Boundary measures for geometric inference

F. Chazal D. Cohen-Steiner Q. Mérigot

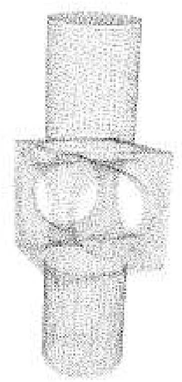
Geometrica

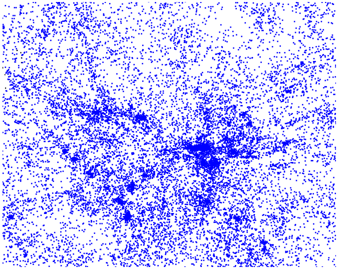
October 16, 2007

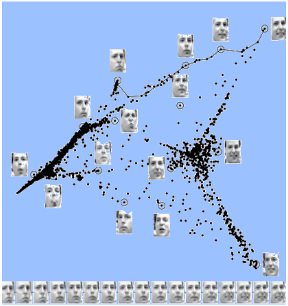
Outline

1. Boundary and curvature measures
2. Some applications in geometric inference
3. Stability of projections & boundary measures

Point cloud geometry

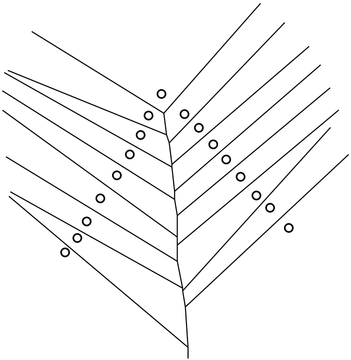






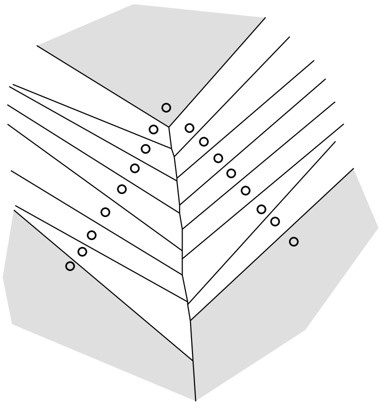
- Given a set of points sampled near an unknown shape, can we infer the geometry of that shape ?

Detecting singularities



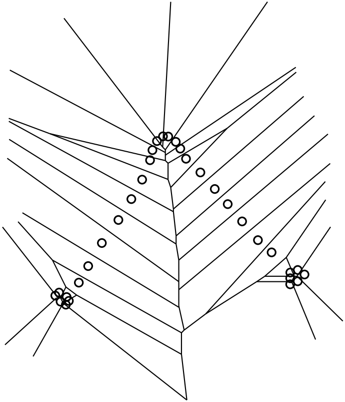
- the volume of a cell is very sensitive to perturbation
- but if one consider the union of Voronoï cells whose site is contained in a given ball...

Detecting singularities



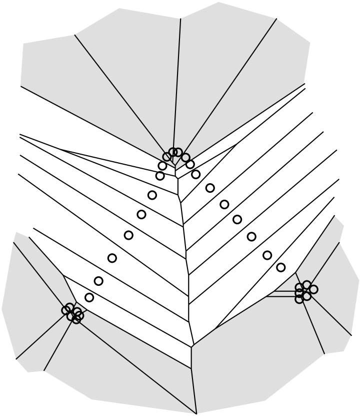
- the volume of a cell is very sensitive to perturbation
- but if one consider the union of Voronoï cells whose site is contained in a given ball...

Detecting singularities



- the volume of a cell is very sensitive to perturbation
- but if one consider the union of Voronoï cells whose site is contained in a given ball...

Detecting singularities

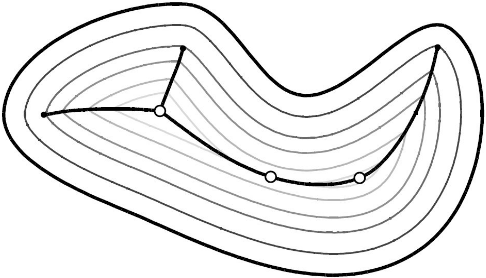


- the volume of a cell is very sensitive to perturbation
- but if one consider the union of Voronoï cells whose site is contained in a given ball...

Projection on a compact set

Definition

The projection $p_K : \mathbb{R}^n \to K \subset \mathbb{R}^n$ maps any point $x \in \mathbb{R}^n$ to its closest point in K. It is defined outside of the *medial axis* of K.



Measures

Definition

A measure is a map μ that takes a subset $B \subset \mathbb{R}^n$ and outputs a nonnegative number $\mu(B)$. It must be additive, *ie* if (B_i) are disjoint subsets,

$$\mu\left(\bigcup_{i\in\mathbb{N}} B_i\right) = \sum_{i\in\mathbb{N}} \mu(B_i)$$

measure = mass distribution

Boundary measure

Definition

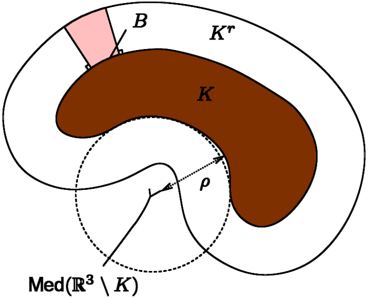
For $E \subset \mathbb{R}^n$, the boundary measure $\mu_{K,E}$ is defined as follows :

$$\forall B \subseteq K,\ \mu_{K,E}(B) = vol^n(\{x \in E \,|\, \mathrm{p}_K(x) \in B\})$$

that is, the n-volume of the part of E that projects on B.

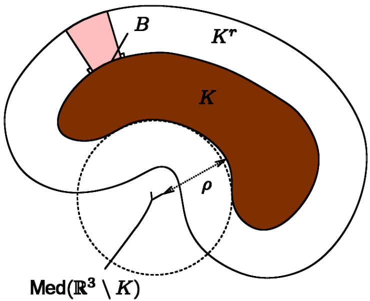
- measure supported in K
- contains a lot of geometric information about K
- defined for *all* compact sets $K \subseteq \mathbb{R}^n$!

Smooth object



- Let $K \subset \mathbb{R}^n$ be an n-dimensional object with smooth boundary.
- The smallest distance between K and its medial axis is called reach(K).
- Take $E = K^r \equiv \{x \in \mathbb{R}^n; \mathrm{d}(x, K) \leq r\}$, assuming $r <$ reach(K).

Smooth object



- $K^r \setminus K$ is the (one to one) image of the map

$$\begin{array}{rcl} f : \partial K \times [0,r] & \rightarrow & \mathbb{R}^n \\ (x,t) & \mapsto & x + t\overrightarrow{n_K}(x) \end{array}$$

Smooth object

- $K^r \setminus K$ is the (one to one) image of the map

$$\begin{aligned} f : \partial K \times (0, r] &\rightarrow \mathbb{R}^n \\ (x, t) &\mapsto x + t\overrightarrow{n_K}(x) \end{aligned}$$

- Hence

$$vol^n(K^r \setminus K) = \int_{\partial K \times (0,r]} |det\ df_{(x,t)} f|\ dxdt$$

- A calculation shows that :

$$det\ df_{(x,t)} f = \sum_{k=0}^{n-1} const(n,k)\, \sigma_k(x)\, t^k$$

where $\sigma_k(x)$ is the degree k elementary symmetric function of the principal curvatures at x : $\sigma_k = \sum_{i(1)<\ldots<i(k)} \kappa_{i(1)} \ldots \kappa_{i(k)}$

Smooth object

Tube formula (Steiner, Weyl, Federer)

If $r < reach(K)$:

$$vol^n(K^r) = vol^n(K) + \sum_{k=1}^{n} const(n,k)[\int_{\partial K} \sigma_{k-1}]\ r^k$$

Smooth object

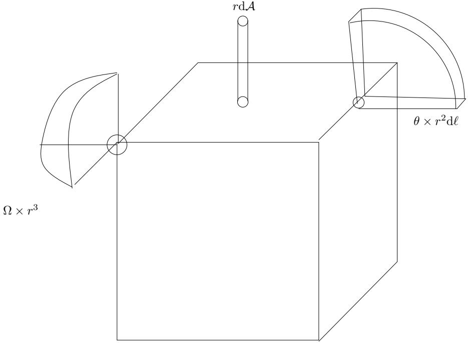
Tube formula (Steiner, Weyl, Federer)

If $r < reach(K)$, for $B \subset K$:

$$\mu_{K,K^r}(B) = \underbrace{vol^n(B)}_{\Phi_K^n(B)} + \sum_{k=1}^{n} const(n,k) \underbrace{\left[\int_{B \cap \partial K} \sigma_{k-1}\right]}_{\Phi_K^{n-k}(B)} r^k$$

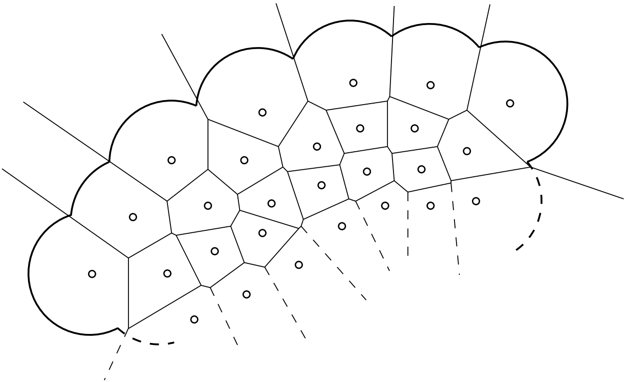
- The Φ_K^i are the (signed) *curvature measures* of K.
- If K is d-dimensional, they vanish identically for $i > d$.
- $\Phi_K^0(K)$ is the Euler characteristic of K.
- They are intrinsic, *i.e.* they do not depend on the embedding.

Convex polyhedron



- the boundary measure of a convex polyhedron K can be decomposed as a sum : $\mu_{K,K^r}(B) = \sum_{k=0}^{n} const(n,k)\, \Phi_{n-k}(B)\, r^k$.
- the curvature measure Φ_K^i is the i-dimensional measure supported on the i-skeleton of K whose density is the local external dihedral angle.

The boundary measure of a point cloud



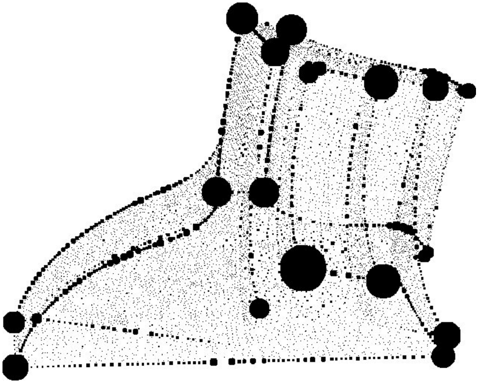
the boundary measure is a sum of weighted Dirac masses :

$$\mu_{C,C^r} = \sum_i vol^n(\mathrm{Vor}(x_i) \cap C^r)\delta_{x_i}$$

Outline

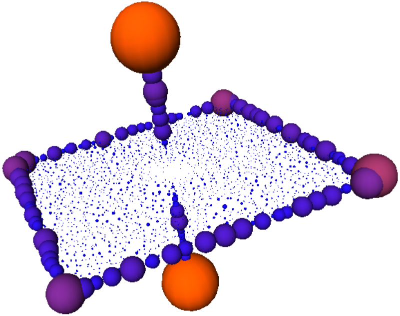
1. Boundary and curvature measures
2. Some applications in geometric inference
3. Stability of projections & boundary measures

Recovering sharp features using boundary measures



Here, the volumes of the Voronoï cells are evaluated using a Monte-Carlo method. Cost scales linearly with ambient dimension. Approximation error does not depend on the ambient dimension.

Dimension estimation using boundary measures

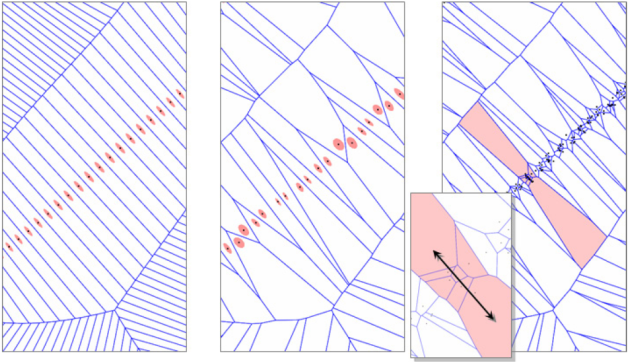


Here, the volumes of the Voronoï cells are evaluated using a Monte-Carlo method. Cost scales linearly with ambient dimension. Approximation error does not depend on the ambient dimension.

The extremal measure of a point cloud

- If we let r go to infinity, we obtain a measure that says how extreme a point is in the data.
- it is actually the 0^{th} curvature measure of the convex hull of the point cloud.
- Monte Carlo : repeatedly pick a random direction and increment the mass of the lowest point in that direction.

Normal & principal directions estimation



Replace **volumes** of Voronoï cells by their **covariance matrices**.
This gives a tensor-valued measure.

- small/large eigenvalues $\simeq$ tangent/normal space
- principal curvatures/directions ?

Outline

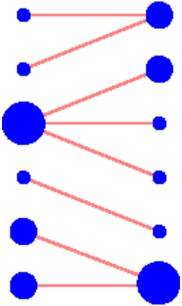
Stability

- Assume point cloud C samples compact K well, *e.g.* $d_H(K, C) \leq \varepsilon$. This means that

$$C \subset K^\varepsilon \text{ and } K \subset C^\varepsilon$$

- Are $\mu_{K,E}$ and $\mu_{C,E}$ close ? In which sense ?

Wasserstein distance



- Assume measures μ and ν are discrete : $\mu = \sum_i c_i \delta_{x_i}$, $\nu = \sum_j d_j \delta_{y_j}$ we suppose that mass(μ) = mass(ν)
- a transport plan between is a set of nonnegative coefficients p_{ij} specifying the amount of mass which is transported from x_i to y_j, with

$$\sum_i p_{ij} = d_j \text{ and } \sum_j p_{ij} = c_i$$

- the cost of a transport plan is $C(p) = \sum_{ij} \|x_i - y_j\| \, p_{ij}$
- $W(\mu, \nu) = \inf_p C(p)$

Kantorovich-Rubinstein theorem

Theorem

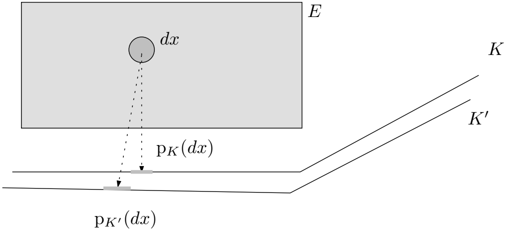
For two measures μ and ν with common finite mass and bounded support,

$$W(\mu, \nu) = \sup_f \left| \int f \, d\mu - \int f \, d\nu \right|$$

where the sup is taken over all 1-Lipschitz functions $\mathbb{R}^n \to \mathbb{R}$.

Convergence for Wasserstein distance implies pointwise convergence after convolving with a “tent” function for example.

Wasserstein distance between boundary measures



We consider the following transport plan : the element of mass $p_K(x)dx$ coming from an element of mass dx at $x \in E$ will be transported to $p_{K'}(x)dx$.
the total cost of this transport is :

$$\int_E \|p_K(x) - p_{K'}(x)\| \, dx = \|p_K - p'_K\|_{L^1(E)}$$

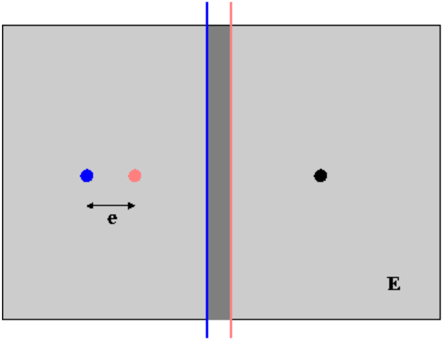
A convergence result for projections

Theorem (Federer)

If (K_n) converges to K for the Hausdorff distance, with $\mathrm{reach}(K_n) \geq \mathrm{reach}(K) > r > 0$, then p_{K_n} converges uniformly to p_K on K^r.

Not sufficient for our purpose.

Example



the projections p_K and p_{K_ε} may differ a lot only on a small neighborhood of the medial axis of K.

A L^1 stability theorem for projections

Theorem

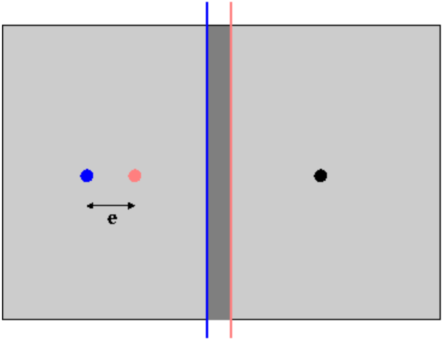
If E is an open set of $\mathbb{R}^n$ with rectifiable boundary, and K and K' are two *close enough* compact subsets :

$$\|p_K - p_{K'}\|_{L^1(E)} := \int_E \|p_K - p_{K'}\|$$
$$\leq C(n)[vol^n(E) + \mathrm{diam}(K) vol^{n-1}(\partial E)]\sqrt{R_K \mathrm{d}_H(K, K')}$$

where $R_K = \sup_{x \in E} \mathrm{d}(x, K)$.

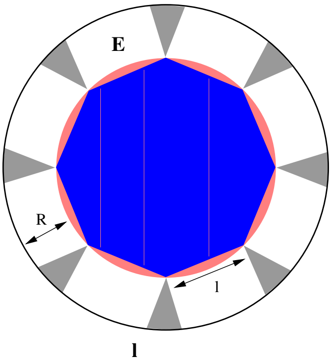
1. *close enough* means that $\mathrm{d}_H(K, K')$ does not exceed $\min(R_K, \mathrm{diam}(K), \mathrm{diam}(K)^2 / R_K)$
2. $C(n) = O(\sqrt{n})$

Optimality of the stability theorem for projections (1)



- taking E included in the dark region shows that the term $vol^{n-1}(\partial E)$ is necessary.

Optimality of the stability theorem for projections (2)

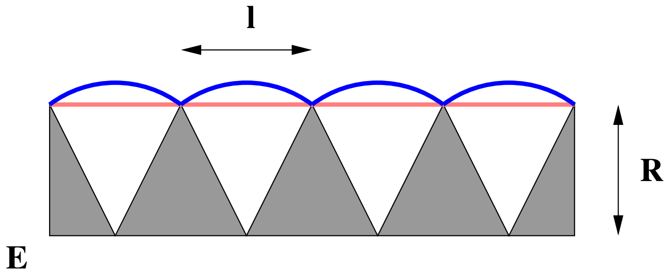


K = unit disk in $\mathbb{R}^2$
K_ℓ = polygonal approximation with sidelength ℓ

- $d_H(K, K_\ell) \simeq \ell^2$.
- A certain fraction of the mass of $\mu_{K_\ell,E}$ is concentrated at the vertices. Hence $W(\mu_{K_\ell,E}, \mu_{K,E}) = \Omega(\ell)$
- So $\|p_K - p_{K_\ell}\|_{L^1(E)} \simeq d_H(K, K_\ell)^{1/2}$

Optimality of the stability theorem for projections (2)

K = unit line segment in $\mathbb{R}^2$
K_ℓ = circle arcs whose center lie is at distance $R > 0$ of K



- $d_H(K, K_\ell) \simeq \ell^2$.
- Almost half of the mass of $\mu_{K_\ell,E}$ is concentrated at the vertices. Hence $W(\mu_{K_\ell,E}, \mu_{K,E}) = \Omega(\ell)$
- So $\|p_K - p_{K_\ell}\|_{L^1(E)} \simeq d_H(K, K_\ell)^{1/2}$

Sketch of proof

Lemma

Function $v_K : x \mapsto \|x\|^2 - d_K^2(x)$ is convex.

$$\begin{aligned} v_K(x) &= \|x\|^2 - \inf_{y \in K} \|x - y\|^2 \\ &= \sup_{y \in K} [\|x\|^2 - \|x - y\|^2] \end{aligned}$$

which is a sup of affine functions.

Sketch of proof

Lemma

$\nabla v_K = 2p_K$ almost everywhere.

$$\begin{aligned}\nabla v_K(x) &= 2x - 2d_K(x).\nabla d_K(x) \\ &= 2x - 2d_K(x)\frac{x - p_K(x)}{d_K(x)} \\ &= 2p_K(x)\end{aligned}$$

We want to bound the L^1 distance between the gradients of two convex functions.

Sketch of proof

What do we know about v_K and $v_{K'}$?
We assume $\mathrm{d}_H(K, K') = \varepsilon \leq \mathrm{diam}(K)$ and let $R_K = \sup_E \mathrm{d}_K$

1. v_K and $v_{K'}$ are uniformly close :

$$
\begin{aligned}
|v_K - v'_K| &= |d_K^2 - d_{K'}^2| \\
&= |d_K - d_{K'}|(d_K + d_{K'}) \\
&\leq 3\, R_K \varepsilon
\end{aligned}
$$

2. ∇v_K and $\nabla v_{K'}$ are both contained in $2K \cup 2K'$, which has diameter $\leq 2\,\mathrm{diam}(K) + 4\,\varepsilon \leq 6\,\mathrm{diam}(K)$.

An inequality

Lemma

If $f, g : E \to \mathbb{R}$ are convex and $k = \mathrm{diam}(\nabla f(E) \cup \nabla g(E))$ then

$$\|\nabla f - \nabla g\|_{\mathrm{L}^1(E)} \leq C(n)\,[vol^n(E) + k\ vol^{n-1}(\partial E)]\ \|f - g\|_\infty^{1/2}$$

$$+\ C(n)\ vol^{n-1}(\partial E)\ \|f - g\|_\infty$$

1-dimensional case

Lemma

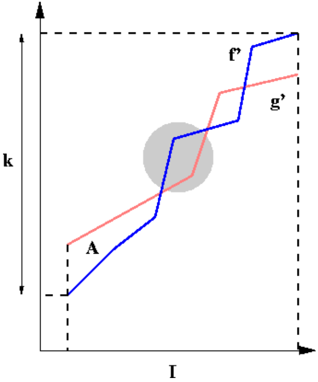
Let $f, g : I \to \mathbb{R}$ be two convex functions such that que $\forall x \in I,\ |f(x) - g(x)| \leq \delta$ and $\mathrm{diam}(f'(I) \cup g'(I)) \leq k$. Then,

$$\int_I |f' - g'| \leq C[(\mathrm{length}(I) + k)\sqrt{\delta} + \delta]$$

Proof of the 1-dimensional case

Lemma

$$\int_I |f' - g'| \leq C[(\text{length}(I) + k)\sqrt{\delta} + \delta]$$



1. No disk of radius $\sqrt{2\delta/\pi}$ can fit between the graphs of f' and g'.
2. Hence $A \subset \partial A^{\sqrt{2\delta/\pi}}$.
3. $\text{length}(\partial A) \leq 2(k + \text{length}(I))$
4. For any curve $S \subset \mathbb{R}^2$:

$$\text{area}(S^r) \leq \frac{9\pi}{4}[\text{length}(S).r + r^2]$$

Proof of the general case

Considering restrictions of f and g to an affine line $\ell \subset \mathbb{R}^n$ with direction v ($\|v\| = 1$), we get :

$$\int_{\ell \cap E} \left| f'_{|\ell \cap E} - g'_{|\ell \cap E} \right| \leq C[\,\text{length}(\ell \cap E) + k.\,(\#(\ell \cap \partial E) + \sqrt{\delta})\,]\sqrt{\delta}$$

$$\int_{\ell \cap E} |\langle \nabla f - \nabla g | v \rangle| \leq C[\,\text{length}(\ell \cap E) + k.\,(\#(\ell \cap \partial E) + \sqrt{\delta})\,]\sqrt{\delta}$$

integrating over all lines and using the Crofton formulas gives the n-dimensional bound.

Stability of boundary measures

Theorem

If K is a fixed compact set, and E an open set with smooth boundary, then

$$\mathrm{W}(\mu_{K,E}, \mu_{K',E}) \leq C(n, E, K)\, \mathrm{d}_H(K, K')^{1/2}$$

as soon as K' is close enough to K.

A similar result holds for μ_{K,K^r} and μ_{K',K'^r}.

Estimating curvature measures

1. for any K with positive reach, there exists measures $\Phi_{K,i}$ such that for $r < \mathrm{reach}(K)$, $\mu_{K,r}(B) = \sum_{i=1}^{n} \Phi_K^{n-i}(B) r^i$
2. can be computed knowing only the boundary measures for $n+1$ values $r_0 < \ldots < r_n$: denote the result by $\Phi_{K,i}^{(r)}$.

Corollary

If $\mathrm{reach}(K) > r_n$ and K' is *close* to K, there is a constant $C(K, n, (r))$ such that

$$\mathrm{d_{bL}}\left(\Phi_{K,i}, \Phi_{K',i}^{(r)}\right) \leq C(K, n, (r))\, \mathrm{d}_H(K, K')^{1/2}$$

$\mathcal{C}^0$ (Hausdorff) closeness implies closeness of differential properties at a given scale.

Discussion

1. The boundary measure and its tensor version encode “much” of the geometry of a compact set.
2. these measures depend continuously on the compact set for the Hausdorff distance (whatever the compact).
3. what can we do when the underlying shape has zero reach ?
4. what happens if we replace nearest neighbors by approximate nearest neighbors in the Monte-Carlo algorithm ?
5. how can we deal with outliers ?