Prominence of the peaks of a real-valued function (a.k.a. persistence theory in degree 0)

X: topological space $\qquad f : X \to \mathbb{R}$

(no assumption on f, not even continuity)

Let $x \in X$ be a peak of f

(i.e. $\exists\, U \ni x$ open in X s.t. $f(x) = \max\limits_U f$)

$\boxed{\text{Def:}}$ $\forall\, t \leq f(x)$, let $C_t(x)$ = path-connected component of $f^{-1}((-\infty, t])$ that contains x.

<u>Note:</u> $t' \leq t \leq f(x) \Rightarrow C_{t'}(x) \supseteq C_t(x)$.

$\boxed{\text{Def:}}$ $h(x) := \sup I(x) \in \mathbb{R} \cup \{-\infty\}$

where $I(x) := \{ t \leq f(x) \mid \exists\, y \text{ peak of } f \text{ s.t. } f(y) > f(x) \text{ and } C_t(y) = C_t(x) \}$

$\hookrightarrow$ "birth" of $x := f(x)$; "death of x" $:= h(x)$

"prominence"/"persistence" of $x := f(x) - h(x) \in \mathbb{R}^+ \cup \{+\infty\}$

"persistence interval"/"lifespan" of $x := (h(x), f(x)]$

can be $-\infty$

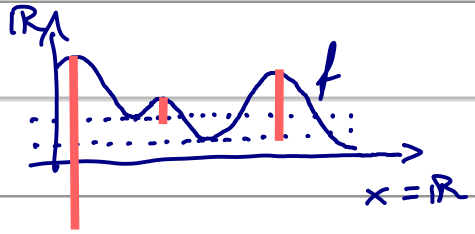
$\boxed{\text{Def:}}$ Barcode of f: $B(f) := \{ (h(x), f(x)] \mid x \text{ peak of } f \}$.

multi-set, with multiplicities

Example:

(see slide)

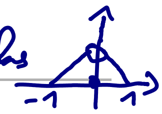
—: $B(f)$



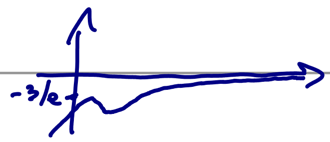
Note: for $f : X \to \mathbb{R}^+$, infinite prominence $\iff$ prominence > height

▶ pathological cases (which one would like to avoid):

- $\mathrm{id} : \mathbb{R} \to \mathbb{R}$ $\rightsquigarrow$ no peaks $\Rightarrow B(f) = \emptyset$
- $x \mapsto 1 - |x|$ over $[-1, 1] \setminus \{0\}$ and $0 \mapsto 0$ $\rightsquigarrow$ no peaks



- $-(x^3+2)\,e^{-x}$ over $\mathbb{R}^+$ $\rightsquigarrow$ unique peak at $x = 1$, of height $-3/e$ and of infinite prominence whereas $\lim_{t \to +\infty} f = 0$.



↳ $\boxed{\text{Hyp.:}}$ (i) the path-connected components of the peaks generate the entire super-level sets of f:

$$\forall t \in \mathbb{R},\quad f^{-1}((-\infty, t]) = \bigcup_{\substack{x:\ \text{peak} \\ f(x) \geq t}} C_t(x).$$

ensures the entire super-level sets are accounted for

(ii) f has a finite number of peaks

for simplicity

(iii) a total order on the peaks of f is given, which is compatible with the pre-order defined by f.

allows us to break ties among peaks of same height (may require axiom of choice)

▶ typical use cases (imply (i) is satisfied):

- X compact, f continuous
- $X = \mathbb{R}^d$, f continuous, non-negative, vanishes at infinity.

Def: For any peak x such that $h(x) > -\infty$, the set

$$\{ y : \text{peak of } f \mid f(y) > f(x) \text{ and } C_t(y) = C_t(x) \ \underline{\underline{\forall t \in I(x)}} \}$$

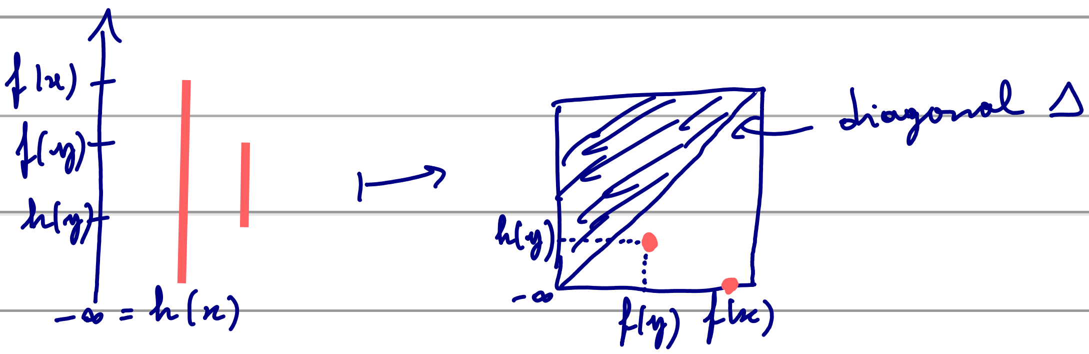
is finite and non-empty, hence it has a unique maximum, called the <u>parent of x</u>.

<u>Remarks:</u>

- $(i) + (ii) \Rightarrow$ the set is non-empty for any x s.t. $h(x) > -\infty$.
- $(ii) + (iii) \Rightarrow$ the maximum exists and is unique as soon as the set is non-empty.

Def: Barcode $B(f) \longleftrightarrow$ <u>Persistence diagram</u> $D(f)$

$$(h(x), f(x)] \longmapsto (f(x), h(x)) \in \mathbb{R} \times [-\infty, +\infty)$$



Theorem: (stability)

$\forall f, g : X \to \mathbb{R}$ satisfying (i)–(iii),

$$d_b\big(D(f), D(g)\big) \leq \|f-g\|_\infty$$

where:

- $\|f-g\|_\infty := \|f-g\|_{C^0(X)} = \sup_{x \in X} |f(x)-g(x)|$
- $d_b\big(D(f), D(g)\big) := W_\infty\big(D(f) \cup \Delta,\ D(g) \cup \Delta\big)$

Wasserstein distance

infinite multiplicity

$$= \inf_{\substack{\gamma : D(f)\cup\Delta \to D(g)\cup\Delta \\ \text{bijection}}} \sup_{x \in D(f)} \|x - \gamma(x)\|_\infty$$

Algorithm: (variant of Kruskal's MST algorithm)

(enough to compute path-connected components)

Input: graph $G=(V,E)$, function $f : V \sqcup E \to \mathbb{R}$

Hyp: graph filtration, i.e. $f(u,v) \leq \min\{f(u), f(v)\}$ for every edge $(u,v) \in E$.

Pre-processing:

⚠ only a pre-order (need to break ties)

- sort $V \sqcup E$ by increasing lexicographic order $(-f\text{-value}, \dim)$

  ↳ get a sequence $\sigma_1, \sigma_2, \dots, \sigma_m$ of vertices / edges such that $f(\sigma_i) > f(\sigma_j)$ or $\big(f(\sigma_i) = f(\sigma_j)$ and $\dim \sigma_i \leq \dim \sigma_j\big)$ $\forall i \leq j$.

- initialize a union-find data structure $\mathcal{U}$.

Main loop :

for $i = 1$ to m do :

if σ_i is a vertex v then :

- create new entry $e_v := \{v\}$ in $\mathcal{U}$

(records the birth of v as a connected comp.)

else σ_i is an edge (u, v) and then do :

- find the entries e_{u_0} and e_{v_0} in $\mathcal{U}$ that contains u and v respectively

if $u_0 \neq v_0$ then : (assume wlog that $u_0 < v_0$ in the sorted sequence of vertices/edges)

- merge e_{v_0} into e_{u_0} in $\mathcal{U}$
- print out "(" + $f(\sigma_i)$ + "," + $f(v_0)$ + "]"

(lifespan of v_0 as an independent comp)

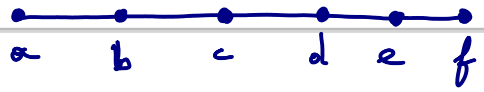
Post-processing :

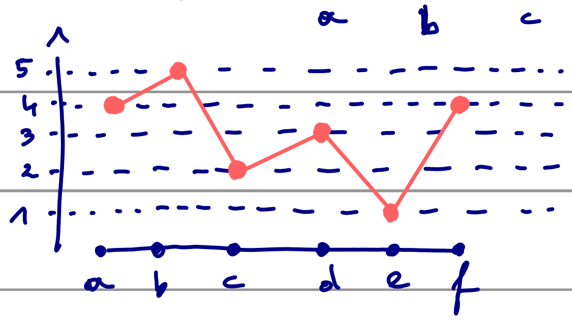
For each remaining entry e_v in $\mathcal{U}$ do :

- print out "$(-\infty$," + $f(v)$ + "]"

(component of v is never merged into the component of a higher peak $\Rightarrow$ persistence of v is infinite)

Example :

$G =$ 

$f =$ 

$f(u,v) := \min\{f(u), f(v)\}$

$\forall\, u,v \in \{a,b,c,d,e,f\}$

↳ lexicographic order : $b < a = f < ab < d < c < bc = cd < e < de = ef$

↳ barcode : $\{\ \underset{(a)}{\cancel{(4,4]}}\ ;\ \underset{(c)}{\cancel{(2,2]}}\ ;\ \underset{(d)}{(2,3]}\ ;\ \underset{(e)}{\cancel{(1,1]}}\ ;\ \underset{(f)}{(1,4]}\ ;\ \underset{(b)}{(-\infty,5]}\ \}$

— : remove empty intervals from the barcode