New Results on the Stability of Persistence Diagrams

Primoz Skraba
Jozef Stefan Institute

TAGS - Linking Topology to Algebraic Geometry and Statistics joint work with

Katharine Turner
and
D. Yogeshwaran
G. Thoppe

Stability

Setting: K - finite simplicial complex

$f, g : K \to \mathbb{R}$

Dgm(f) - sub-levelset PD ($f^{-1}(-\infty, \alpha]$)

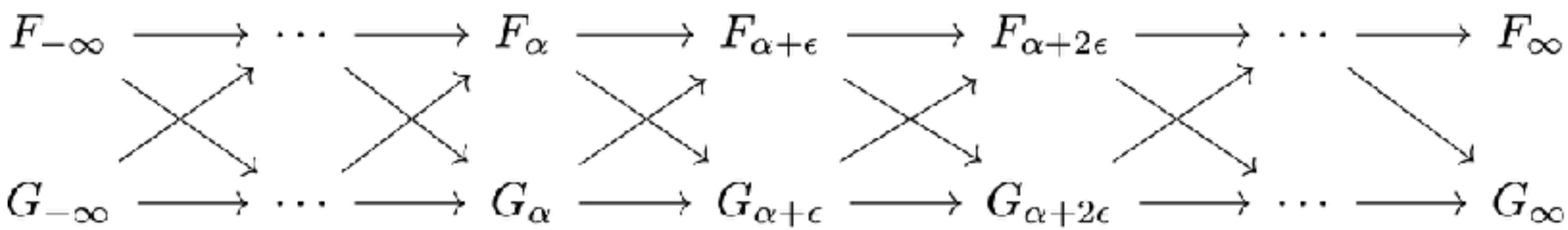
Stability Theorem [CSEH06]

Let X be a triangulable space with continuous tame functions $f, g : X \to \mathbb{R}$. Then the persistence diagrams Dgm(f) and Dgm(g) for their sublevel set filtrations satisfy

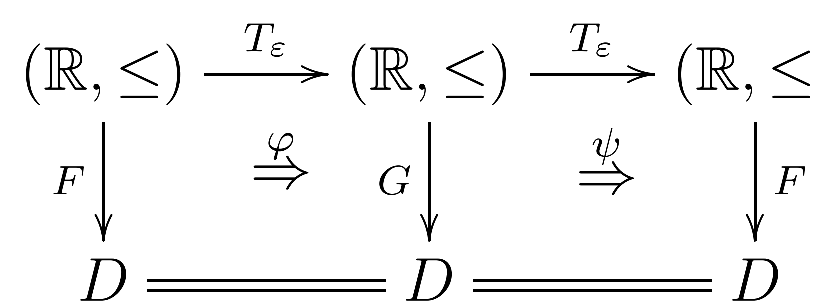
$$d_B(\text{Dgm}(f), \text{Dgm}(g))) \leq ||f - g||_\infty.$$

Extensions

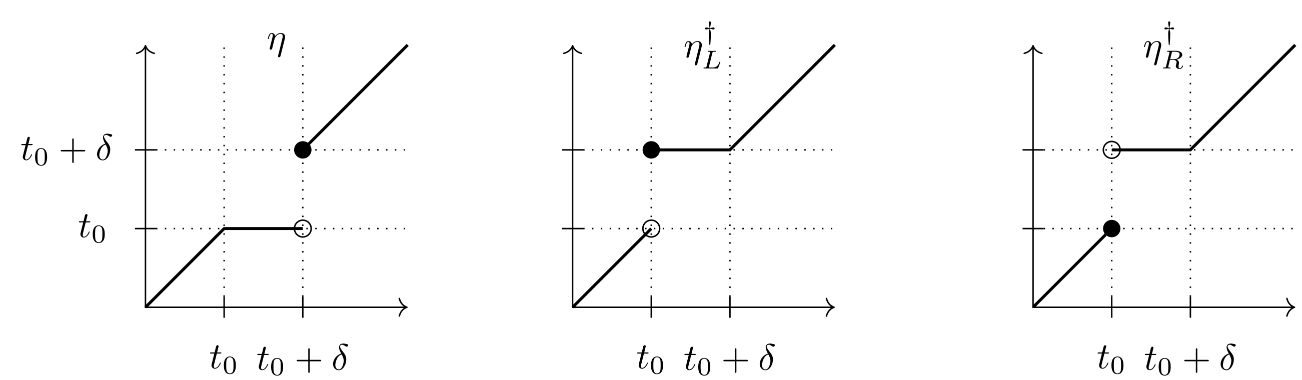
- Interleaving filtrations



- Categorical formulation



- Non-uniform interleaving



Applications of Stability

- Homological reconstruction
- Statistical inference

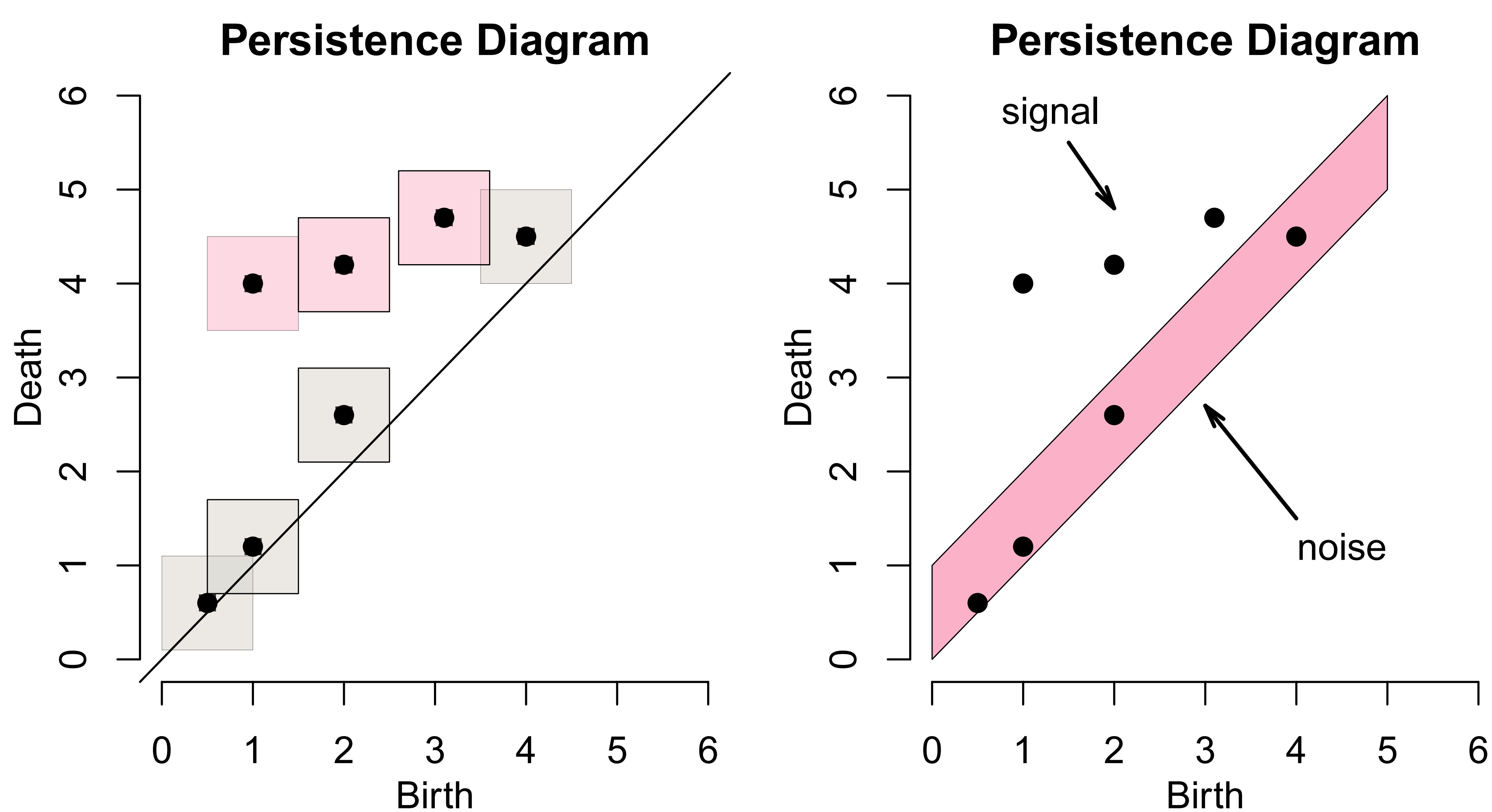
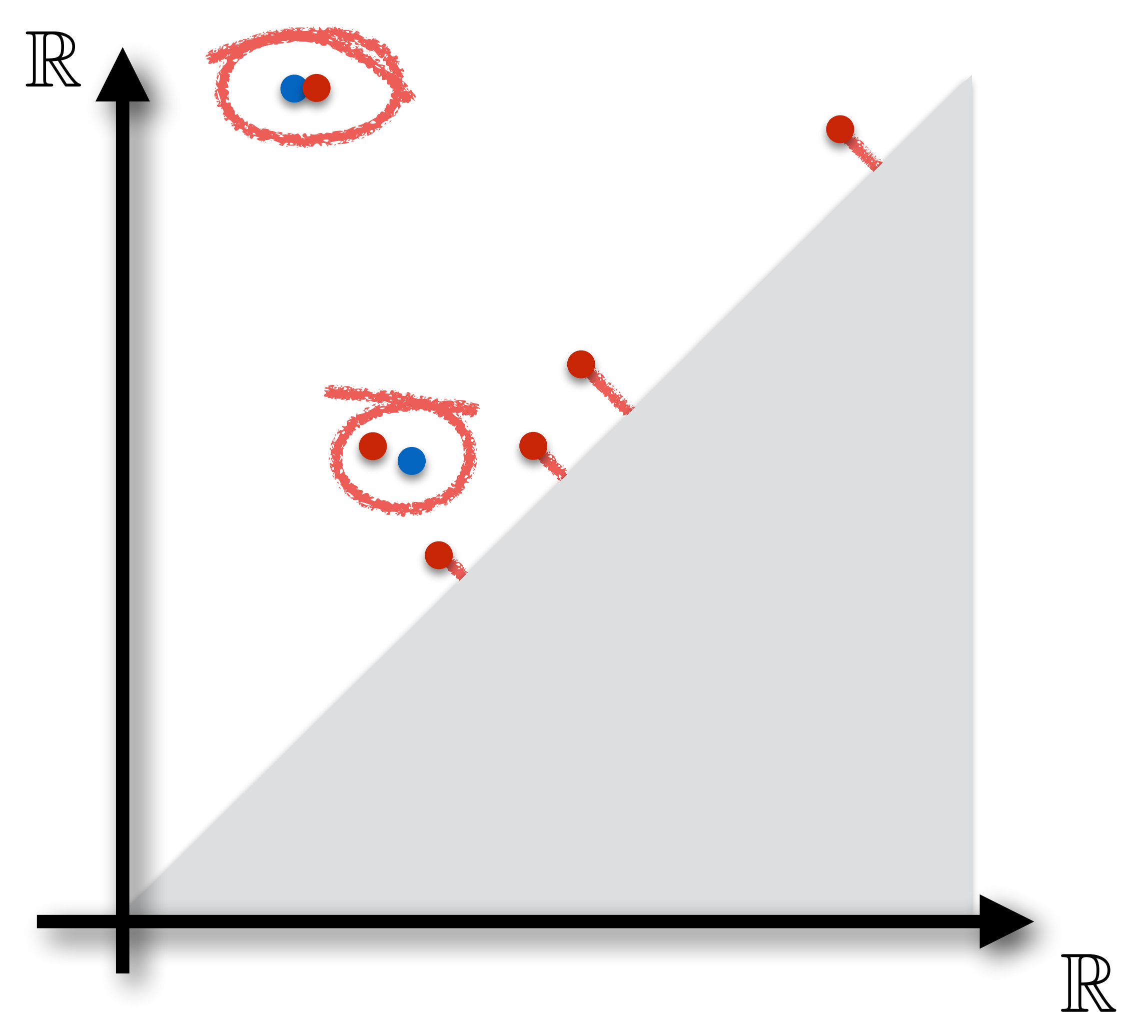


Image from Fasy, Brittany Terese, et al. "Confidence sets for persistence diagrams." The Annals of Statistics 42.6 (2014): 2301-2339.

Bottleneck Distance

Definition

$$d_\infty(X,Y) = \inf_{\substack{\phi: X \to Y \\ \phi \in \text{bijections}}} \sup_{x \in X} \|x - \phi(x)\|_\infty .$$



Problem: Outliers

- Bottleneck distance is defined by worst case
- To prove convergence, requires no outliers w.h.p.

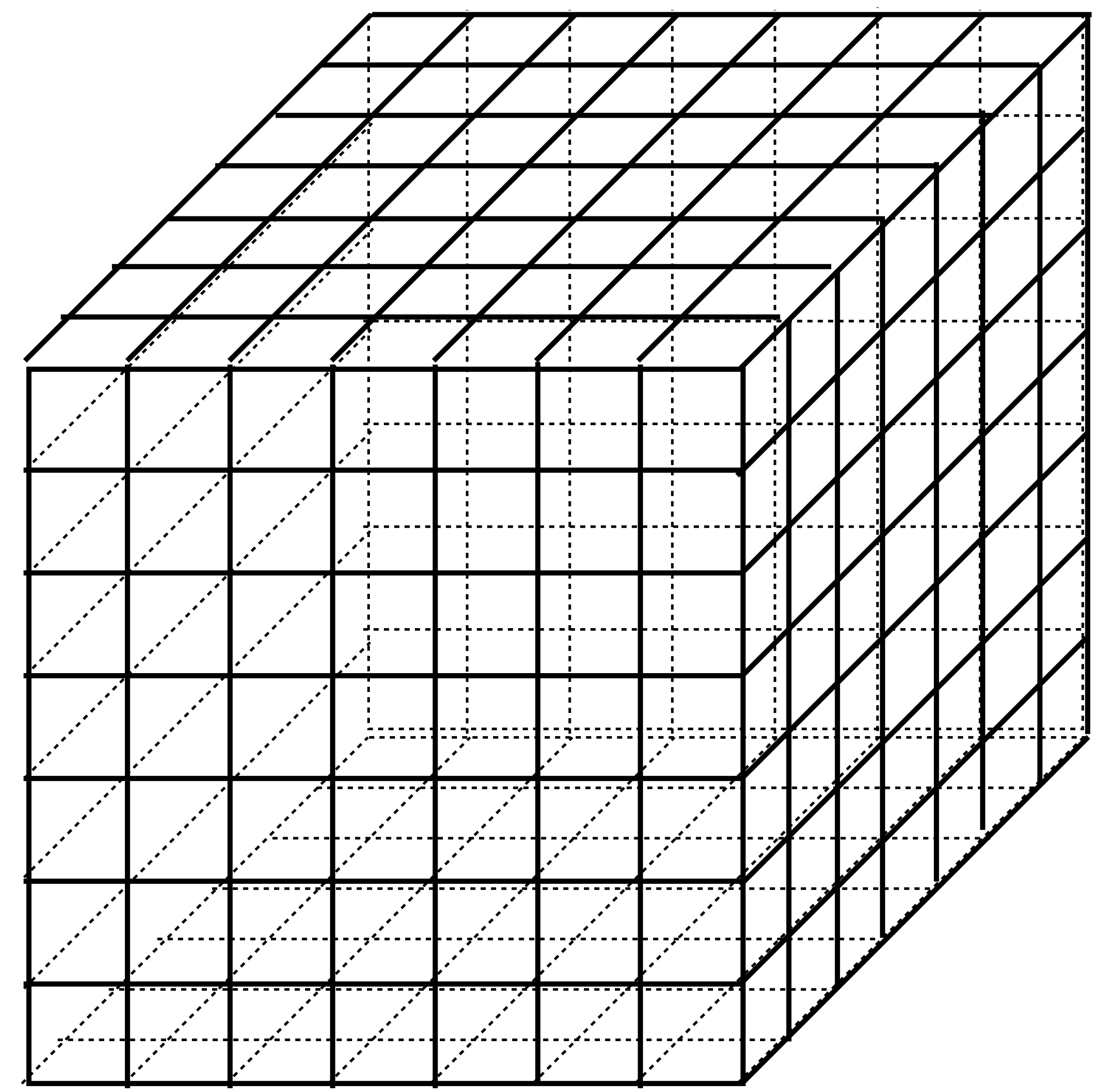
Example

Let $\mathcal{P}$ be drawn by sampling M in i.i.d. fashion according to the uniform probability measure on M. Then with probability greater than $1-\delta_\varepsilon$ we have that $\mathcal{P}$ is $\frac{\varepsilon}{2}$-dense $(\varepsilon < \frac{\tau}{2})$ in M provided

$$|\mathcal{P}| > \beta_1(\log(\beta_2) + \log(\frac{1}{\delta}))$$

Problem: Outliers

- Sub/super level set persistence
- Most errors are small, some are large
- Bottleneck distance is large



p-Wasserstein Distance

Definition

The p-Wasserstein distance between two PDs X and Y is defined as

$$d_p(X,Y) = \left(\inf_{\substack{\phi : X \to Y \\ \phi \in \text{bijections}}} \sum_{x \in X} \|x - \phi(x)\|_p^p \right)^{1/p}$$

Often used in applications

p-Wasserstein Distance

Definition

The p-Wasserstein distance between functions f and d defined on a simplicial complex K is defined as

$$\|f - g\|_p^p = \sum_{\Delta \in K} |f(\Delta) - g(\Delta)|^p.$$

Simplicial norm (compare to standard definition)

Related Work

Wasserstein stability for Lipschitz functions

Theorem

Let X be a triangulable space with continuous tame functions $f, g : X \to \mathbb{R}$. Then the persistence diagrams $\mathrm{Dgm}(f)$ and $\mathrm{Dgm}(g)$ for their sublevel set filtrations satisfy

$$W_p(\mathrm{Dgm}(f), \mathrm{Dgm}(g))) \leq C||f - g||_\infty^{1-\frac{k}{p}}.$$

C depends on total persistence

Uses equivalence of norms (number of points)

D. Cohen-Steiner, H. Edelsbrunner, J. Harer and Y. Mileyko, *Lipschitz Functions Have Lp-stable Persistence,* FOCM, April 2010

Goal

Theorem

Let X be a triangulable space with continuous tame functions $f, g : X \to \mathbb{R}$. Then the persistence diagrams $\text{Dgm}(f)$ and $\text{Dgm}(g)$ for their sublevel set filtrations satisfy

$$W_p(\text{Dgm}(f), \text{Dgm}(g))) \leq C||f - g||_p$$

Goal

Theorem

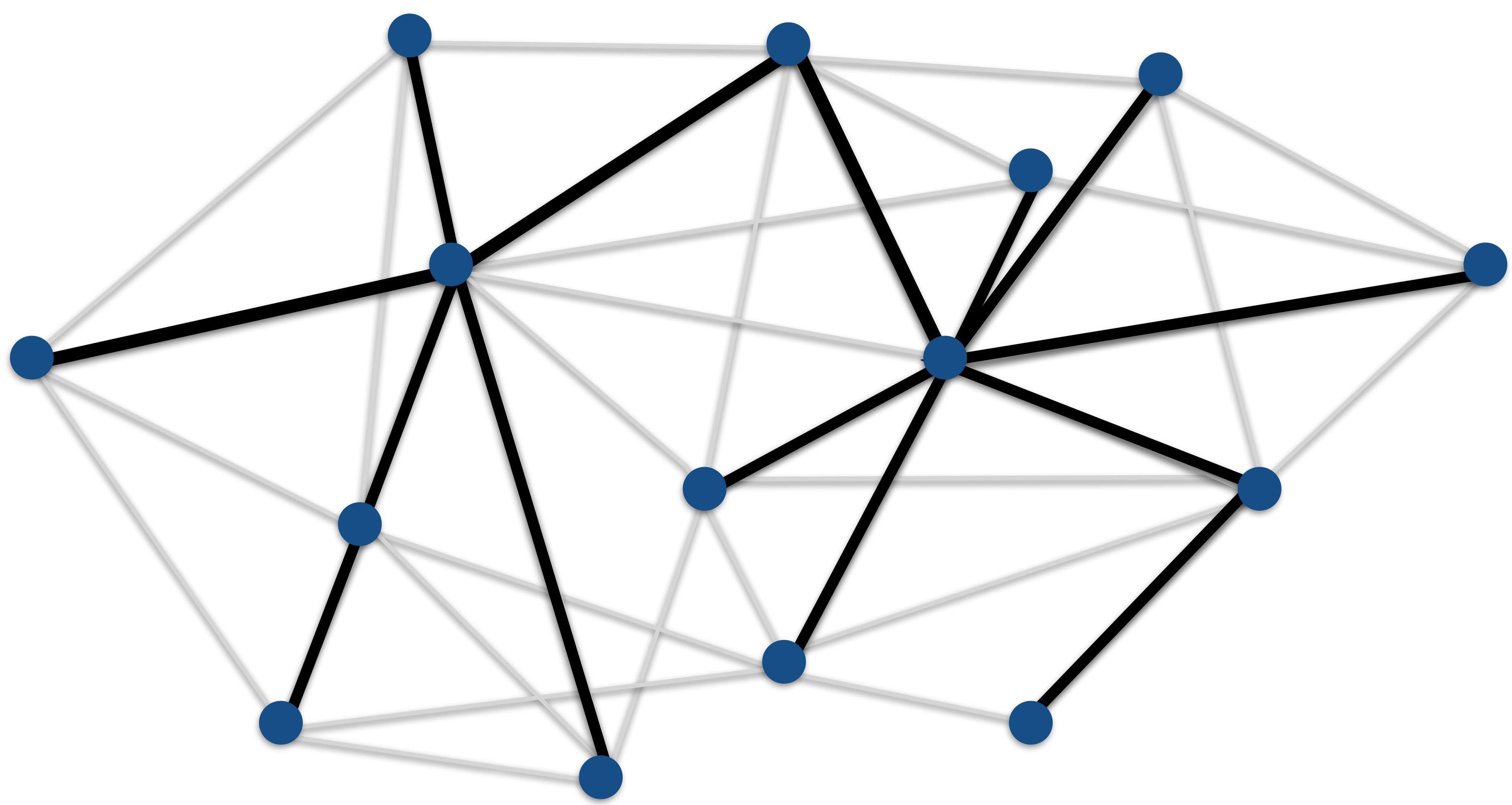
Let X be a triangulable space with continuous tame functions $f, g : X \to \mathbb{R}$. Then the persistence diagrams $\mathrm{Dgm}(f)$ and $\mathrm{Dgm}(g)$ for their sublevel set filtrations satisfy

$$W_p(\mathrm{Dgm}(f), \mathrm{Dgm}(g))) \leq C||f - g||_p$$

Why is this reasonable?

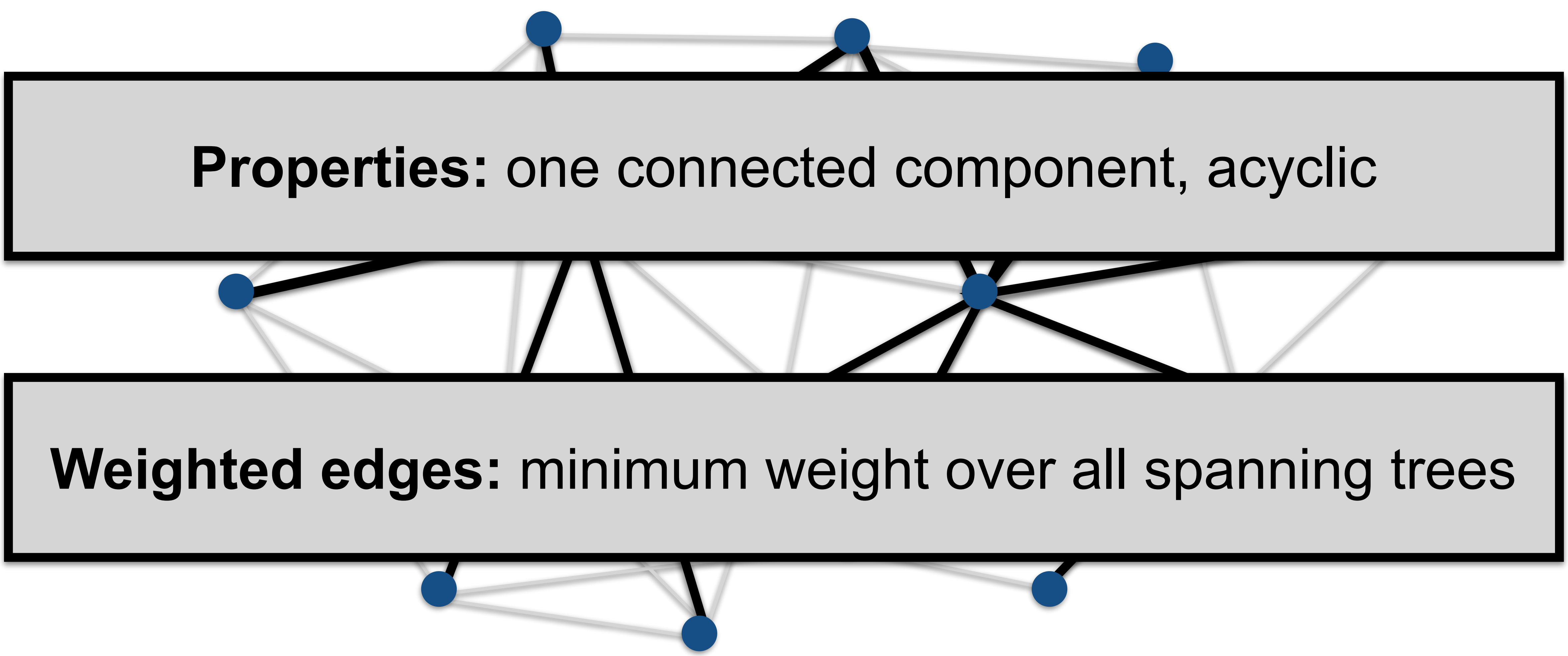
Minimum Spanning Acycles

- Generalization of a minimum spanning tree



Minimum Spanning Acycles

- Generalization of a minimum spanning tree



Minimum Spanning Acycles

Fix dimension d - consider d-skeleton of a simplicial complex

Algebraic properties:

Acyclicity: $\beta_d = 0$

Spanning: $\beta_{d-1} = 0$

Minimum Spanning Acycles

Fix dimension d - consider d-skeleton of a simplicial complex

Algebraic properties:

Acyclicity: $\beta_d = 0$

Spanning: $\beta_{d-1} = 0$

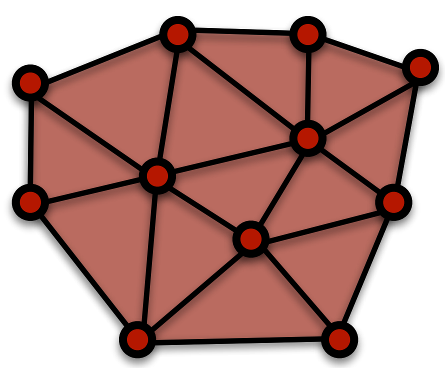
If $(d\text{-}1)$-Betti number cannot be zero, higher dimensional generalisation of spanning forest.

Minimum Spanning Acycles

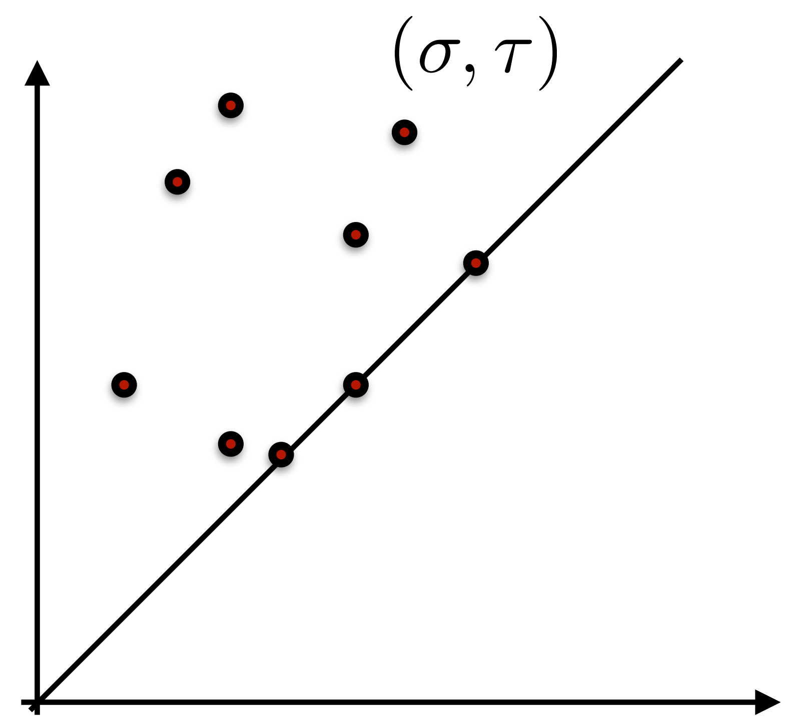
Weighted version: d-simplices have weights

Use weight function as a filtration

$f :$



$\to \mathbb{R}$



Observation

Negative simplicies (& death times) correspond to simplicies in MSA

Birth and Death Times

Theorem

Let K be a finite complex with two monotone functions f, g. For any $p \in \{0, \ldots, \infty\}$,

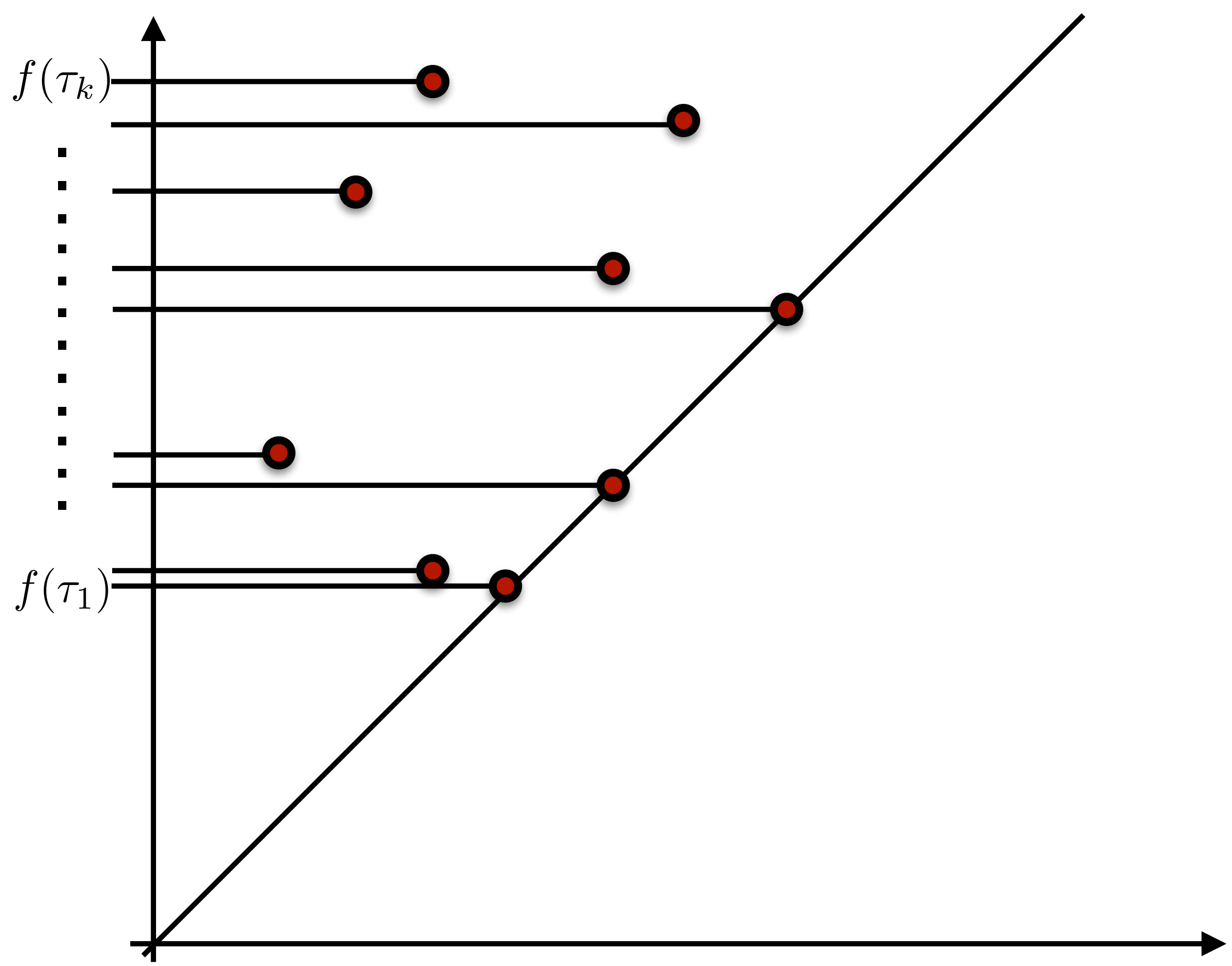
$$\inf_{\pi \in \Pi_D} \sum_i |D_i - \pi(D_i')|^p \leq \sum_{\sigma \in K} |f(\sigma) - g(\sigma)|^p$$

$$\inf_{\pi \in \Pi_B} \sum_i |B_i - \pi(B_i')|^p \leq \sum_{\sigma \in K} |f(\sigma) - g(\sigma)|^p$$

Outline of Proof

Death times represent when no new classes are created (including instantaneous classes)

Each simplex + or - $\Rightarrow$ consequence of Mayer-Vietoris



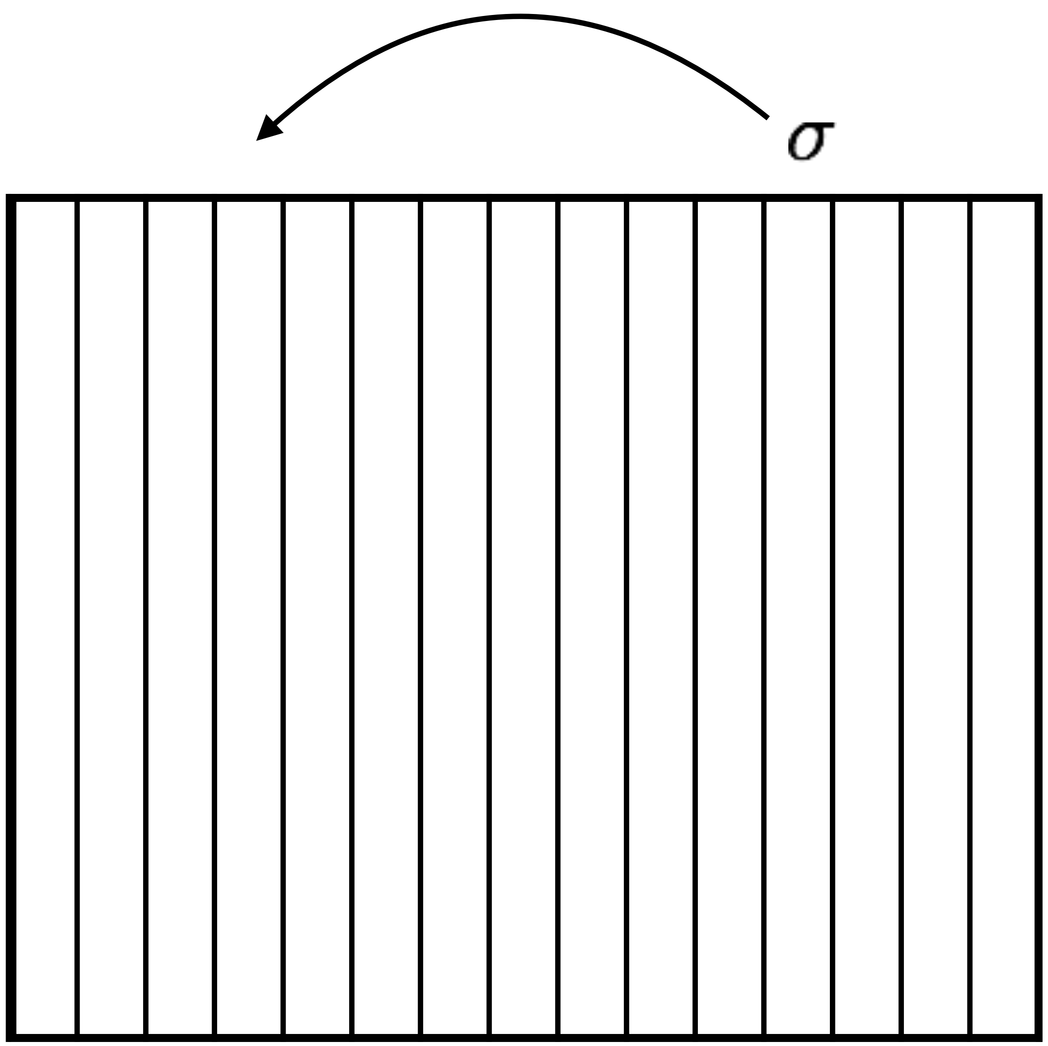
Outline of Proof

4 cases:

1. Moving a positive simplex forward
2. Moving a negative simplex backward
3. Moving a positive simplex backward
4. Moving a negative simplex forward

Outline of Proof

Moving a positive simplex backward



- If remains positive - trivial
- If negative - birth time moves less than how much the simplex moved

Towards Persistence Diagrams

In persistence, need to take care of pairing between positive and negative simplicies

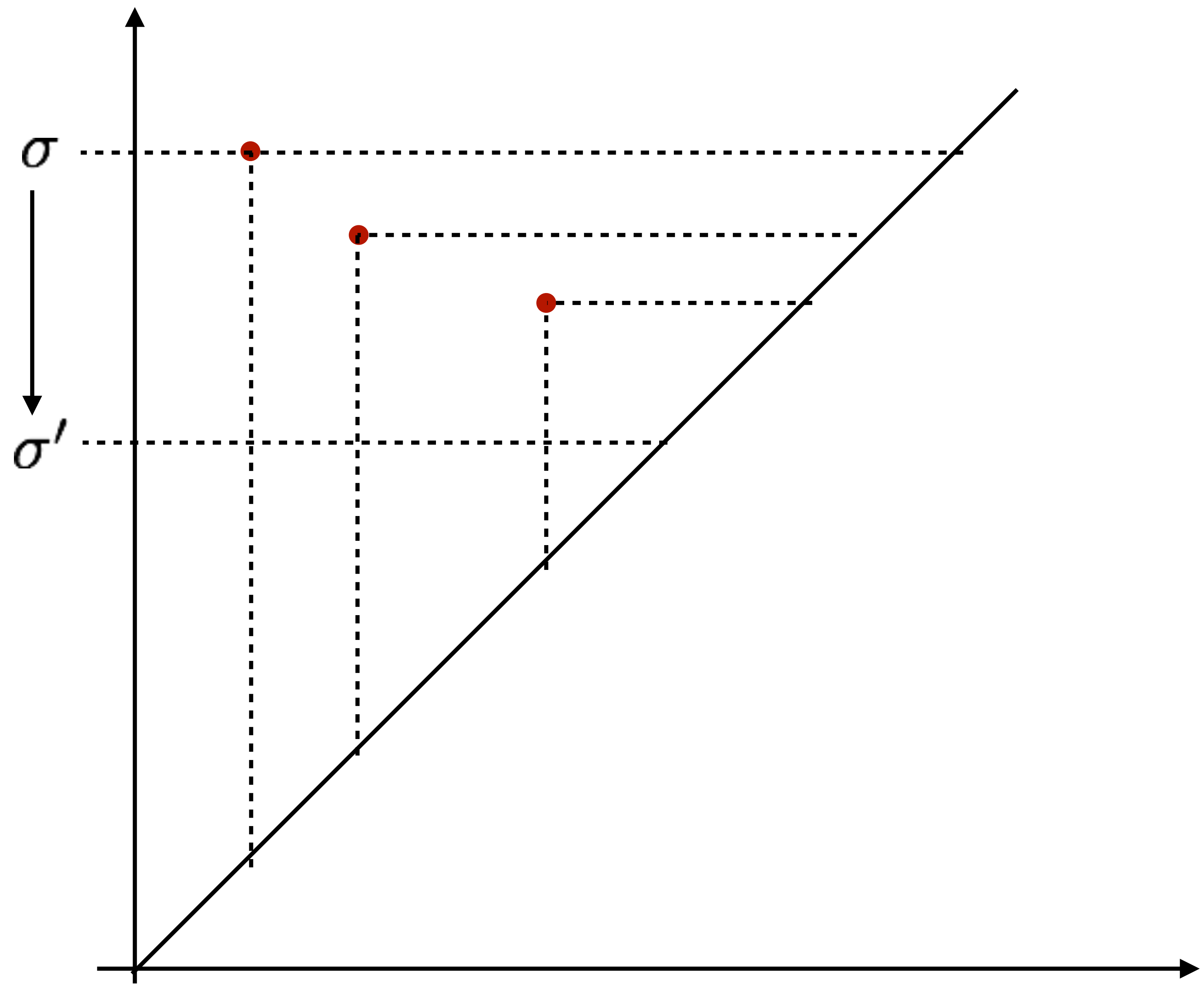
Lemma

Let $f : K \to \mathbb{R}$ be a monotone function over a simplicial complex K. There exists a map from Dgm(f) to pairs of simplices (σ, τ) such that for $p \in$ Dgm(f), if $\pi(p) = (\sigma, \tau)$, then $f(\sigma) = p_x$ and $f(\tau) = p_y$.

How can this map change?

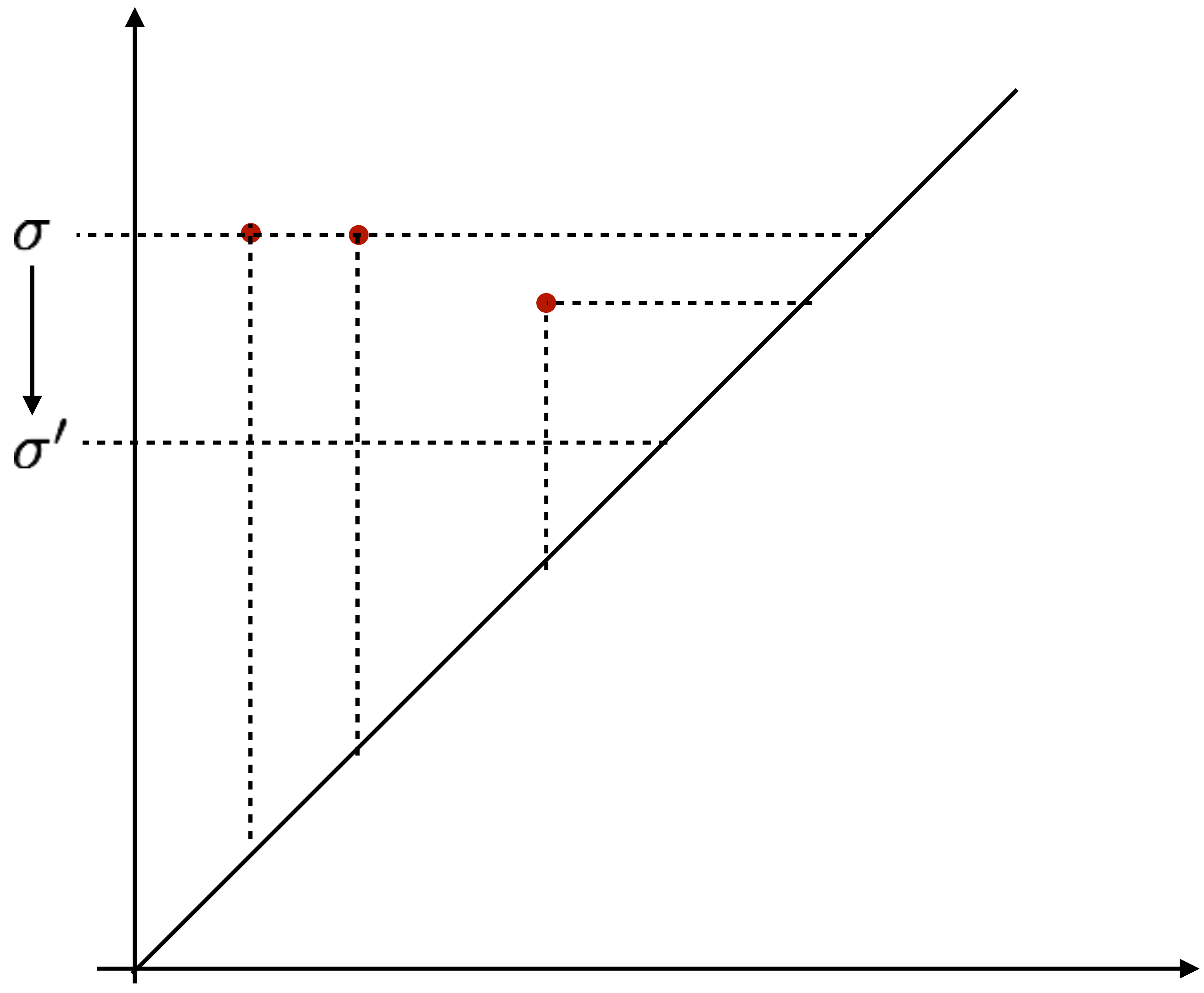
Geometric Picture

Example: moving a negative simplex back in the filtration



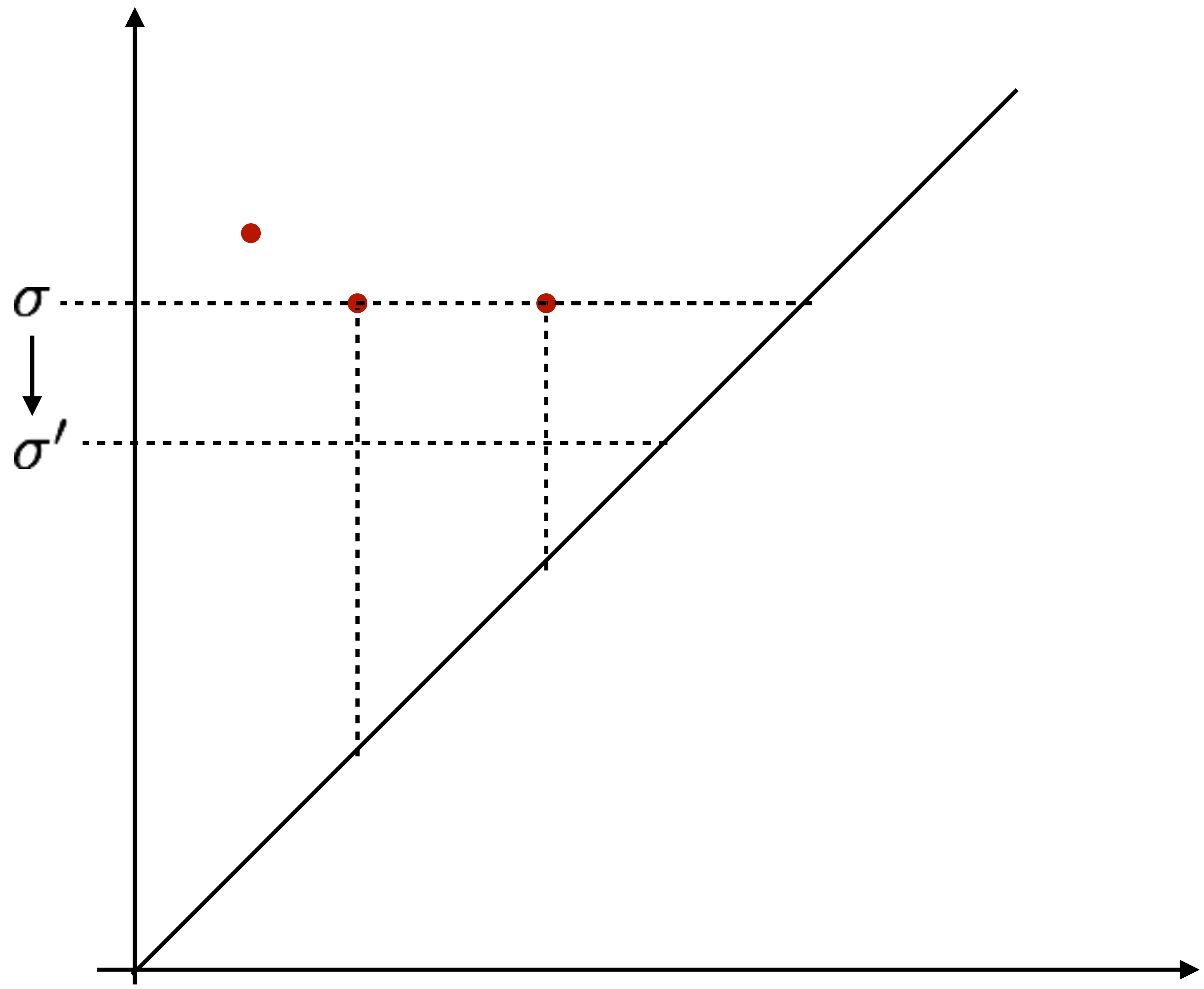
Geometric Picture

Example: moving a negative simplex back in the filtration



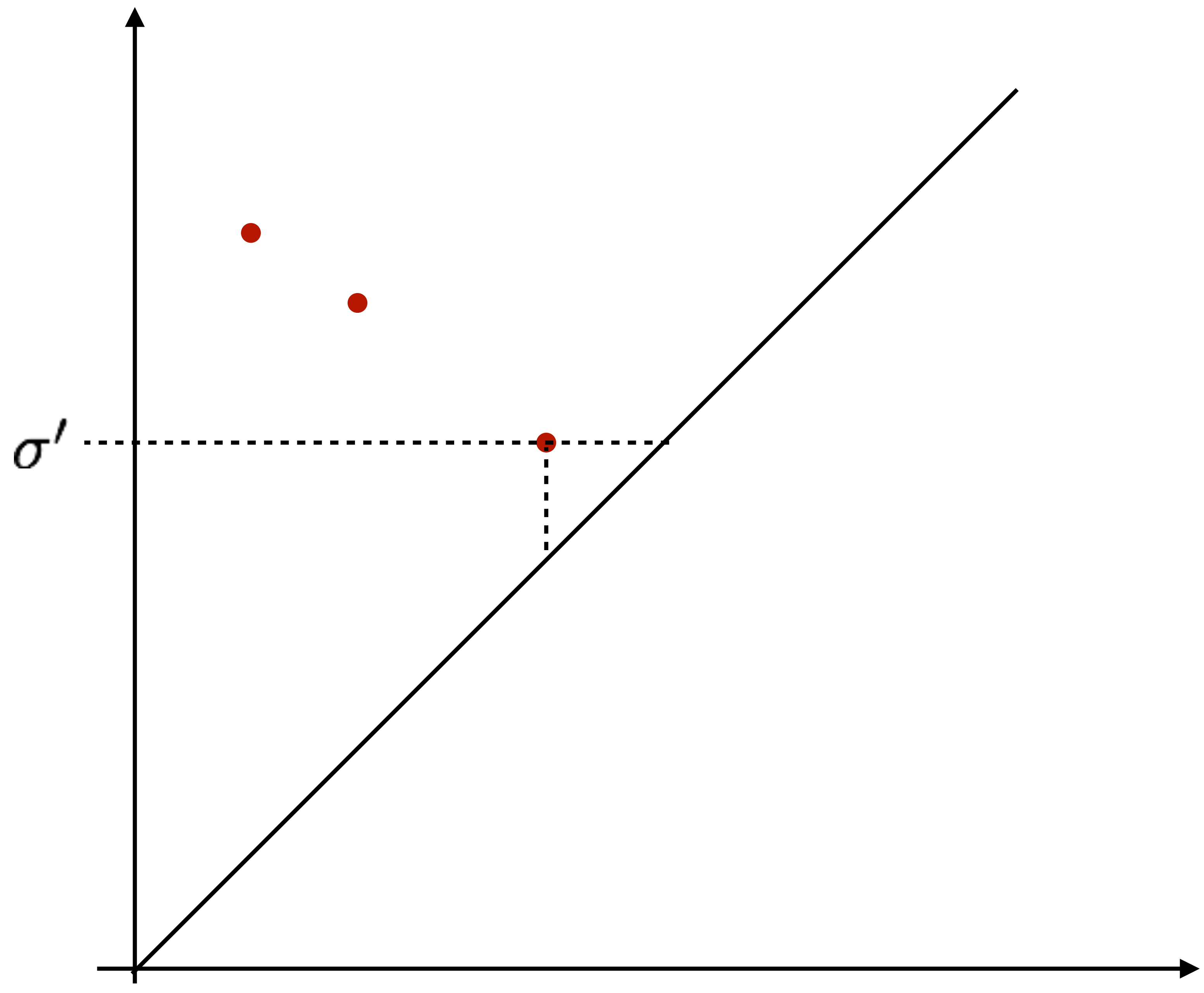
Geometric Picture

<u>Example</u>: moving a negative simplex back in the filtration



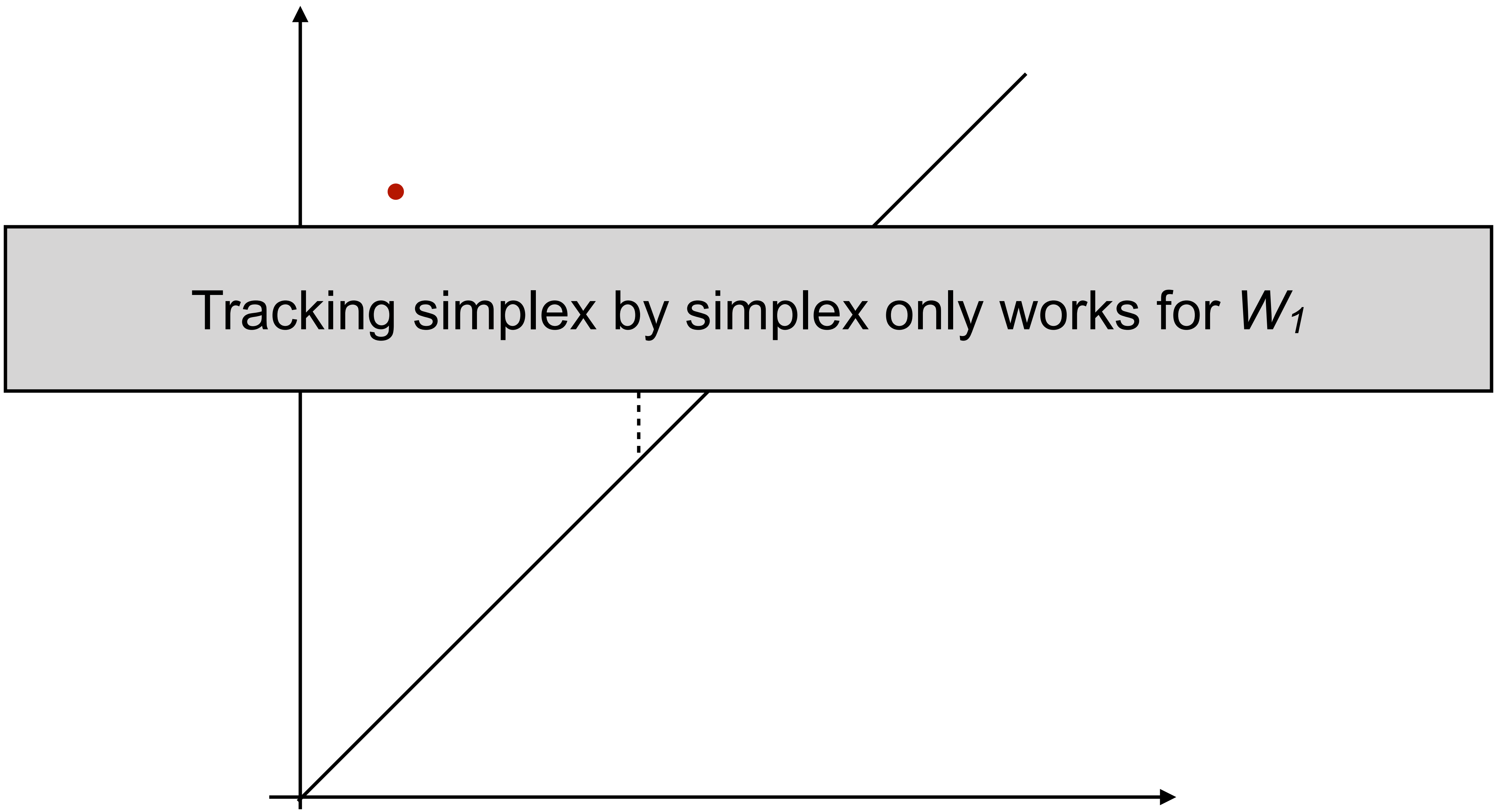
Geometric Picture

Example: moving a negative simplex back in the filtration



Geometric Picture

Example: moving a negative simplex back in the filtration



Critical Pairs

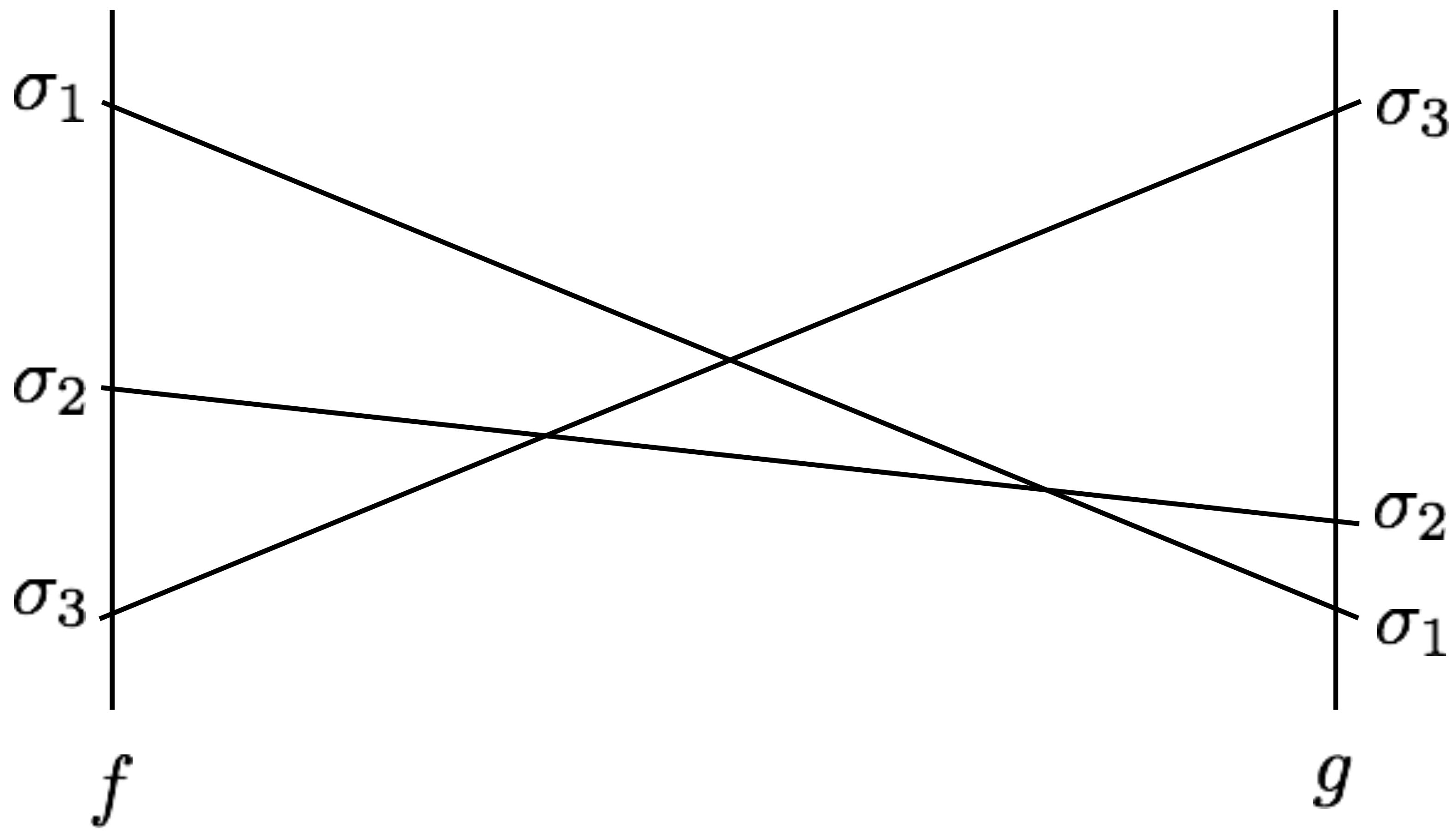
Definition

A *critical pair* is a pair of simplicies in the image of π such that $f(\tau) - f(\sigma) > 0$. We call any simplex crtitical if it part of a critical pair.

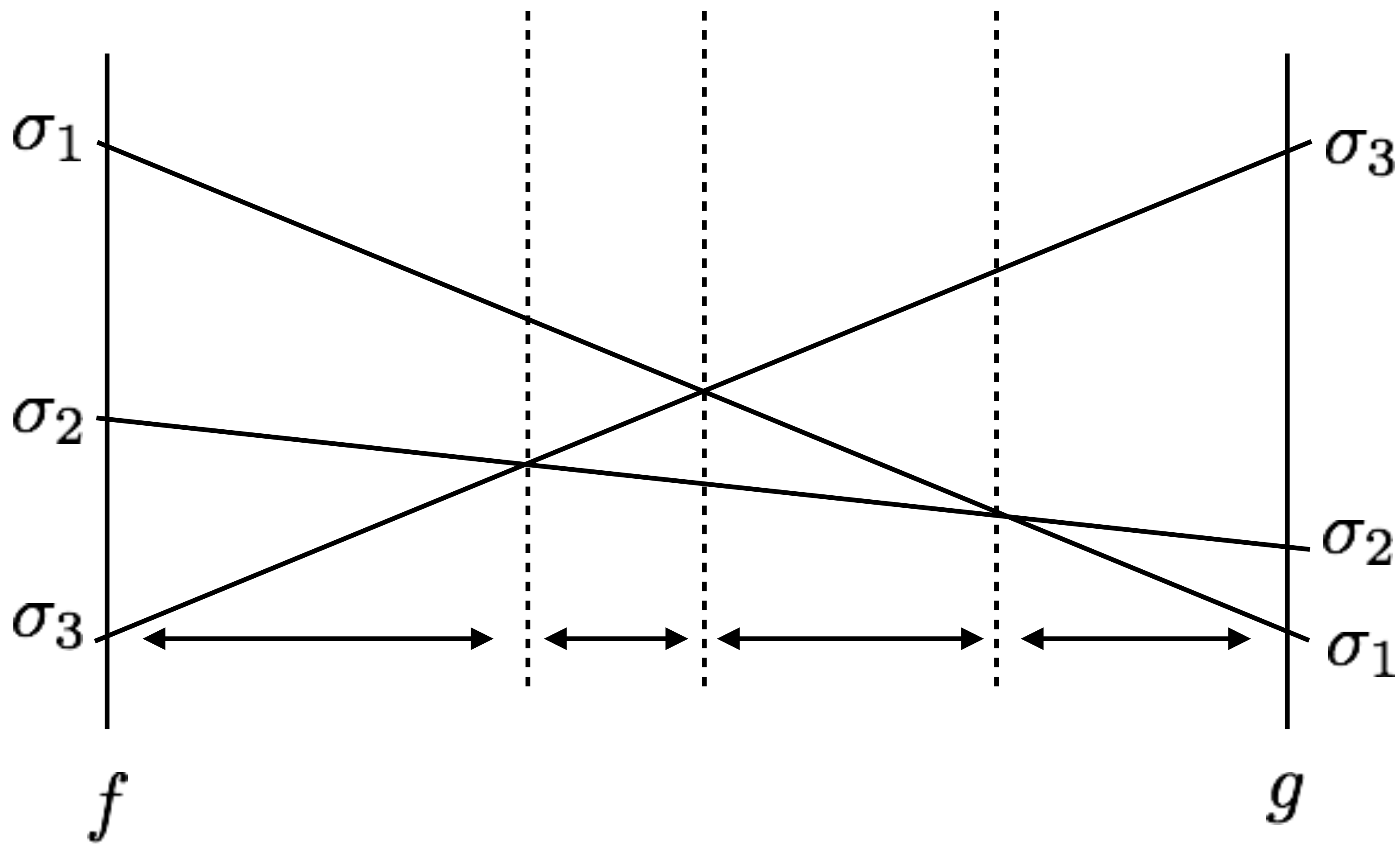
Correspondence with points off the diagonal

Interpolation

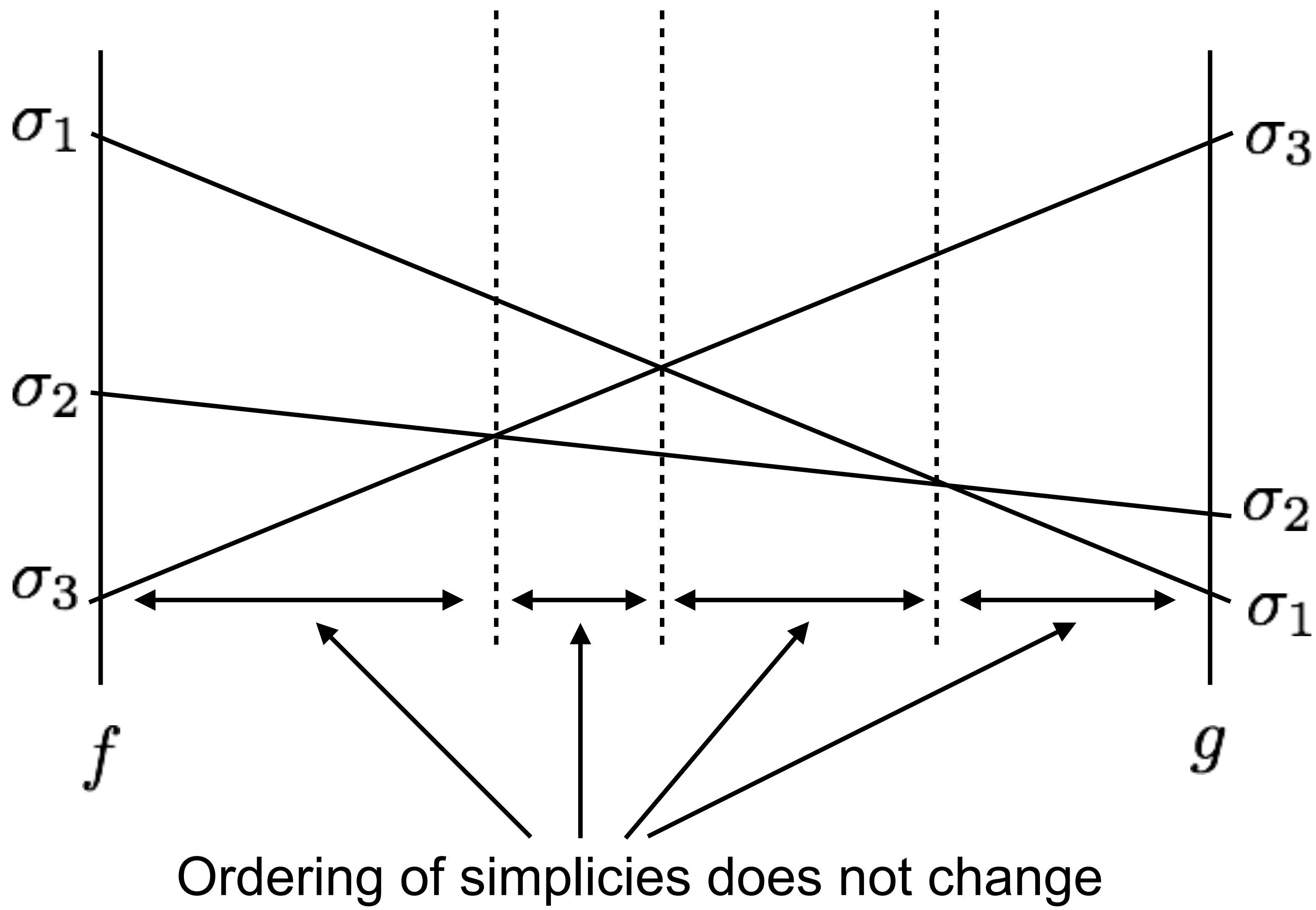
Linearly interpolate between functions



Dividing Up the Problem



Dividing Up the Problem



Easy Case

Lemma

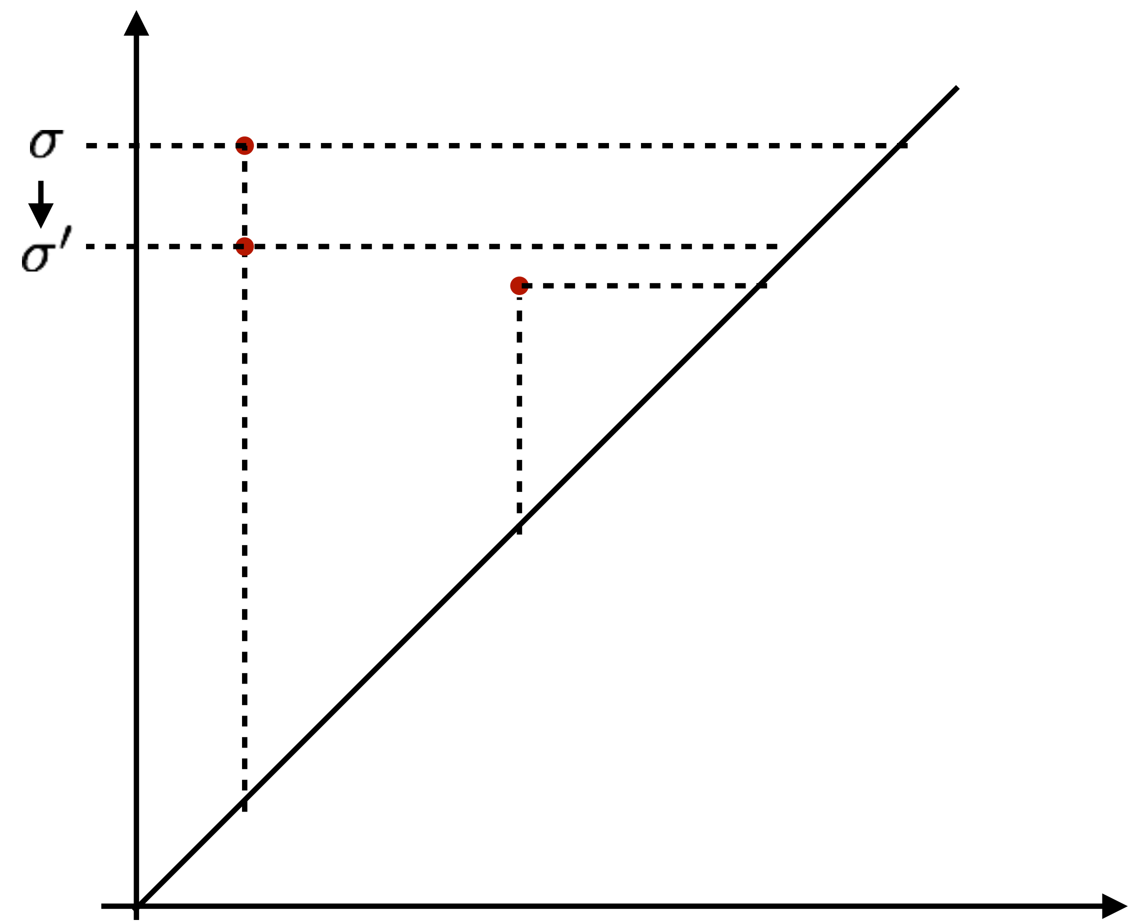
Let $f_t : K \to \mathbb{R}$, $t \in [a, b]$ be a continuous family of monotone functions over a simplicial complex K such that for all $a < s < b$ the order (potentially with equality) of the function values of the simplices remains the same. Then

$$W_p(\text{Dgm}(f_a), \text{Dgm}(f_b)) \leq ||f_a - f_b||_p.$$

Proof Idea

For $t \in (a, b)$, the ordering does not change so pairing map does not change (we extend to a and b by taking one-sided limits)

Since pairing map does not change, movement of points is equal to movement of simplices



Combining Intervals

Within each interval we have the desired bound

For each interval (a, b), $0 \leq a \leq b \leq 1$ we have

$$||f_a - f_b||_p = |b - a| \cdot ||f - g||_p$$

$$\begin{aligned} W_p(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) &\leq \sum_{i=0}^{n} W_p(\mathrm{Dgm}(f_{a_i}), \mathrm{Dgm}(f_{a_{i+1}})) \\ &\leq \sum_{i=0}^{n} \|f_{a_i} - f_{a_{i+1}}\|_p \\ &= \sum_{i=0}^{n} (a_{i+1} - a_i)\|f - g\|_p \\ &= \|f - g\|_p \end{aligned}$$

Main Result

Theorem

Let $f, g : K \to \mathbb{R}$ be monotone functions,

$$W_p(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leq \|f - g\|_p.$$

Observations

- Outliers do not affect Wasserstein distances between persistence diagrams too much
- Requires fixed simplicial complex (since we use a simplicial norm)
- It does not use the equivalence of norms

Vietoris-Rips Filtrations

- Often we build complexes from point sets
- Can a similar result hold if we move/perturb points?

Result

Theorem

Fix k, d If $C_{d,k}$ is finite then for all $p \geq 1$, assuming $X_0, X_1 \subset \mathbb{R}^d$

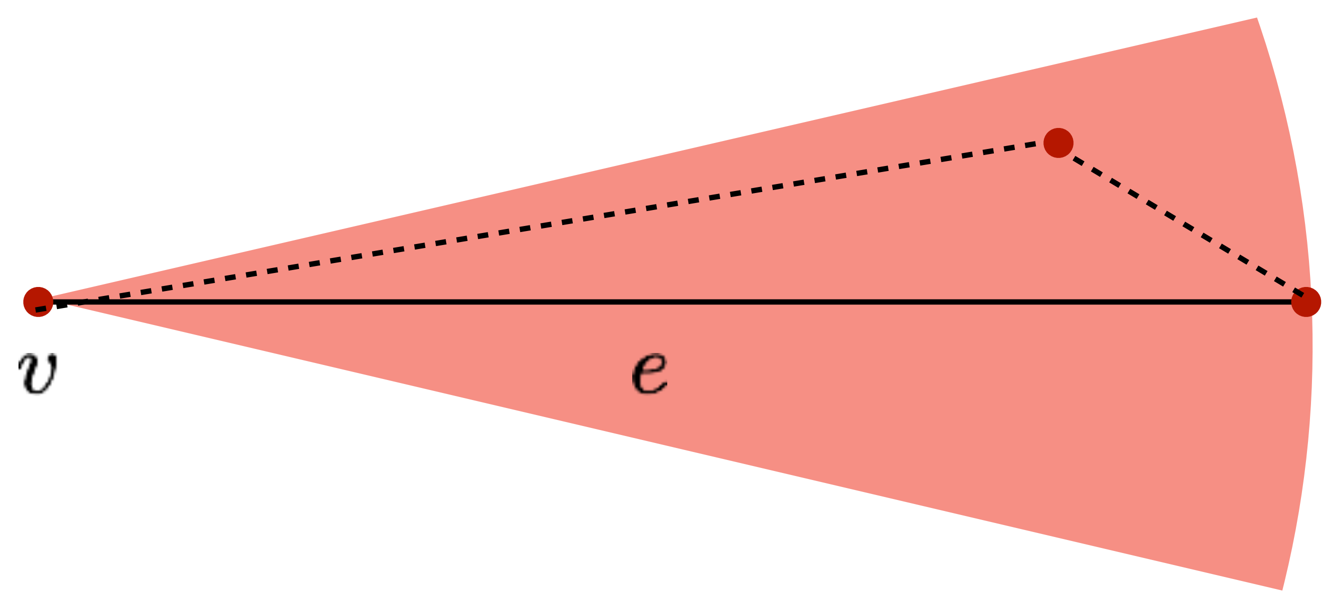
$$W_p(\mathrm{Dgm}_k(X_0), \mathrm{Dgm}_k(X_1) \leq C_{d,k}^{1/p} D_p(S_0, S_1).$$

where $\mathrm{Dgm}_k(X_i)$ is the k-dimensional persistence diagram for the Vietoris-Rips filtration on the point set X_i.

When can we bound $C_{k,d}$?

Components(H_0)

- Cannot have too many classes around one vertex
- We do not points, so only consider critical edges (which for 0-dim only kill homology classes)
- Assuming *e* is critical (adjacent to a critical simplex), we can exclude a cone



Čech Filtrations

Distance filtration:

Let $\mathcal{P} \subset \mathbb{R}^d$ be a finite point set, let $f : \mathbb{R}^d \to \mathbb{R}$, defined by

$$f(x) = \min_{p \in \mathcal{P}} d(x, p)$$

Homotopic to the Čech Filtration

Counterexample

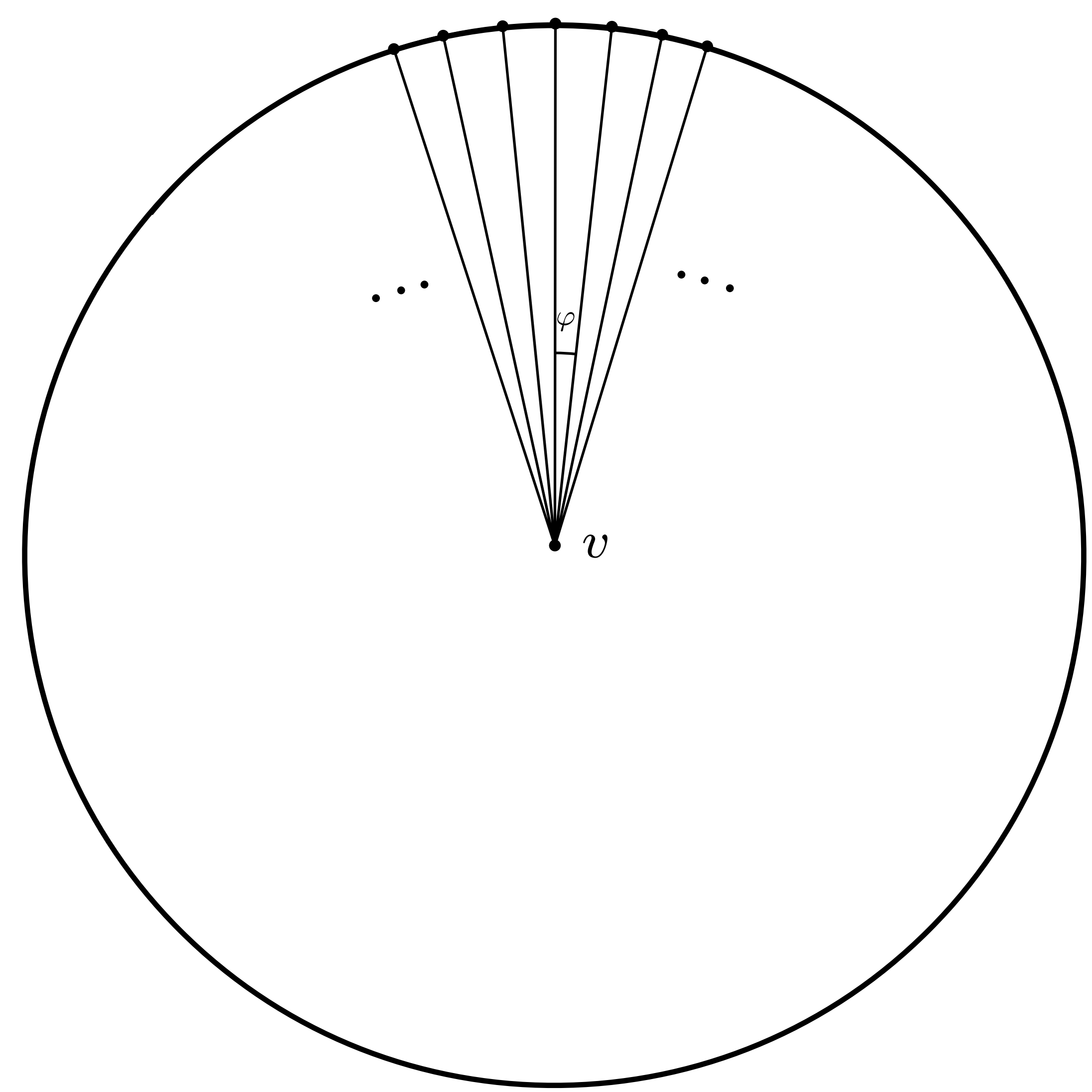
Points placed along a circle

Configuration is generic

Constant depends on n,

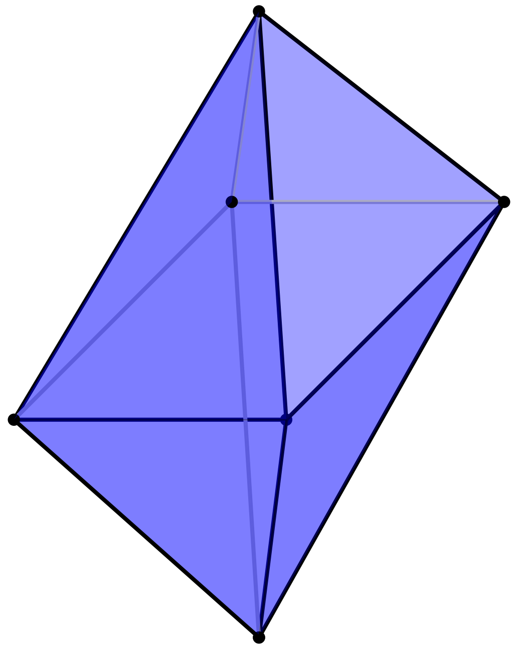
$$C_{2,2} = O(n)$$

$$C_{2,1} = O(n)$$

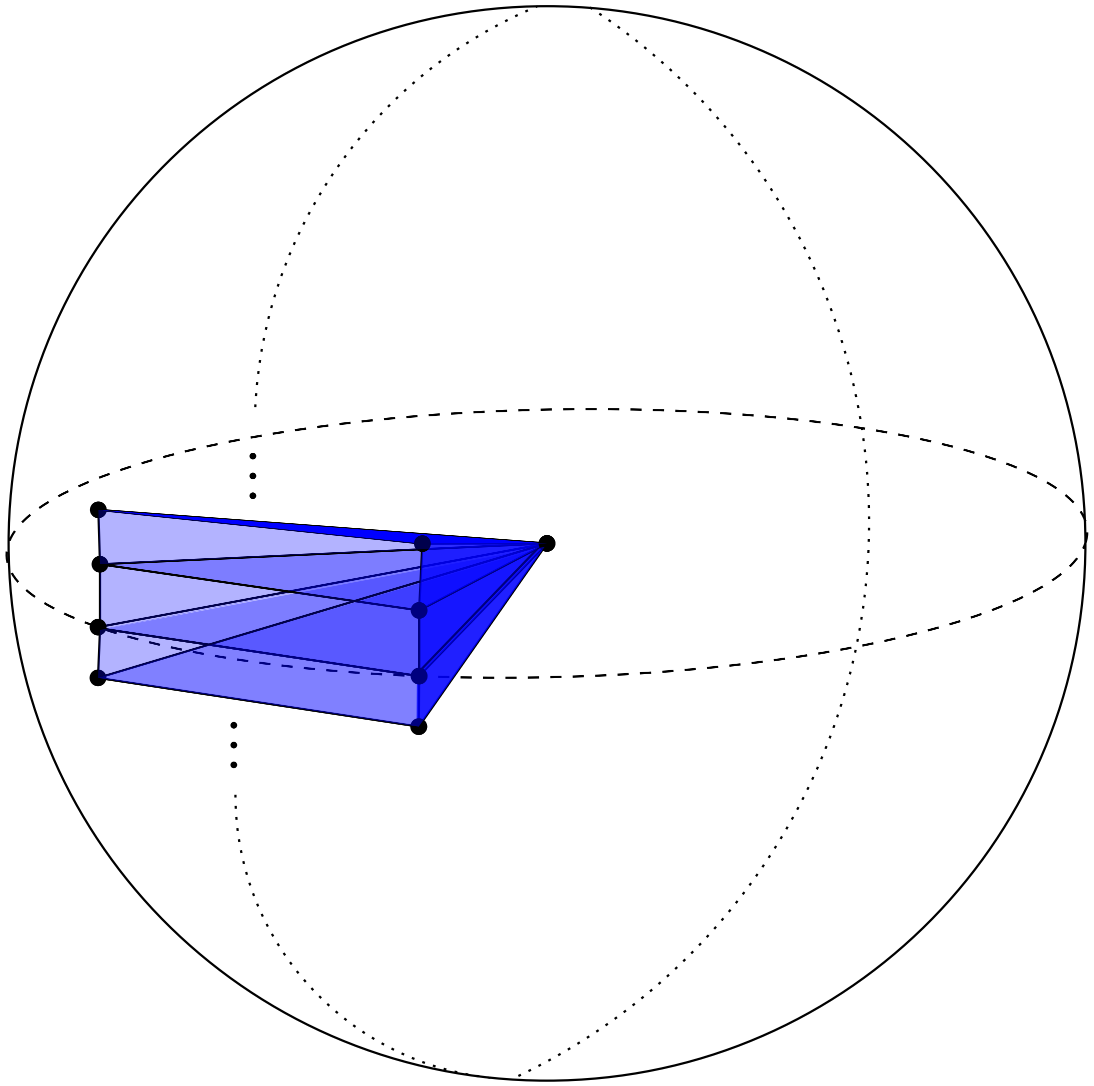


Bounding $C_{d,k}$

- Counterexample



minimal nontrivial 2-cycle



packing critical simplifies

Bounding $C_{2,k}$

Generalize 0-dimensional case: cannot be adjacent to too many critical simplifies

Lemma

If $\tilde{H}_k(\text{Lnk}(e)) = 0$ for all k, then e cannot be a critical edge.

Proof: Mayer-Vietoris

Bounding $C_{2,1}$

Corollary

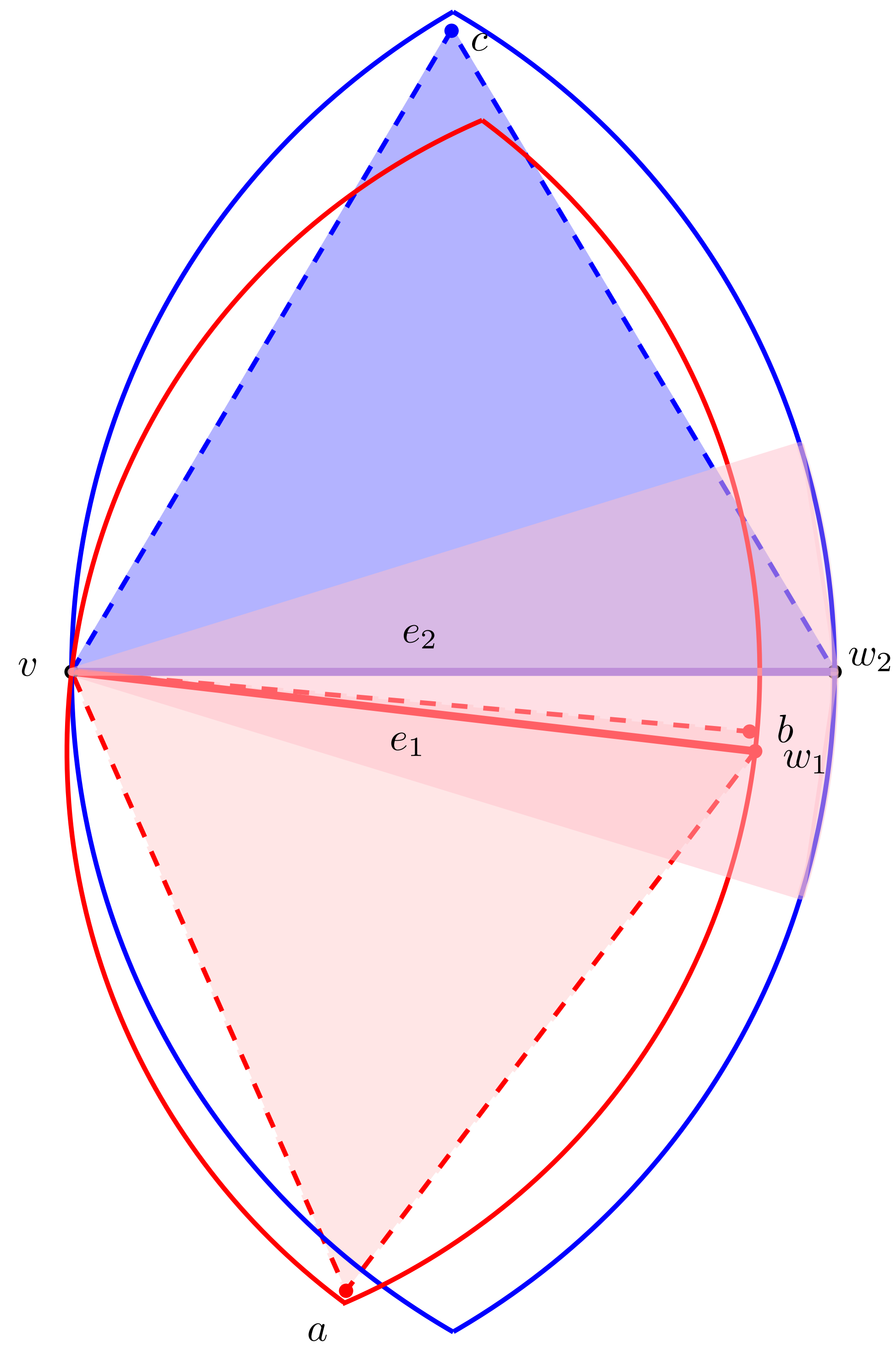
If e an edge is critical, either Lnk(e) consists only of the endpoints of e or Lnk(e) (and so Cl St(e)) must contain at least two vertices, v_1 and v_2 such that $v_1, v_2 \notin e$ and $d(v_1, v_2) > f(e)$. In $\mathbb{R}^2$, this implies that v_1, lies in the half-plane above e and v_2 in the half-plane below e.

Implication: There are at most three 1-critical edges within an angle of $\pi/12$ of each other.

Geometric Picture

If there are four 1-critical edges, then any triangle which includes the 4th (longest) edge, is homologous to an existing triangle.

Should extend to $C_{2,k}$



What's Next?

- Relating simplicial norm to more classical norm, e.g. Sobelov norm, recent work by Polterovich et. al. (Persistence barcodes and Laplace eigenfunctions)

- Expected Wasserstein bounds (bad cases are generic but unlikely)

- Combining with approximate Nerve Theorem