Multidimensional discrete Morse function for persistent homology computation

Tomasz Kaczynski

joint work with
Madjid Allili, Claudia Landi, and Filippo Masoni

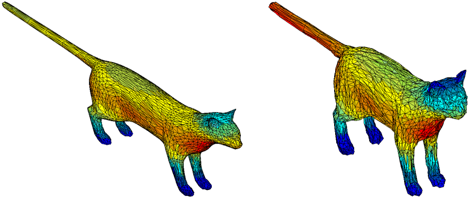
Départment de mathématiques
Université de Sherbrooke

Multiparameter Persistent Homology
BIRS–Oaxaca 2018

Outline

Shape similarity measures

How big is $\operatorname{dist}((X, f), (Y, g))$ between two shapes?



Shape descriptors based on *1D filtrations:* Filter images X by *sublevel sets* $X_{f \leq a}$ of a *measuring function* $f : X \to \mathbb{R}$.
Record changes in topology as a increases.

- Ideally, f should express features of interest
  — provided by users.
- Typical choices of f for testing purposes:
  - Coordinate projections;
  - Distance to a half-space, the gravity center, an axis of inertia, . . .,

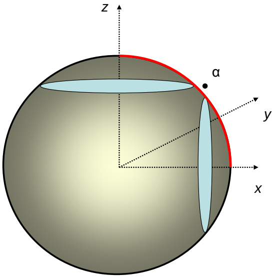
Multiparameter filtration: Study several features of compared shapes at once. Filter models X by *partially ordered sublevel sets* $X_{f \preceq a}$ of a *measuring function* $f : X \to \mathbb{R}^k$.

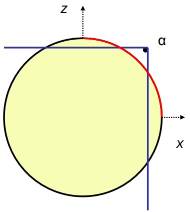
Record changes in topology induced by inclusions

$$j^{(a,b)} : X_a \hookrightarrow X_b,$$

where $a \preceq b$, i.e. $a_i \leq b_i$ for all $i = 1, 2, \ldots, k$.

History: *Pareto optimal points* in Economy, $\sim$ 1900's
Problem: Simultaneously maximize several functions.

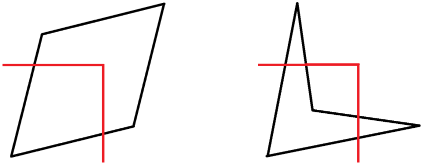




Why f with values in $\mathbb{R}^k$ and not k separate tests for 1D functions?

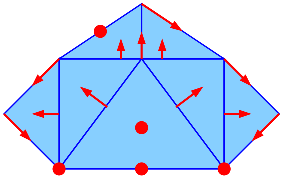
Example

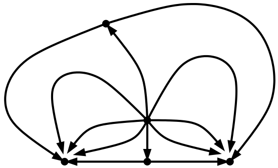
Using one coordinate projection per time does not permit distinguishing these two contours:



- Benefits from *MultiD* descriptors: More accurate shape similarity measures.
- Issues: More costly to compute. Distance (*matching, Wasserstein*, etc) computation in development — at this workshop!
- Explored direction: Reduce the complex representing the shape. Reduction should be filtration-preserving.

Morse-Forman Theory





Our goals:

- Extend Forman's concept of *discrete Morse function* to $\mathbb{R}^k$*-valued* functions.
- Construct a multifiltration-compatible *discrete vector field*.
- Compute a reduced complex with the same persistent homology.
- Use the above as a *pre-processing* for the distance computation to come.

MD Morse function

$\mathcal{K} = \{\mathcal{K}_p\}$ simplicial complex. Given $g : \mathcal{K} \to \mathbb{R}^k$ and $\alpha \in \mathcal{K}_p$, we set

$$H_g(\alpha) = \{\beta \in \mathcal{K}_{p+1} \mid \beta > \alpha \text{ and } g(\beta) \preceq g(\alpha)\};$$

$$T_g(\alpha) = \{\gamma \in \mathcal{K}_{p-1} \mid \gamma < \alpha \text{ and } g(\alpha) \preceq g(\gamma)\}.$$

H stands for *heads* and T for *tails*.

Definition

$g : \mathcal{K} \to \mathbb{R}^k$ is a multidimensional discrete Morse (mdm) function, *if*

(1) $\operatorname{card} H_g(\alpha) \leq 1$;

(2) $\operatorname{card} T_g(\alpha) \leq 1$;

(3) *If $\beta^{(p+1)} > \alpha$ is not in $H_g(\alpha)$, then $g(\alpha) \precneqq g(\beta)$;*

(4) *If $\gamma^{(p-1)} < \alpha$ is not in $T_g(\alpha)$, then $g(\gamma) \precneqq g(\alpha)$.*

Proposition

For any simplex $\alpha \in \mathcal{K}$, $\operatorname{card} H_g(\alpha) \cdot \operatorname{card} T_g(\alpha) = 0$.

Recall: A discrete vector field (dvf) V on $\mathcal{K}$ is the set of pairs

$$\left\{\left(\alpha^{(p)}, \beta^{(p+1)}\right)\right\} \text{ with } \alpha^{(p)} < \beta^{(p+1)}$$

such that each simplex of $\mathcal{K}$ is in at most one pair of V.

Definition

Let $g : \mathcal{K} \to \mathbb{R}^k$ be mdm. $\gamma \in \mathcal{K}$ is critical if $H_g(\gamma) = \emptyset = T_g(\gamma)$.

The sets

$$\mathtt{A} = \{\alpha \in \mathcal{K} \mid \operatorname{card} H_g(\alpha) = 1\},$$

$$\mathtt{B} = \{\beta \in \mathcal{K} \mid \operatorname{card} T_g(\beta) = 1\},$$

$$\mathtt{C} = \{\gamma \in \mathcal{K} \mid \operatorname{card} H_g(\gamma) = 0 = \operatorname{card} T_g(\gamma)\}.$$

form a partition of $\mathcal{K}$. The map $\mathtt{m} : \mathtt{A} \to \mathtt{B}$ given by

$$\mathtt{m}(\alpha) = \text{ unique } \beta \in H_g(\alpha),$$

defines a dvf V called the *gradient field* of g.
$(\mathtt{A}, \mathtt{B}, \mathtt{C}, \mathtt{m})$ is also called *partial matching*.

MDM function idea: back to 2012

What we could do:

- Prove an analogy of the sublevel set *deformation lemma*.

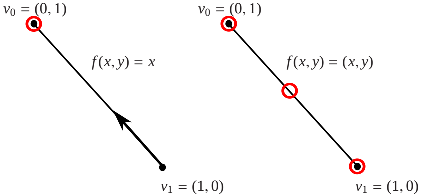
What we could not do:

- Provide a full extension of *Forman-Morse theory* in this setting;
- Design an algorithm producing an mdm function from *data on vertices*.

Chosen approach:

- Forget our MDM function;
- Design a function f on simplices as in 1D case;
- Declare *unpaired* simplices *critical*.

Algorithm design progress



- [King-Knudson-Mramor 2005] analogy (2015):
  Too many *unprocessed* simplices declared critical and dropped to $\mathcal{C}$.
- [Robins-Wood-Sheppard 2011] analogy (2017):
  More successful in reducing $\mathcal{C}$ thanks to processing cells of all dimensions, not only vertices.
- We now can get an *f-compatible* MDM function g.

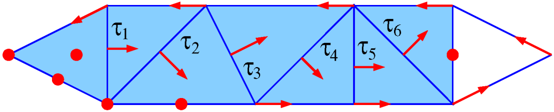
Partial matching

A *partial matching* $(\mathtt{A}, \mathtt{B}, \mathtt{C}, \mathtt{m})$ on a simplicial complex $\mathcal{K}$ is a partition $\mathtt{A}, \mathtt{B}, \mathtt{C}$ of $\mathcal{K}$ with a bijection

$$\mathtt{m} : \mathtt{A} \to \mathtt{B} \qquad \tau < \mathtt{m}(\tau).$$

such that $\mathtt{m}(\tau)$ is a cofacet of τ for all $\tau \in \mathtt{A}$.

$\mathtt{m} : \mathtt{A} \to \mathtt{B}$ Forman's *discrete vector field*, $\mathtt{C}$ *critical cells*



An *m–path* is a sequence

$$\tau_0 \mapsto \sigma_0 > \tau_1 \mapsto \sigma_1 > \ldots \tau_p \mapsto \sigma_p > \tau_{p+1}$$

A partial matching is *acyclic* if there is no closed $\mathtt{m}$–path.

Filtration, lower stars and indexing

Multifiltration is given initially on vertices $f : \mathcal{K}_0 \to \mathbb{R}^k$. It may be assumed that f is *component-wise injective*.

We extend it to $f : \mathcal{K} \to \mathbb{R}^k$ on all cells:

$$f(\sigma) = (f_1(\sigma), \ldots, f_k(\sigma)) \qquad \text{with} \qquad f_i(\sigma) = \max_{v \in \mathcal{K}_0(\sigma)} f_i(v).$$

The *sublevel set filtration* of $\mathcal{K}$

$$\mathcal{K}^a = \{\sigma \in \mathcal{K} \mid f(v) \preceq a \text{ for all } v \in \sigma\}, \quad a \in \mathbb{R}^k.$$

The *lower star* of $\sigma \in \mathcal{K}$ is $L(\sigma) = \{\alpha \in \mathcal{K} \mid \sigma \subseteq \alpha \quad \text{and} \quad f(\alpha) \preceq f(\sigma)\}$,
The *strict lower star* is $L_*(\sigma) = L(\sigma) \setminus \{\sigma\}$.

Topological Sorting Algorithm $\Rightarrow$ construction of an *indexing map* on $\mathcal{K}$, compatible with f:

A bijective map $I : \mathcal{K} \to \{1, 2, \ldots, N\}$, $N = \overline{\overline{\mathcal{K}}}$, such that

$$\sigma, \tau \in \mathcal{K}, \;\; \sigma \neq \tau, \;\; \sigma \subseteq \tau \text{ or } f(\sigma) \precneqq f(\tau) \quad \Rightarrow \quad I(\sigma) < I(\tau).$$

Goal: build a multifiltration-compatible partition of $\mathcal{K}$ into $\mathtt{A}$, $\mathtt{B}$, and $\mathtt{C}$, $\mathtt{m} : \mathtt{A} \to \mathtt{B}$, $\mathtt{C}$ declared critical.

- Process all cells σ of $\mathcal{K}$ increasingly with indexing I.
- Extra routines:
  - States $\mathtt{classified}(\sigma)$=true/ false, to avoid re-processing cells from lower stars of other cells and sets $\mathtt{unclass\_facets}_\sigma(\alpha)$, for $\alpha \in L_*(\sigma)$.
  - Priority queues $\mathtt{PQzero}$ and $\mathtt{PQone}$, to store cells with 0 and 1 available unclassified facets.
- σ is added to $\mathtt{C}$, if $L_*(\sigma) = \emptyset$. Otherwise, σ is paired with the cofacet $\delta \in L_*(\sigma)$ of minimal index $I(\delta)$.
- Additional pairings interpreted as building $L_*(\sigma)$ with simple homotopy expansions or reducing it with contractions:
  - α is a candidate for pairing when $\mathtt{unclass\_facets}_\sigma(\alpha)$ contains exactly one λ that belongs to $\mathtt{PQzero}$.
  - If no pairing of α is possible, add it to $\mathtt{C}$ and continue from that cell.
  - When $\mathtt{PQone} \neq \emptyset$, its front is popped out and either inserted into $\mathtt{PQzero}$ or paired with its single available unclassified facet.
  - When $\mathtt{PQone} = \emptyset$, the front cell of $\mathtt{PQzero}$ is added to $\mathtt{C}$.

Matching Algorithm

Algorithm 2 Matching

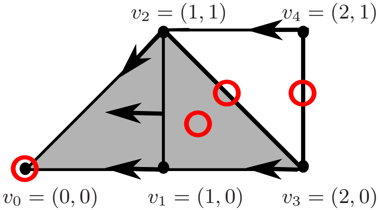
Input: A finite simplicial complex $\mathcal{K}$ with an admissible function $f : \mathcal{K} \to \mathbb{R}^k$ and an indexing map $I : \mathcal{K} \to \{1, 2, \ldots, N\}$ on its simplices compatible with f.
Output: Three lists A, B, C of simplices of $\mathcal{K}$, and a function m : A $\to$ B.
for $i = 1$ to N **do**
  $\sigma := I^{-1}(i)$
  if classified(σ)=**false then**
    if $L_*(\sigma)$ contains no cells **then**
      add σ to C, classified(σ)=**true**
    else
      $\delta :=$ the cofacet in $L_*(\sigma)$ of minimal index $I(\delta)$
      add σ to A and δ to B and define $\mathtt{m}(\sigma) = \delta$, classified($\sigma$)=**true**, classified($\delta$)=**true**
      add all $\alpha \in L_*(\sigma) - \{\delta\}$ with num_unclass_facets$_\sigma(\alpha) = 0$ to PQzero
      add all $\alpha \in L_*(\sigma)$ with num_unclass_facets$_\sigma(\alpha) = 1$ and $\alpha > \delta$ to PQone
      while PQone $\neq \emptyset$ or PQzero $\neq \emptyset$ **do**
        while PQone $\neq \emptyset$ **do**
          $\alpha :=$ PQone.pop_front
          if num_unclass_facets$_\sigma(\alpha) = 0$ **then**
            add α to PQzero
          else
            add $\lambda \in$ unclass_facets$_\sigma(\alpha)$ to A, add α to B and define $\mathtt{m}(\lambda) = \alpha$, classified($\alpha$)=**true**, classified($\lambda$)=**true**
            remove λ from PQzero
            add all $\beta \in L_*(\sigma)$ with num_unclass_facets$_\sigma(\beta) = 1$ and either $\beta > \alpha$ or $\beta > \lambda$ to PQone
          end if
        end while
        if PQzero $\neq \emptyset$ **then**
          $\gamma :=$ PQzero.pop_front
          add γ to C, classified(γ)=**true**
          add all $\tau \in L_*(\sigma)$ with num_unclass_facets$_\sigma(\tau) = 1$ and $\tau > \gamma$ to PQone
        end if
      end while
    end if
  end if
end for

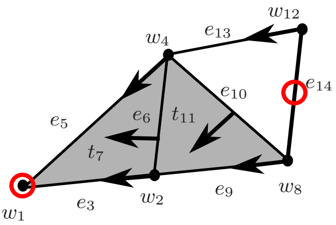
Theorem

The algorithm produces a multifiltration-compatible partial matching (A, B, C, m) that is acyclic.
The worst case processing cost is $O(N \cdot \gamma \log \gamma)$, where

$$N := \overline{\overline{\mathcal{K}}},\ \gamma := \max_{\sigma \in \mathcal{K}} \overline{\overline{\mathrm{cbd}(\sigma)}},\ \textit{and}\ \mathrm{cbd}(\sigma) := \{\tau \in \mathcal{K} \mid \sigma \leq \tau\}.$$

Example





$i = 1$	$L_*(w_1) = \emptyset$, $w_1 \in \mathtt{C}$.	$5, 6, 7$	e_5, e_6, t_7 classified.
$i = 2$	$L_*(w_2) = \{e_3\}, \mathtt{m}(w_2) = e_3$.	$i = 8$	$L_*(w_8) = \{e_9\}, \mathtt{m}(w_8) = e_9$.
$i = 3$	e_3 classified.	$i = 9$	e_9 classified.
$i = 4$	$L_*(w_4) = \{e_5, e_6, t_7\}$,	$i = 10$	$L_*(e_1 0) = \{t_1 1\}, \mathtt{m}(e_1 0) = t_1 1$.
	$\mathtt{m}(w_4) = e_5$,	$i = 11$	$t_1 1$ classified.
	$e_6 \in \mathtt{PQzero}$, $t_7 \in \mathtt{PQone}$,	$i = 12$	$L_*(w_{12}) = \{e_{13}, e_{14}\}$,
	line 15, $\alpha = t_7$ leaves		$\mathtt{m}(w_{12}) = e_{13}$,
	PQone,		$e_{14} \in \mathtt{PQzero}$, $\mathtt{PQone} = \emptyset$,
	line 19, $\lambda = e_6, \mathtt{m}(e_6) = t_7$,		line 25, $\gamma = e_{14} \in \mathtt{C}$.
	e_6 leaves PQzero.	$13, 14$	e_{13}, e_{14} classified.

Red circles: Cells left in C; Left: 2015 algorithm; Right: 2017 algorithm.

Lefschetz complex reductions

$\mathtt{S} = \{\mathtt{S}_q\}$ *cells*, $\tau < \sigma$ *facets*, $\kappa(\sigma, \tau)$ *incidence* $\Rightarrow$ $C_*(\mathtt{S}, \partial^\kappa)$ chain complex.

$\{\mathtt{S}^a\}_{a \in \mathbb{R}^k}$ is a *multi-filtration* of $\mathtt{S}$ if

- $a \preceq b \;\Rightarrow\; \mathtt{S}^a \subseteq \mathtt{S}^b$,
- $\sigma \in \mathtt{S}^a,\ \tau \leq \sigma \;\Rightarrow\; \tau \in \mathtt{S}^a$.

Persistent homology

$$H_q^{a,b}(\mathtt{S}) := \operatorname{im} H_q(j^{(a,b)}) \;,\; j^{(a,b)} : \mathtt{S}^a \hookrightarrow \mathtt{S}^b.$$

$(\mathtt{A}, \mathtt{B}, \mathtt{C}, \mathtt{m})$ on $(\mathtt{S}, \kappa)$, $\sigma \in \mathtt{A}$ $\Rightarrow$ *reduced complex* $(\overline{\mathtt{S}}, \overline{\kappa})$,
$\overline{\mathtt{S}} = \mathtt{S} \setminus \{\mathtt{m}(\sigma), \sigma\}$, and $\overline{\kappa} : \overline{\mathtt{S}} \times \overline{\mathtt{S}} \to R$,

$$\overline{\kappa}(\eta, \xi) = \kappa(\eta, \xi) - \kappa(\eta, \sigma)\kappa(\mathtt{m}(\sigma), \xi)\kappa^{-1}(\mathtt{m}(\sigma), \sigma).$$

Isomorphism Lemma

$$\begin{array}{ccc} H_*(\mathtt{S}^a) & \xrightarrow{H_*(j^{(a,b)})} & H_*(\mathtt{S}^b) \\ \downarrow \cong & & \downarrow \cong \\ H_*(\overline{\mathtt{S}}^a) & \xrightarrow{H_*(j^{(a,b)})} & H_*(\overline{\mathtt{S}}^b) \end{array} \quad , \; a \preceq b.$$

Iterated reductions

$$\mathcal{K}^a =: \mathrm{S}^a(0) \supset \mathrm{S}^a(1) \supset \ldots \supset \mathrm{S}^a(n) = \mathrm{C}^a\ .$$

Corollary

For every $a \preceq b$, $H_*^{a,b}(\mathcal{C}) \cong H_*^{a,b}(\mathcal{K})$. *Moreover, the diagram*

$$\begin{array}{ccc} H_*(\mathcal{K}^a) & \xrightarrow{H_*(j^{(a,b)})} & H_*(\mathcal{K}^b) \\ \downarrow \cong & & \downarrow \cong \\ H_*(\mathcal{C}^a) & \xrightarrow{H_*(j^{(a,b)})} & H_*(\mathcal{C}^b) \end{array} \qquad a \preceq b.$$

commutes.

Worst case cost $O(N\,\gamma\, m^2)$, $m := \overline{\overline{\mathrm{C}}}$.

Best results when grid is fixed ($\Rightarrow\ \gamma$ constant) and m small w.r.t. N.

f-compatible mdm functions g

Recall that $g : \mathcal{K} \to \mathbb{R}^k$ is *mdm* if

(1) $\operatorname{card} H_g(\alpha) \leq 1$;

(2) $\operatorname{card} T_g(\alpha) \leq 1$;

(3) If $\beta^{(p+1)} > \alpha$ is not in $H_g(\alpha)$, then $g(\alpha) \precneqq g(\beta)$;

(4) If $\gamma^{(p-1)} < \alpha$ is not in $T_g(\alpha)$, then $g(\gamma) \precneqq g(\alpha)$.

Proposition

Any $f : \mathcal{K} \to \mathbb{R}^k$ used as input in the Matching Algorithm satisfies conditions (3) and (4).

In general, (1) and (2) may fail.

$\sigma \in \mathcal{K}$ is *primary*, if it is classified by Matching Algorithm at the beginning of processing its own lower star at lines 7 or 10.
$P = \{\sigma_{i_j}\}$ all primary simplices ordered increasingly by I.

Proposition

The lower stars of primary simplices $L(\sigma_{i_j})$ form a partition of $\mathcal{K}$.

Definition

$g : \mathcal{K} \to \mathbb{R}^k$ *is* f-`compatible` *provided that*

(1) $f(\alpha) \precneqq f(\beta) \ \Rightarrow \ g(\alpha) \precneqq g(\beta)$; *and*

(2) *if* $\alpha, \beta \in L(\sigma_{i_j})$ *for a primary* σ_{i_j} *and* α *is classified earlier than* β, *then* $g(\alpha) \precneqq g(\beta)$.

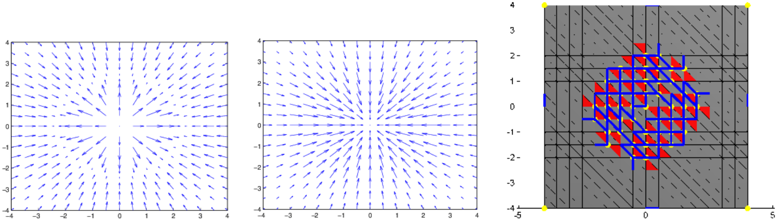
Theorem

Let $g : \mathcal{K} \to \mathbb{R}^k$ *be* f-`compatible`. *Then* g *is an mdm function, and its partial matching coincides with that produced by Matching Algorithm.*

Theorem

There exist f-`compatible` *functions* g.

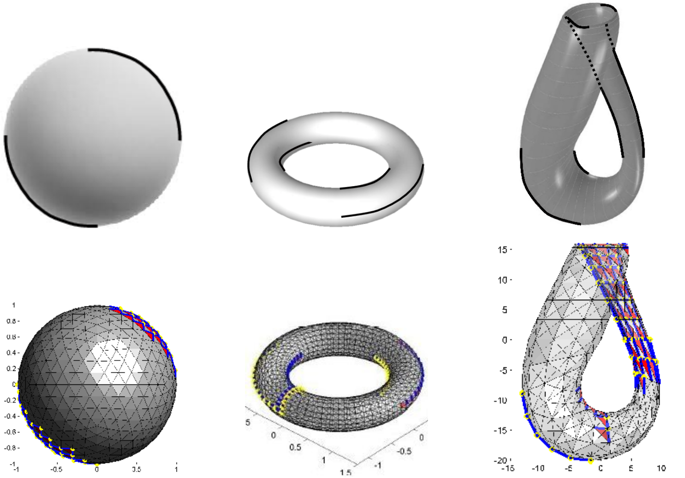
Interpretation of retrieved critical cells



Left and center: gradient vector fields of two scalar functions f_1, f_2.

Right: critical cells of dimension 0 in yellow, dimension 1 in blue and dimension 2 in red for $f = (f_1, f_2)$ as retrieved by Matching Algorithm.

Pareto: smooth and discrete



Left: Pareto critical curves for two projection maps.
Right: Critical cells retrieved by the algorithm: vertices - yellow, edges - blue, triangles - red.

Future work

- Improve construction of f-compatible mdm functions g.
- Continue developing extension of the combinatorial Morse theory to multidimensional functions.
- Further experiments, applications, and optimization.

M. Allili, TK, and C. Landi, *Reducing complexes in multidimensional persistent homology theory*, J. Symb. Comp. **78** (2017), 61–75.

—, —, —, and F. Masoni, *Algorithmic construction of acyclic partial matchings for multidimensional persistence*, in DGCI 2017.

—, —, —, —, *Acyclic Partial Matchings for Multidimensional Persistence: Algorithm and Combinatorial Interpretation* , preprint 2018.

Gracias por su atención!