Reminder: Please submit short project proposal on Canvas.

If you would like a partner, and haven't found one, just say so in your proposal, and I will try to match people based on interests.

Today: When can we recover homology from samples?

Recall: Hausdorff distance btw. sets

$$d_H(X,Y) = \max\left\{\sup_{x\in X}\inf_{y\in Y} d(x,y),\ \sup_{y\in Y}\inf_{x\in X} d(y,x)\right\}$$

Last time:

$$d_I(PH_k(R(X)), PH_k(R(Y))) \leq 2d_H(X,Y)$$

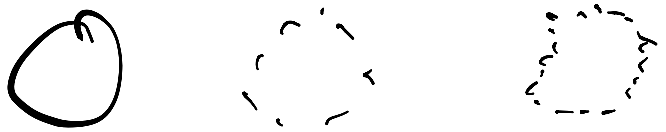
this doesn't say anything abt "ground truth"

Oudot Ch. 4.

Let K be some compact set in Euclidean space $\mathbb{R}^d$. Examples: sub-manifolds of $\mathbb{R}^d$, convex sets, etc.

K is "ground truth"

we sample points $X \sim K$, or $X \sim K + \underset{\text{noise}}{\varepsilon}$



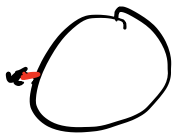
K $\quad$ X $\quad$ X

K has some ground truth $H_k(K)$

can we recover $H_k(K)$ from samples X?

we'll look at offsets:

i) we define distance to K: $d_K(x) = \min_{y \in K} \|x-y\|$



K

$d_K : \mathbb{R}^d \to \mathbb{R}_+$

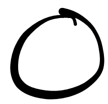
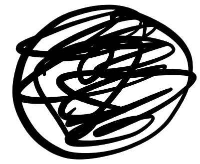
r-offset of K: $K_r := d_K^{-1}((-\infty, r])$

$= \emptyset$ if $r<0$

$= K$ if $r=0$

$= \mathbb{R}^d$ if $r=\infty$

$K_r \subseteq K_s \ \forall r \leq s \to$ filtration.

K:  S^1 $\to K_{0.1}$  annulus $\quad K_5$: 

Note: r-offset vs Čech.

$K_r \neq \check{C}(K;r)$ $\to$ simplicial complex

$\hookrightarrow$ continuous

Reach:

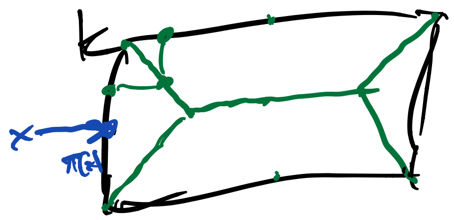
Projection set of $x \in \mathbb{R}^d$

$$\Pi_K(x) = \{ y \in K \mid d(x,y) = d_K(x) \}$$

for many $x \in \mathbb{R}^d$, this is a single pt.

on medial axis $|\Pi_K(x)| \geq 2$.

$$M(K) = \{ x \in \mathbb{R}^d \mid |\Pi_K(x)| \geq 2 \}.$$



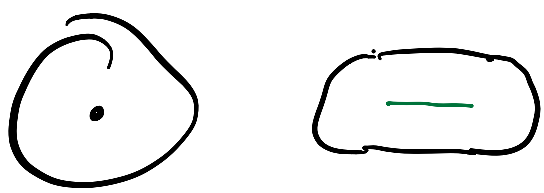
$\leftarrow \mathrm{rch}(K) = 0$

Reach of K: min. distance btw. K and $M(K)$

at a point $\mathrm{rch}(K, x) = d_{M(K)}(x)$

$$\mathrm{rch}(K) = \min_{x \in K} \mathrm{rch}(K, x)$$

$$= \min_{x \in K} d_{M(K)}(x)$$



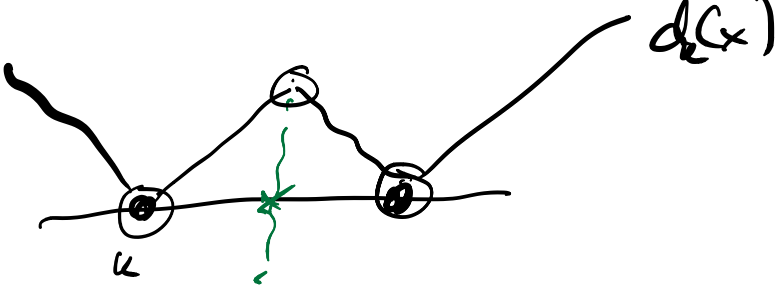
When K has $\mathrm{rch}(K) > 0$, there is a nbhd of K which retracts onto K: if $r < \mathrm{rch}(K)$

$$K_r \to K \quad \text{via} \quad x \mapsto \Pi_K(x)$$

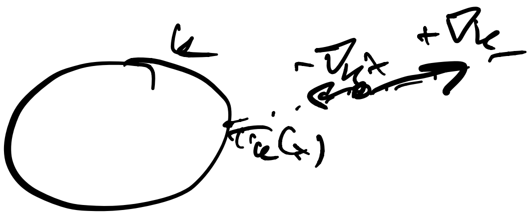
Flow along $\nabla d_K(x)$

idea: d_K is 1-lipschitz over $\mathbb{R}^d$

differentiable over $\mathbb{R}^d \setminus (K \cup M(K))$



$$D_K(x) := \frac{x-\pi_K(x)}{\|x-\pi_K(x)\|} = \frac{x-\pi_K(x)}{d_K(x)}$$



this is cts, so $-D_K$ can be integrated into a flow.

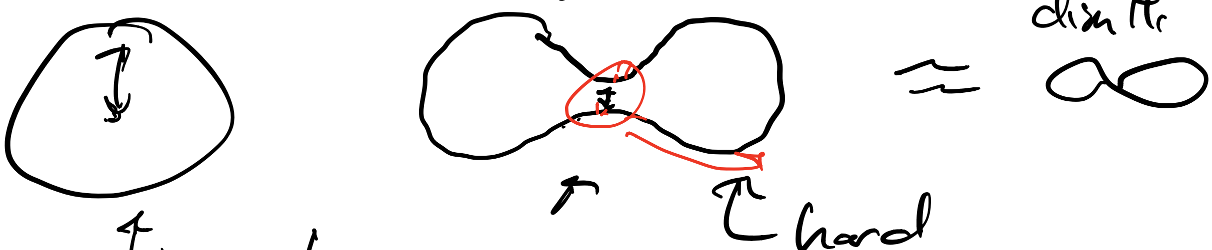
if $\mathrm{rch}(K) > 0$, K_r retracts onto K using $-D_K$

Thm: Niyogi, Smale, Weinberger 05/06, 08.
Let K be cpt. submanifold of $\mathbb{R}^d$, with $\mathrm{rch}(K) > 0$. Let P be a pointcloud $P \subseteq \mathbb{R}^d$ s.t. $d_H(K, P) = \varepsilon < \sqrt{\frac{3}{20}}\,\mathrm{rch}(K)$. Then for any $r \in (2\varepsilon, \sqrt{\frac{3}{5}}\,\mathrm{rch}(K)]$ P_r deformation retracts onto K.

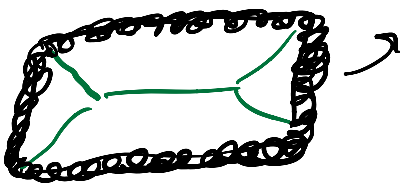
$K \simeq P_r$

$H_k(K) \cong H_k(P_r)$

dim $H_1 = 1$ dim $H_1 = 2$



easy to reconstruct H_k hard

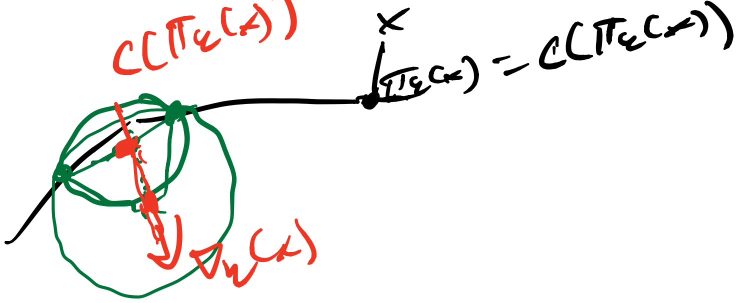


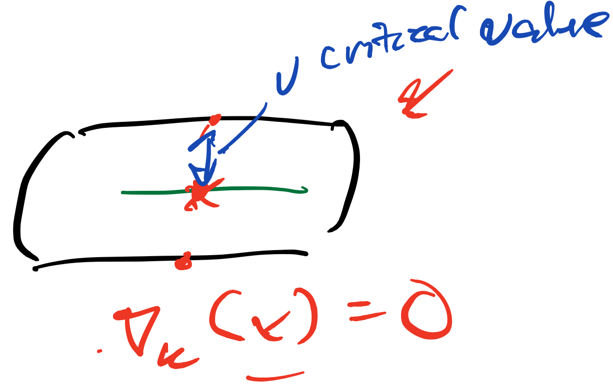
result above is useless

weak feature size. Chazal, Cohen-Steiner, Lieutier ~ 2009

can extend D_K to medial axis.

$D_K(x) = \dfrac{x - c(\Pi_K(x))}{d_K(x)}$ $c(\Pi_K(x))$: center of $\Pi_K(x)$





def: A point $x \in \mathbb{R}^d$ is critical if its generalized gradient $\nabla_K(x) = 0$, or equivalently if $x \in$ cvx hull of $\Pi_K(x)$

A value is critical if $v = d_K(x)$ for some critical point x

---

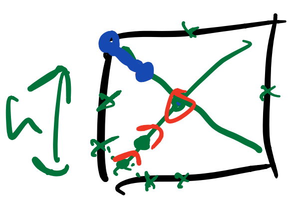
Lemma: if $0 \le r < s$ are s.t. there are no crit. values of d_K in $[r, s]$, then K_s def. ref. onto K_r $\quad K_r \simeq K_s$

weak feature size:

$$wfs(K) = \inf\{r > 0 \mid r \text{ is a crit. value of } d_K\}$$

comment: critical points all live on $M(K)$

$\Rightarrow rch(K) \le wfs(K)$



$rch(K) = 0$ (one crit. pt. at center)

$wfs(K) = \frac{h}{2}$

Moral: wfs is really what governs htpy equivalence

Thm (C, C-S, L): Let K be a compact set in $\mathbb{R}^d$, w/ wfs > 0, let $P \subseteq \mathbb{R}^d$,

$d_H(K, P) = \varepsilon < \frac{1}{4}\,\mathrm{wfs}(K)$

Then for any r, s s.t. $\varepsilon < r < r + 2\varepsilon \leq s < \mathrm{wfs}(K) - \varepsilon$

the img $H_*(P_r) \to H_*(P_s)$ induced by $P_r \to P_s$

is isomorphic to $H_*(K_r)$ $\forall\, r \in (0, \mathrm{wfs}(K))$

$$\begin{gathered} H_*(K_r) \\ \cong \\ K \\ H_*(K) \end{gathered}$$

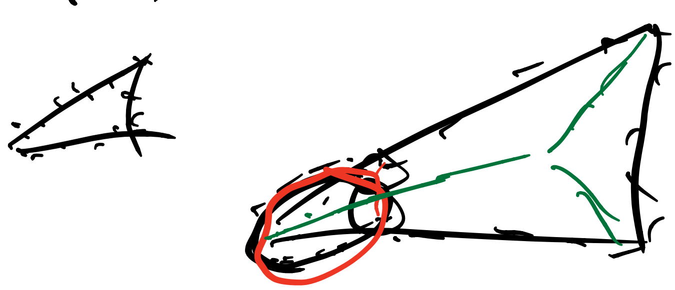
$$\begin{tikzcd}
\mathrm{PH}(K_r) \arrow[r] & K_{r-\varepsilon} \arrow[rr] \arrow[dr] & & K_{r+\varepsilon} \arrow[rr] \arrow[dr] & & K_{s+\varepsilon} \\
\mathrm{PH}(P_r) \arrow[rr] & & P_r \arrow[ur] \arrow[rr] & & P_s \arrow[ur] &
\end{tikzcd}$$

any bars in barcode longer than 2ε in range $(0, \mathrm{wfs}(K))$ are "real"

Topological SNR: $\dfrac{\mathrm{wfs}(K) - 2\varepsilon}{2\varepsilon}$ ← signal bars ← noise bar

---

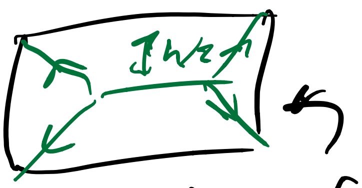
Note: P does not need to be $\subseteq K$ to have $d_H(P, K) \leq \varepsilon$.



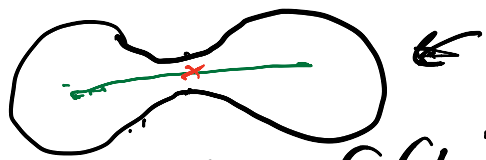
in presence of noise, no single value of r may have $P_r \simeq K$, but we can still recover $H_k(K)$ using persistent homology.

---

wfs: what governs the ability to reconstruct.



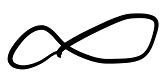
$\mathrm{rch}(K) \neq \mathrm{wfs}(K)$



still in trouble.

$\mathrm{rch}(K) = \mathrm{wfs}(K)$

$\mathbb{R}$



need a lot of samples w/ no noise

or noise $\ll \mathrm{rch}(K)$

P_r offset space.

$\check{C}(P;r)$

P_r are continuous, unions of balls

$$P_r = \bigcup_{x \in P} B(x;r)$$

Čech complex: Nerve of cover of P_r given by these balls.

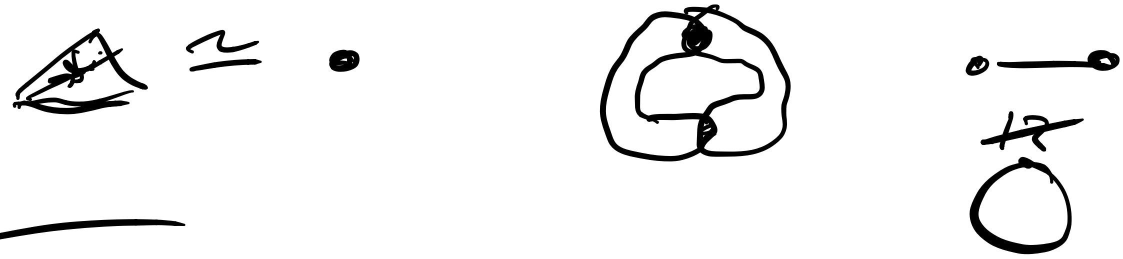
$$\check{C}(P;r) = N(\{B(x;r)\})$$

Recall: simplex $(x_0 \dots x_k) \in \check{C}(P; r)$ if $\bigcap B(x_i; r) \neq \emptyset$

Recall Nerve Thm: Let X be paracompact space, and $\mathcal{U}$ an open cover of X, s.t. intersections of sets either empty or contractible. Then $N(\mathcal{U}) \simeq X$

$B(x; r)$ is convex. So if $\bigcap_{x \in S} B(x; r) \neq \emptyset$ then $\bigcap \dots$ is also convex.

Convex sets are contractible:



Persistent Nerve thm:

Let $X \subseteq X'$ paracompact, and $\mathcal{U}, \mathcal{U}'$ be open covers w/ same parameter set A. Assume that $U_a \subseteq U'_a$ $\forall a \in A$, and $\mathcal{U}, \mathcal{U}'$ satisfy conditions of Nerve thm. Then htpy equivalences commute w/ inclusions

$$\begin{tikzcd} H_k(X) \arrow[r] & H_k(X') \\ H_k(N(\mathcal{U})) \arrow[r] \arrow[u, "\cong"] \arrow[ur, phantom, "\circlearrowleft"] & H_k(N(\mathcal{U}')) \arrow[u, "\cong"'] \end{tikzcd}$$

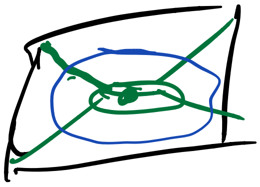
Corollary: Let K be compact set in $\mathbb{R}^d$ w/ $wfs(K) > 0$. Let $P \subseteq \mathbb{R}^d$, $d_H(P, K) = \varepsilon$, $\varepsilon < \frac{1}{4} wfs(K)$, then there is a "sweet range" T: $(\varepsilon, wfs(K) - \varepsilon)$, where barcode of $\check{C}(P; r)$ has property:

bars w/ length $> 2\varepsilon$ in sweet range encode $H_k(K)$

Moral: Čech complex filtrations have nice theoretical properties.

μ-reach: a point is called μ-critical if $\|\nabla_K(x)\| \leq \mu$. $M_\mu(K)$ the μ-medial axis.

μ-reach: $r_\mu(K) = \min_{x \in K} d_{M_\mu(K)}(x)$



interpolates btw. $\mathrm{rch}(K)$ and $\mathrm{wfs}(K)$.

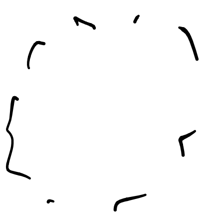
Thm: Let K be compact w/ positive μ-reach for some $\mu \in (0,1]$ let $d_H(P,K) = \varepsilon$
$\varepsilon < \frac{\mu^2}{5\mu^2+2} r_\mu(K)$. Then for any r s.t. $\frac{4\varepsilon}{\mu^2} \le r \le r_\mu(K) - 3\varepsilon$,

$$P_r \simeq K_s \text{ for any } s \in (0, \mathrm{wfs}(K))$$

$\Rightarrow$ μ-reach tells us there is a nice range where there is <u>no</u> topological noise in barcode.

Difficulties:

1) if we don't know K, we don't nec. know $\mathrm{rch}(K)$ or $\mathrm{wfs}(K)$ but might be able to bound



2) we need $d_H(P, K) < \varepsilon$

we can accomodate bounded noise models,

but we don't have theory to deal with tails in noise.

$X \sim K + N(0, \sigma^2)$

not robust to <u>outliers</u>.

3) $\check{C}(P; r)$ is hard to compute.

checking intersections of balls grows exp. in d. where $\mathbb{R}^d$ is ambient space.

$\check{C}(P, \infty)$ is very large.

$D(P, r)$: restriction of $\check{C}$ to delaunay triangulation is htpy eq.

construction in high-dims difficult.