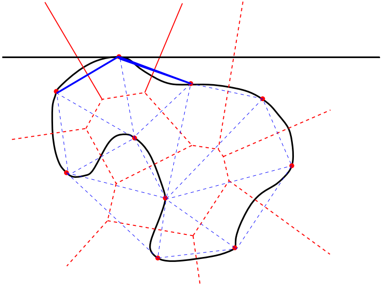
1. Construct local Delaunay triangulations (stars)
2. Make the local triangulations compatible by perturbing the input point set

   $\Rightarrow$ a simplex belongs to the stars of all its vertices
3. Stitch the stars $\Rightarrow$ a combinatorial manifold $\hat{\mathbb{M}}$
4. Exhibit a homeomorphism $\hat{\mathbb{M}} \rightarrow \mathbb{M}$

The tangential Delaunay complex

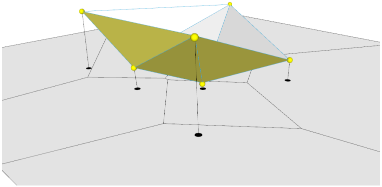
[Freedman 2002], [B.& Flottoto 2004], [B. Ghosh 2014]





Local triangulations

$$\forall p \in \mathcal{P} : \qquad T_p(\mathcal{P}) = \mathrm{star}(p, \mathrm{Del}_{|T_p}))$$



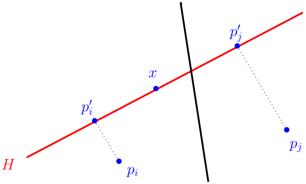
Tangential complex

$$TC(\mathcal{P}) = \{T_p(\mathcal{P}), p \in \mathcal{P}\} \subset \mathrm{Del}(\mathcal{P})$$

Construction of $\mathrm{Del}_{T\mathbb{M}}(\mathcal{P})$

Complexity linear in d, exponential in k

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



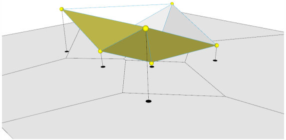
$$\|x - p_i\|^2 \leq \|x - p_j\|^2$$
$$\Leftrightarrow \quad \|x - p'_i\|^2 + \|p_i - p'_i\|^2 \leq \|x - p'_j\|^2 + \|p_j - p'_j\|^2$$

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

$$\mathrm{Vor}(\mathcal{P}) \cap H \quad = \quad \mathrm{Vor}(\psi_p(\mathcal{P})) \qquad \text{(Laguerre diag.)}$$

Corollary : construction of $\mathrm{Del}_{T_p}(\mathcal{P})$

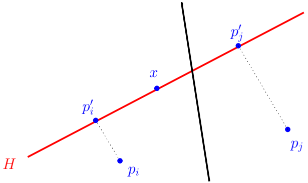
1. project $\mathcal{P}$ in T_p in time $O(dn)$
2. construct $\mathrm{star}_{p_i}(p'_i)$ in $\mathrm{Del}(\pi_{p_i}(\mathcal{P})) \subset T_{p_i}$
3. $\mathrm{star}(p_i) \xleftrightarrow{1-1} \mathrm{star}_{p_i}(p'_i)$



Construction of $\mathrm{Del}_{T\mathbb{M}}(\mathcal{P})$

Complexity linear in d, exponential in k

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



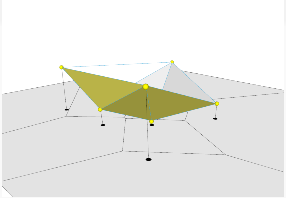
$$\|x - p_i\|^2 \le \|x - p_j\|^2$$
$$\Leftrightarrow \quad \|x - p'_i\|^2 + \|p_i - p'_i\|^2 \le \|x - p'_i\|^2 + \|p_j - p'_j\|^2$$

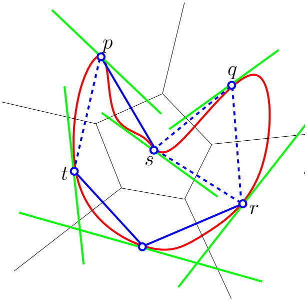
$$\psi_p(p_i) = (p'_i, -\|p_i - p'_i\|^2) \qquad \text{(weighted point)}$$

$$\mathrm{Vor}(\mathcal{P}) \cap H \quad = \quad \mathrm{Vor}(\psi_p(\mathcal{P})) \qquad \text{(Laguerre diag.)}$$

Corollary : construction of $\mathrm{Del}_{T_p}(\mathcal{P})$

1. project $\mathcal{P}$ in T_p in time $O(dn)$
2. construct $\mathrm{star}_{p_i}(p'_i)$ in $\mathrm{Del}(\pi_{p_i}(\mathcal{P})) \subset T_{p_i}$
3. $\mathrm{star}(p_i) \overset{1-1}{\longleftrightarrow} \mathrm{star}_{p_i}(p'_i)$





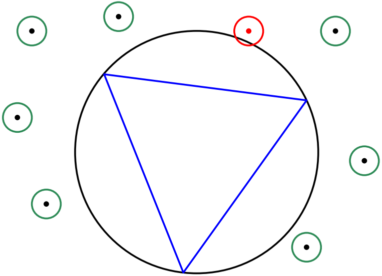
$\mathrm{Vor}(p,t)$ intersects T_t but not T_p

$\Rightarrow$ $p \in \mathrm{star}(t)$

$\Rightarrow$ $t \notin \mathrm{star}(p)$

Conflicting stars

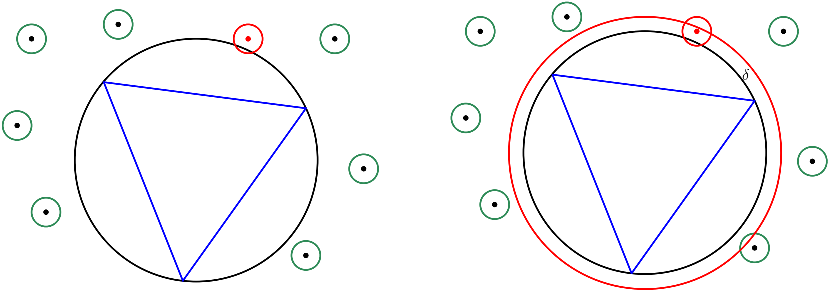
Why do we have conflicts among stars ?



We need the simplices to be protected !

Conflicting stars

Why do we have conflicts among stars ?



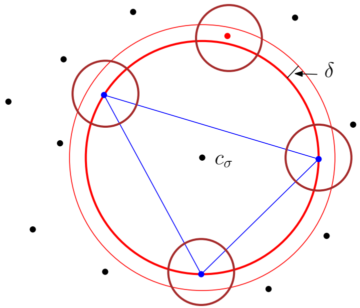
We need the simplices to be protected !

Protection via Random perturbations

A geometric application of the constructive version of the Local Lovasz Lemma [Moser and Tardos 2010]

Random variables : $\mathcal{P}'$ a set of random points $\{p'_1, ..., p'_n\}$

$$p'_i \in B(p_i, \rho), p_i \in \mathcal{P}$$



Algorithm

Input : an ε-net $\mathcal{P}$ of $\mathbb{T}^d$, ρ, δ

while $\exists\ \phi' = (\sigma', p')$ unprotected **do**

resample the points of ϕ'

update $\mathrm{Del}(\mathcal{P}')$

Return $\mathcal{P}'$ and $\mathrm{Del}(\mathcal{P}')$

Let $A_1, ..., A_N$ be a set of bad events
each occurs with $\text{proba}(A_i) \leq p < 1$

Question : what is the probability that none of the events occur ?

The (easy) case of independent events

$$\text{proba}(\neg A_1 \wedge ... \wedge \neg A_N) \geq (1-p)^N > 0$$

What happens if the events are weakly dependent ?

Let $A_1, ..., A_N$ be a set of bad events
each occurs with $\text{proba}(A_i) \leq p < 1$

Question : what is the probability that none of the events occur ?

The (easy) case of independent events

$$\text{proba}(\neg A_1 \wedge ... \wedge \neg A_N) \geq (1-p)^N > 0$$

What happens if the events are weakly dependent ?

Let $A_1, ..., A_N$ be a set of bad events
each occurs with $\text{proba}(A_i) \leq p < 1$

Question : what is the probability that none of the events occur ?

The (easy) case of independent events

$$\text{proba}(\neg A_1 \wedge ... \wedge \neg A_N) \geq (1-p)^N > 0$$

What happens if the events are weakly dependent ?

Lovász local lemma

[Lovász & Erdös 1975]

If, for $i = 1, ..., N$,

1. A_i is independent from all events except $\leq \Gamma$ of them
2. $\text{proba}(A_i) \leq \frac{1}{e\,(\Gamma+1)}$ $\qquad e = 2.718...$

then

$$\text{proba}(\neg A_1 \wedge ... \wedge \neg A_N) > 0$$

A constructive version of LLL

[Moser and Tardos 2010]

$\mathcal{P}$ a finite set of independent random variables

$\mathcal{A}$ a finite set pf events determined by the values of some of the variables of $\mathcal{P}$

Two events are independent iff they don't share any variable

Algorithm

for all $p \in \mathcal{P}$ **do**
  $v_p \leftarrow$ random evaluation of p ;

while some events of $\mathcal{A}$ occur for $(p = v_p, p \in \mathcal{P})$ **do**
  select (arbitrarily) such an event $A \in \mathcal{A}$;
  for all $p \in$ variables(A) **do**
    $v_p \leftarrow$ a new random evaluation of p ;

return $(v_p)_{p \in \mathcal{P}}$;

A constructive version of LLL

[Moser and Tardos 2010]

$\mathcal{P}$ a finite set of independent random variables

$\mathcal{A}$ a finite set pf events determined by the values of some of the variables of $\mathcal{P}$

Two events are independent iff they don't share any variable

Algorithm

for all $p \in \mathcal{P}$ **do**

  $v_p \leftarrow$ random evaluation of p ;

while some events of $\mathcal{A}$ occur for $(p = v_p, p \in \mathcal{P})$ **do**

  select (arbitrarily) such an event $A \in \mathcal{A}$;

  for all $p \in \text{variables}(A)$ **do**

    $v_p \leftarrow$ a new random evaluation of p ;

return $(v_p)_{p \in \mathcal{P}}$;

Moser and Tardos theorem

Hypothesis : for all events A_i, $i \in [1, N]$

1. A_i is independent from all events except $\leq \Gamma$ of them
2. $\text{proba}(A_i) \leq \frac{1}{e\,(\Gamma+1)}$ $\qquad e = 2.718\ldots$

Theorem

- The algorithm assigns values to the variables $\mathcal{P}$ s.t. no event of $\mathcal{A}$ occurs
- The algorithm resamples an event $A \in \mathcal{A}$ at most $\boxed{\frac{1}{\Gamma}}$ times on expectation before finding such an assignment
- The expected total number of resamplings is at most $\boxed{\frac{N}{\Gamma}}$

Hypotheses

- $\mathbb{M}$: a differential submanifold of positive reach of dim. $m \subset \mathbb{R}^d$
- $\mathcal{P}$: an ε-net of $\mathbb{M}$ for some sufficiently small ε
- we know the tangent space T_p at each point $p \in \mathcal{P}$ (of dim. m)

Main result

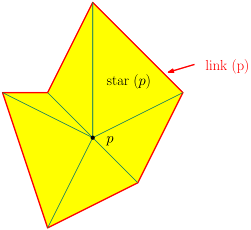
Under those hypotheses, one can choose δ and ρ so that the algorithm terminates and removes all conflicts among stars

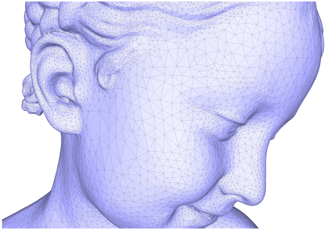
Complexity : $O(d2^{m^2})$, i.e. linear in d and only exponential in m (almost optimal)

Reconstruction of smooth submanifolds

1. For each vertex p, compute its star in $\mathrm{Del}_p(\mathcal{P})$, $\mathrm{star}(p)$
2. Remove conflicts among stars
3. Stitch the stars $\rightarrow \hat{\mathbb{M}}$

$\hat{\mathbb{M}}$ is a PL manifold of dimension m, i.e. the link of each vertex is PL-homeomorphic to a topological sphere of dimension m





The underlying space of a PL manifold is a topological manifold

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

Lemma

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

Proof :

1. Since $\hat{\mathbb{M}}$ contains no conflicting stars, the star of any vertex p of $\hat{\mathbb{M}}$ is identical to $\mathrm{star}_p(p)$, the star of p in $\mathrm{Del}_p(\mathcal{P})$
2. $\mathrm{star}(p) \quad \overset{1-1}{\longleftrightarrow} \quad \mathrm{star}_p(p)$
3. $\mathrm{star}_p(p)$ is a m-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 $\mathrm{Vor}(\psi_p(\mathcal{P})) \;=\; \mathrm{Vor}(p) \cap T_p$, which is bounded)

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

Lemma

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

Proof :

1. Since $\hat{\mathbb{M}}$ contains no conflicting stars, the star of any vertex p of $\hat{\mathbb{M}}$ is identical to $\mathrm{star}_p(p)$, the star of p in $\mathrm{Del}_p(\mathcal{P})$
2. $\mathrm{star}(p) \quad \stackrel{1-1}{\longleftrightarrow} \quad \mathrm{star}_p(p)$
3. $\mathrm{star}_p(p)$ is a m-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 $\mathrm{Vor}(\psi_p(\mathcal{P})) \quad = \quad \mathrm{Vor}(p) \cap T_p$, which is bounded)

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

Lemma

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

Proof :

1. Since $\hat{\mathbb{M}}$ contains no conflicting stars, the star of any vertex p of $\hat{\mathbb{M}}$ is identical to $\mathrm{star}_p(p)$, the star of p in $\mathrm{Del}_p(\mathcal{P})$
2. $\mathrm{star}(p) \quad \overset{1-1}{\longleftrightarrow} \quad \mathrm{star}_p(p)$
3. $\mathrm{star}_p(p)$ is a m-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 $\mathrm{Vor}(\psi_p(\mathcal{P})) \quad = \quad \mathrm{Vor}(p) \cap T_p$, which is bounded)

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

Lemma

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

Proof :

1. Since $\hat{\mathbb{M}}$ contains no conflicting stars, the star of any vertex p of $\hat{\mathbb{M}}$ is identical to $\mathrm{star}_p(p)$, the star of p in $\mathrm{Del}_p(\mathcal{P})$
2. $\mathrm{star}(p) \quad \overset{1-1}{\longleftrightarrow} \quad \mathrm{star}_p(p)$
3. $\mathrm{star}_p(p)$ is a m-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 $\mathrm{Vor}(\psi_p(\mathcal{P})) \quad = \quad \mathrm{Vor}(p) \cap T_p$, which is bounded)

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

Lemma

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

Proof :

1. Since $\hat{\mathbb{M}}$ contains no conflicting stars, the star of any vertex p of $\hat{\mathbb{M}}$ is identical to $\mathrm{star}_p(p)$, the star of p in $\mathrm{Del}_p(\mathcal{P})$
2. $\mathrm{star}(p) \quad \overset{1-1}{\longleftrightarrow} \quad \mathrm{star}_p(p)$
3. $\mathrm{star}_p(p)$ is a m-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 $\mathrm{Vor}(\psi_p(\mathcal{P})) \quad = \quad \mathrm{Vor}(p) \cap T_p$, which is bounded)

Guarantees on the quality of approximation

For ε small enough, there exists Θ_0 s.t.

- $\hat{\mathbb{M}}$ is a PL submanifold of dimension m
- $\hat{\mathbb{M}} \ \subset\ $ tub $(\mathbb{M}, 4\varepsilon^2 \mathrm{rch}(\mathbb{M}))$
- Angles between facets of $\hat{\mathbb{M}}$ and tangent spaces of $\mathbb{M}\ \leq \frac{2\varepsilon}{\Theta_0}$
- All simplices of $\hat{\mathbb{M}}$ are Θ_0-thick
- $\mathrm{proj} : \hat{\mathbb{M}} \to \mathbb{M}$ is a homeomorphism

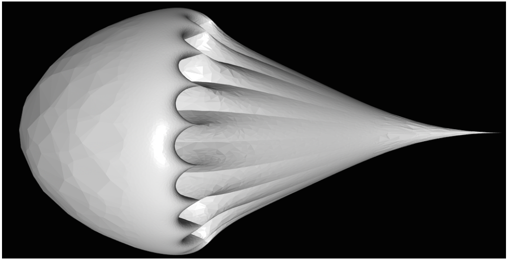
$\Rightarrow \hat{\mathbb{M}}$ is a triangulation of $\mathbb{M}$

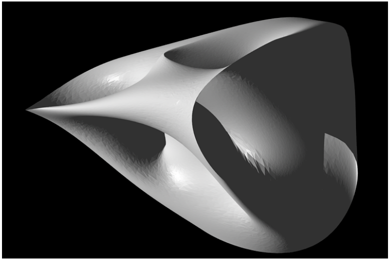
Complexity of the algorithm

- linear in d
- exponential in m

Reconstruction of Rieman surfaces of $\mathbb{R}^8$

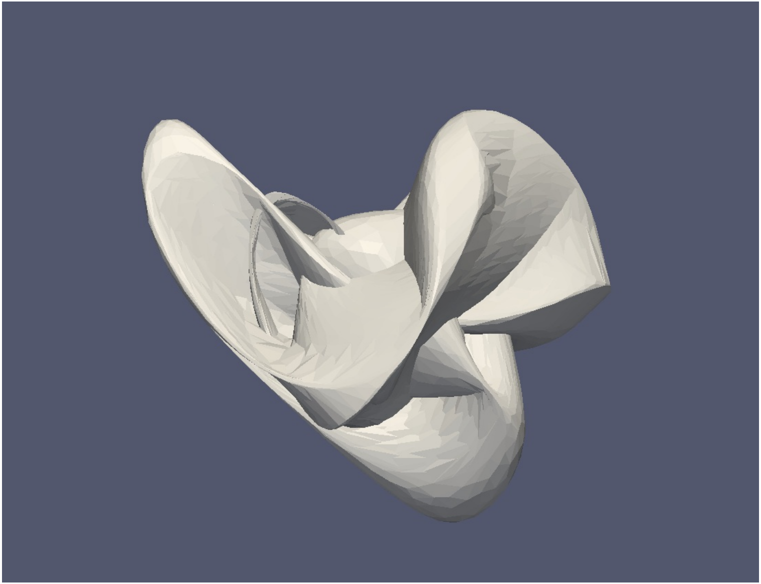
Gudhi implementation [Jamin 2016]





Data provided by A. Alvarez

Triangulation of the space of conformations of C_8H_{16}



Take home messages

Mathematical and algorithmic tools for Geometric Data Analysis

- Manifolds, atlas and charts, reach, nets, manifold hypothesis
- Distance functions, $\check{\mathrm{C}}\mathrm{ech}$ and Delaunay complexes
- Sampling theory and reconstruction algorithms
- Randomized algorithms : RIC, Moser-Tardos algorithm

Good news and questions

- Complexity linear in d : got around the curse of dimensionality
- Practicality is questionnable
- Statistical point of view