Advanced Data Visualization

CS 6965

Spring 2018

Prof. Bei Wang Phillips

University of Utah

Lecture 17

About Project Proposal

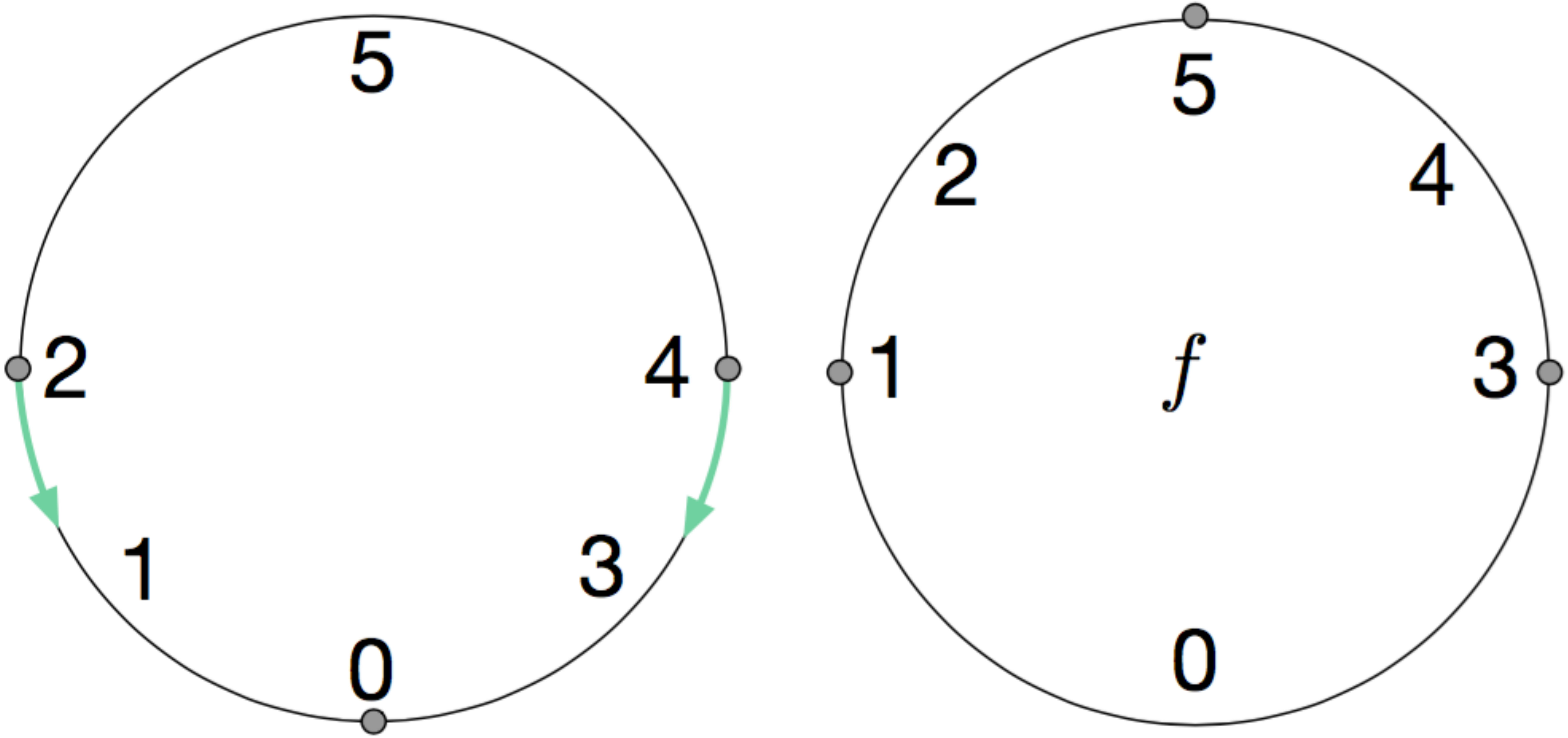
- See details on project proposal requirement.

http://www.sci.utah.edu/~beiwang/teaching/cs6965-spring-2018/final-project-proposal.pdf

DMT Examples Structural Inference of High-dim Data

TOPO

DMT Review



- $U(\alpha) = \{\beta^{(p+1)} > \alpha \mid f(\beta) \leq f(\alpha)\}$
- $L(\alpha) = \{\gamma^{(p-1)} < \alpha \mid f(\gamma) \geq f(\alpha)\}$

Forman 2002

Discrete Morse Function

Definition

A function $f : K \to \mathbb{R}$ is a *discrete Morse function* if for every $\alpha^{(p)} \in K$,

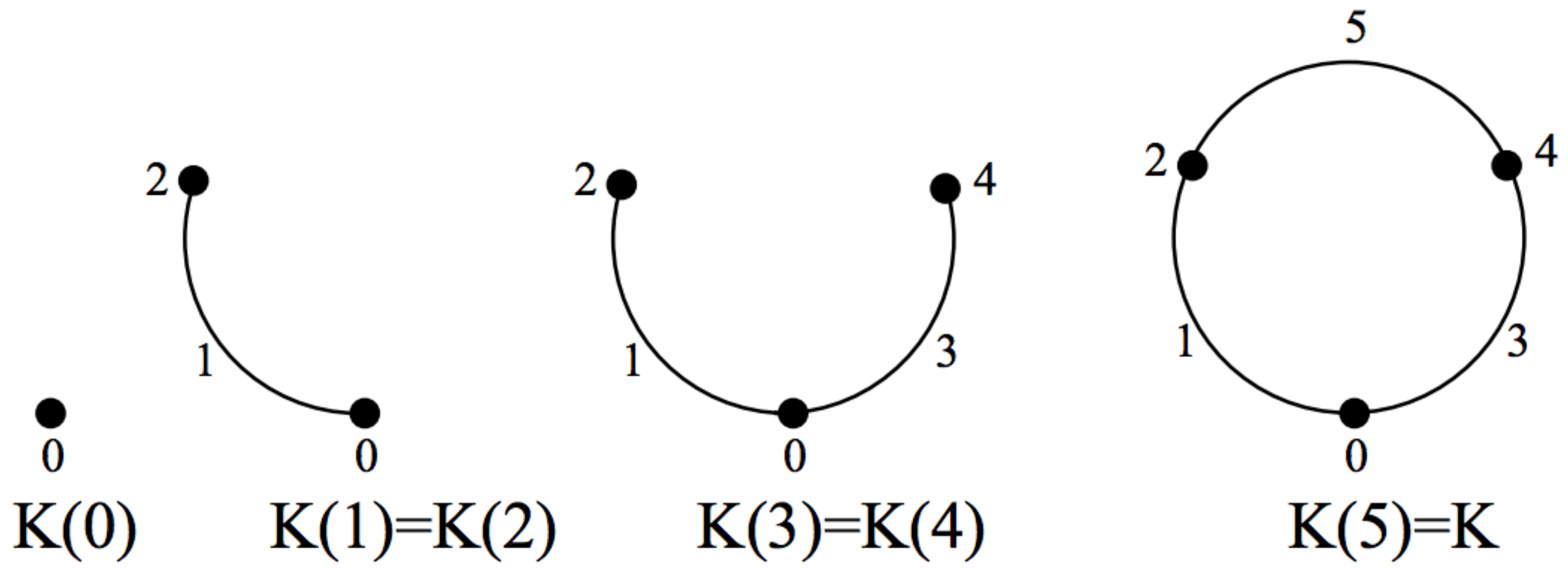
- (i) $|U(\alpha)| \leq 1$ and
- (ii) $|L(\alpha)| \leq 1$.

Definition

A simplex $\alpha^{(p)}$ is *critical* if (i) $|U(\alpha)| = 0$ and (ii) $|L(\alpha)| = 0$. A *critical value* of f is its value at a critical simplex.

Level Subcomplex

$$K_c = \cup_{f(\alpha)\leq c} \cup_{\beta \leq \alpha} \beta.$$



The level subcomplexes of the discrete Morse function shown in Figure 2.2(ii)

Discrete Gradient Vector Fields

- Since any noncritical simplex $\alpha^{(p)}$ has at most one of the sets $U(\alpha)$ and $L(\alpha)$ nonempty, there is a unique face $\nu^{(p-1)} < \alpha$ with $f(\nu) \geq f(\alpha)$ or a unique coface $\beta^{(p+1)} > \alpha$ with $f(\beta) \leq f(\alpha)$.
- Denote by V the collection of all such pairs $\{\sigma < \tau\}$.
- Then every simplex in K is in at most one pair in V and the simplices not in any pair are precisely the critical cells of the function f.
- We call V the *gradient vector field associated to* f.
- We visualize V by drawing an arrow from α to β for every pair $\{\alpha < \beta\} \in V$.

Discrete Gradient Vector Fields

- Data reduction: collapsing the pairs in V using the arrows.
- By a V-*path*, we mean a sequence

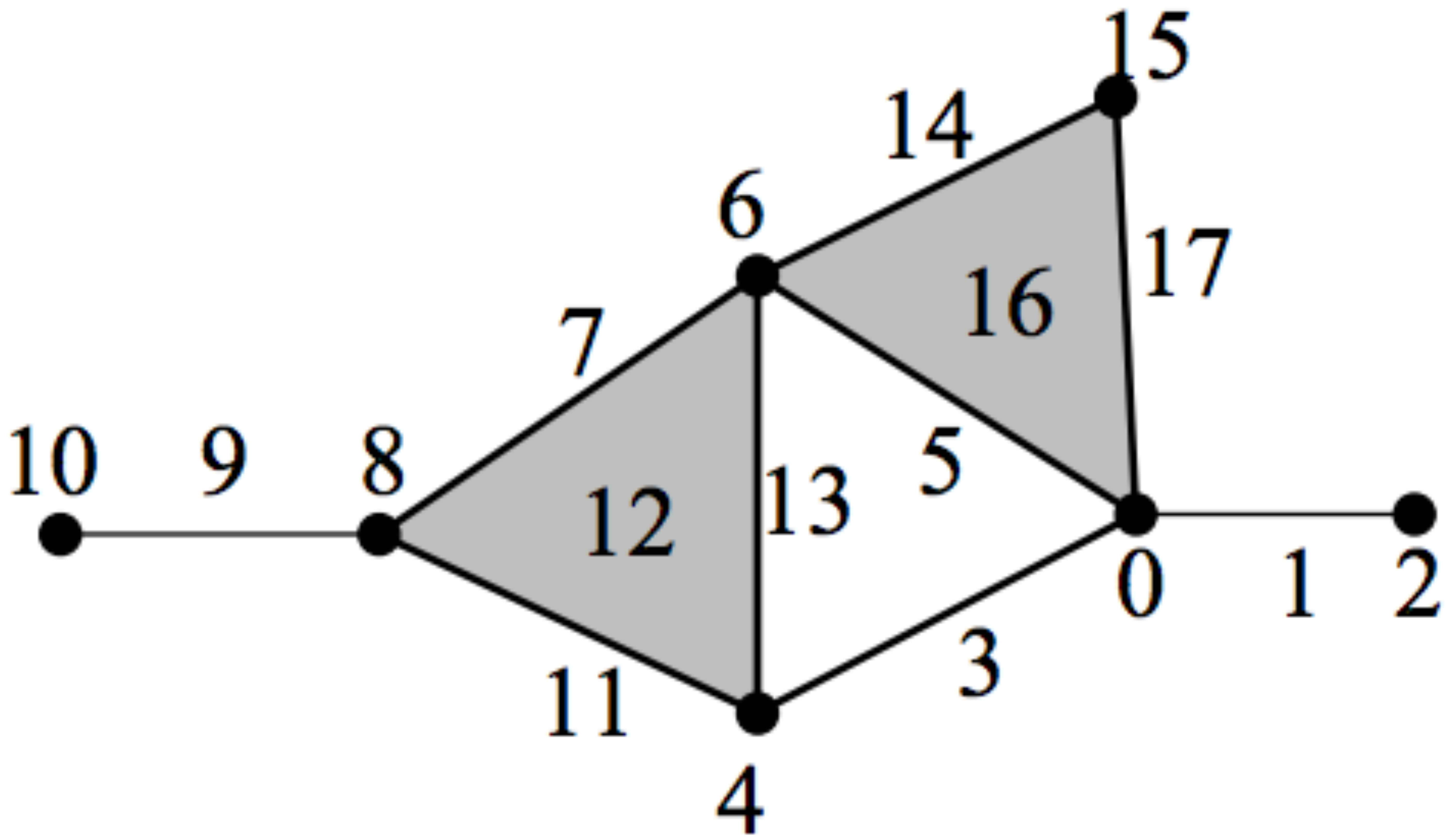
$$\alpha_0^{(p)} < \beta_0^{(p+1)} > \alpha_1^{(p)} < \beta_1^{(p+1)} > \cdots < \beta_r^{(p+1)} > \alpha_{r+1}^{(p)}$$

where each $\{\alpha_i < \beta_i\}$ is a pair in V. Such a path is *nontrivial* if $r > 0$ and *closed* if $\alpha_{r+1} = \alpha_0$.

Theorem ([Forman1998])

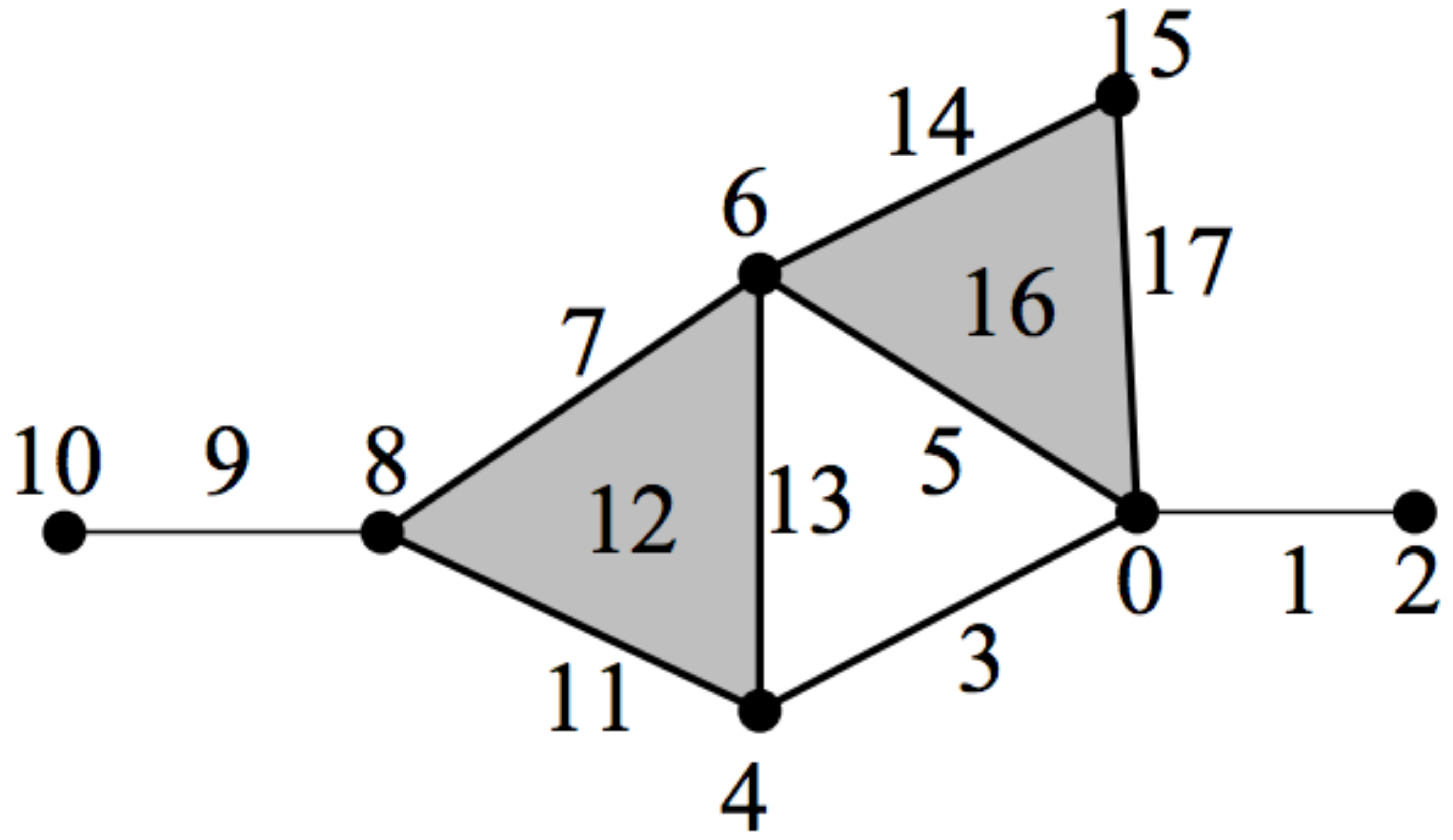
If V is a gradient vector field associated to a discrete Morse function f on K, then V has no nontrivial closed V-paths.

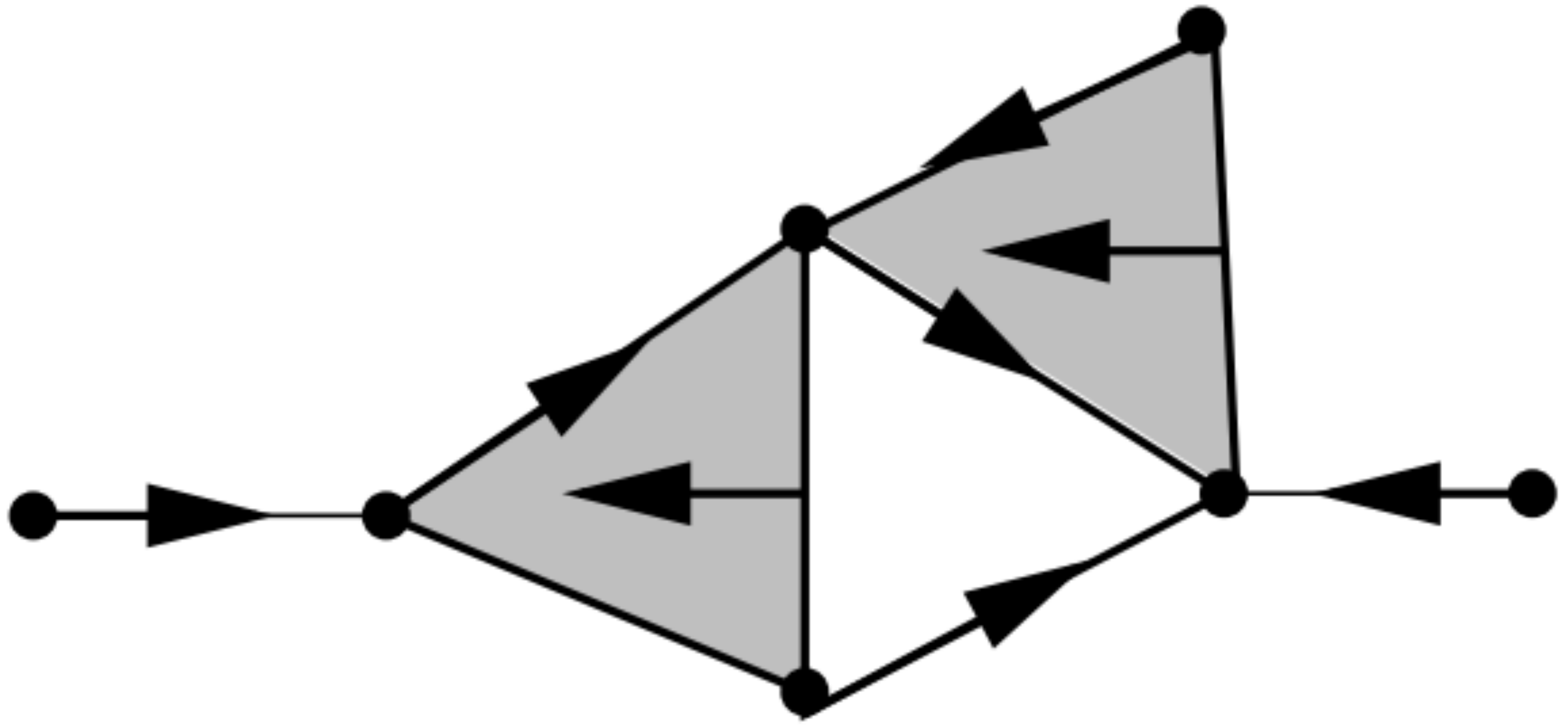
Draw the DVF



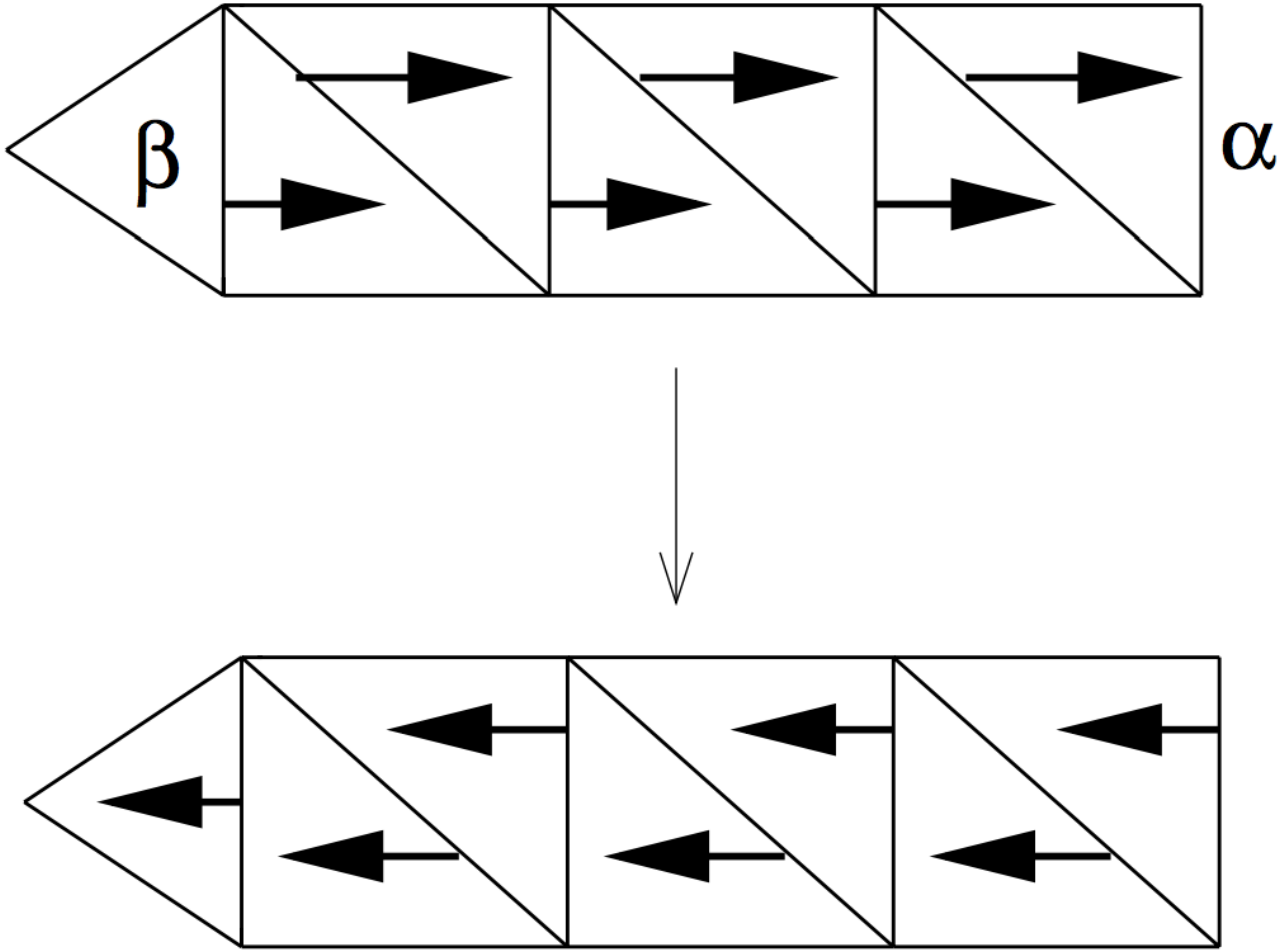
Forman 2002

Which simplices are critical?





Forman 2002



Forman 2002

Cancelling critical points.

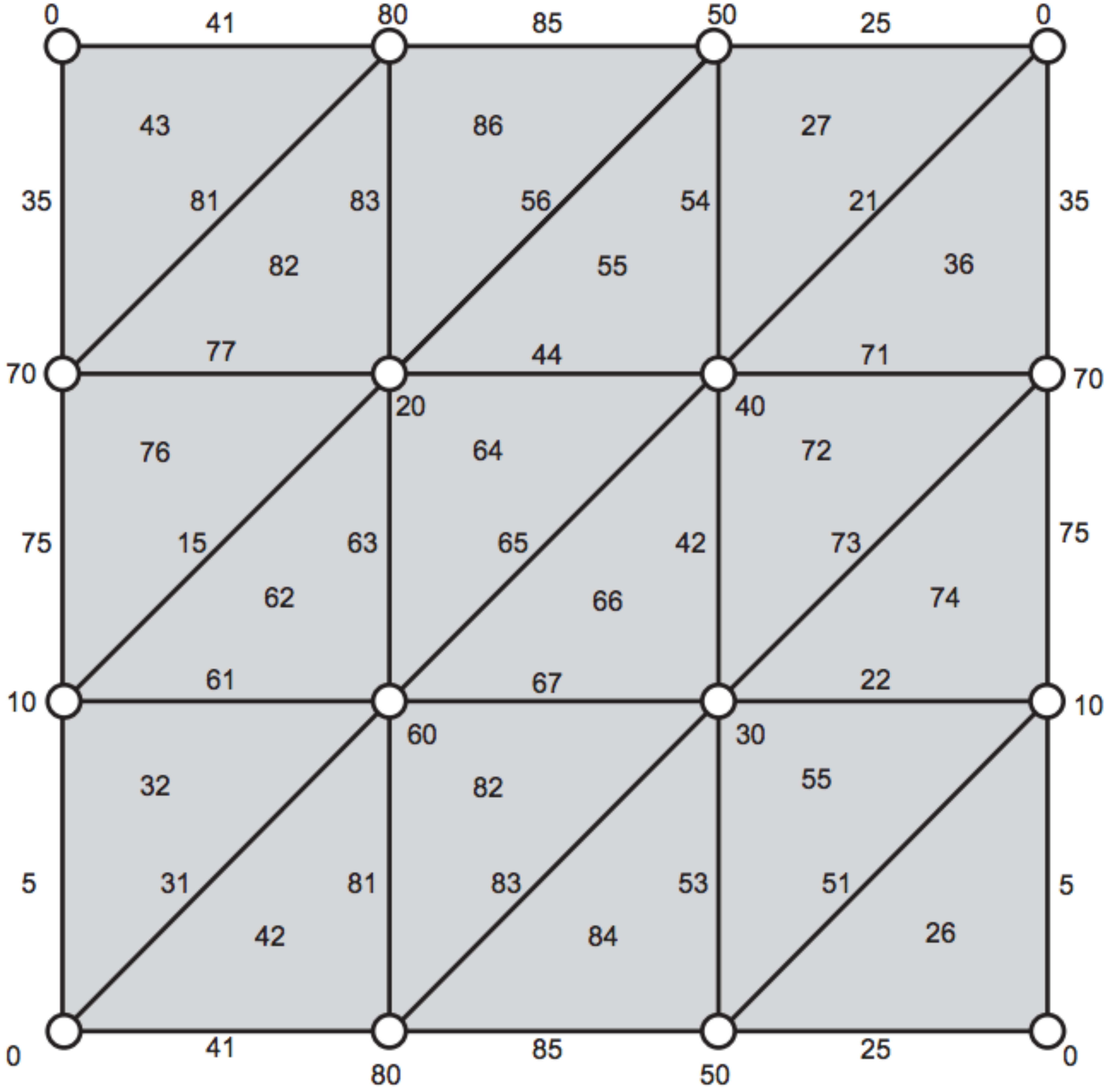


FIGURE 1. A discrete Morse function on the torus.

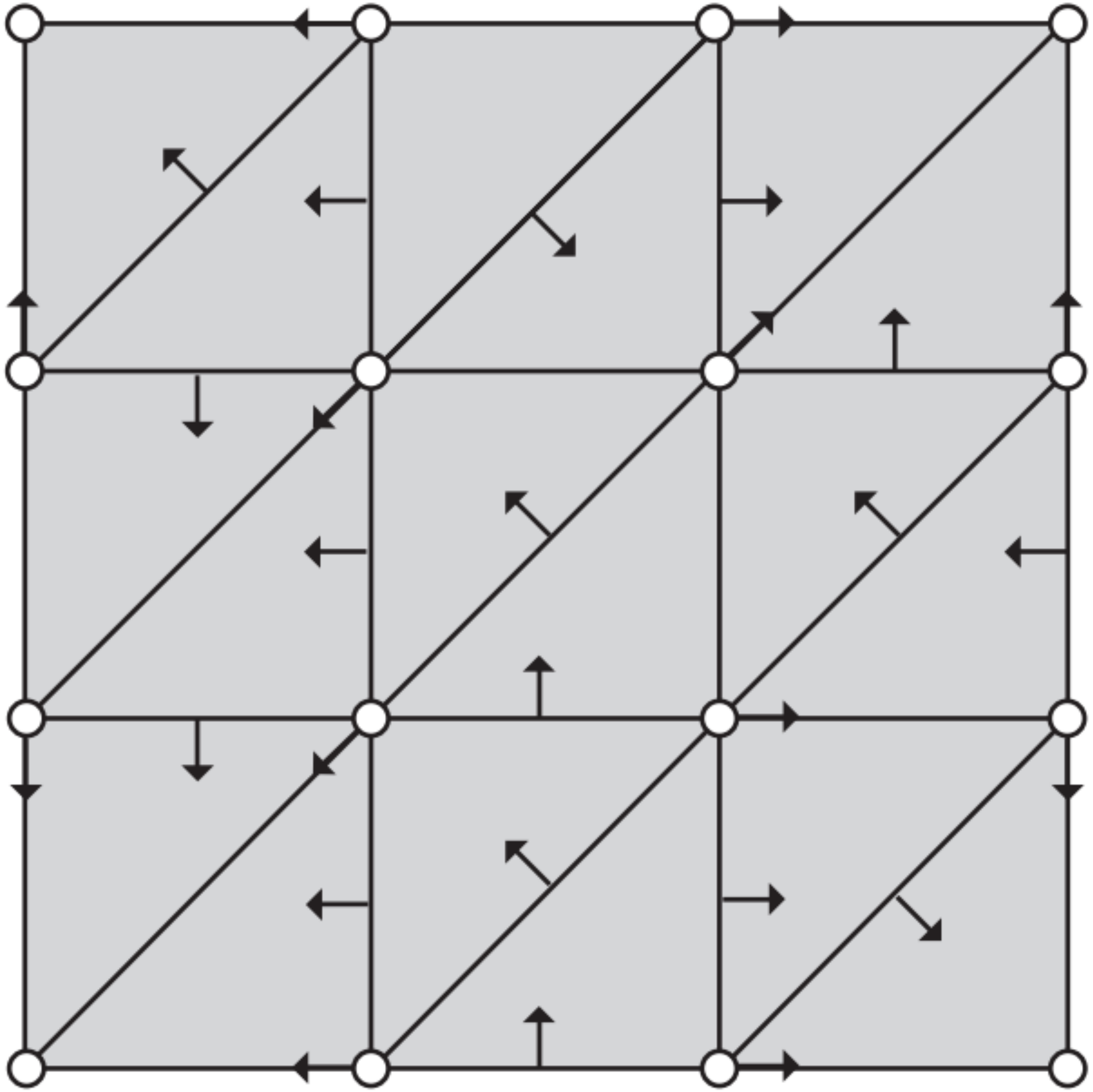
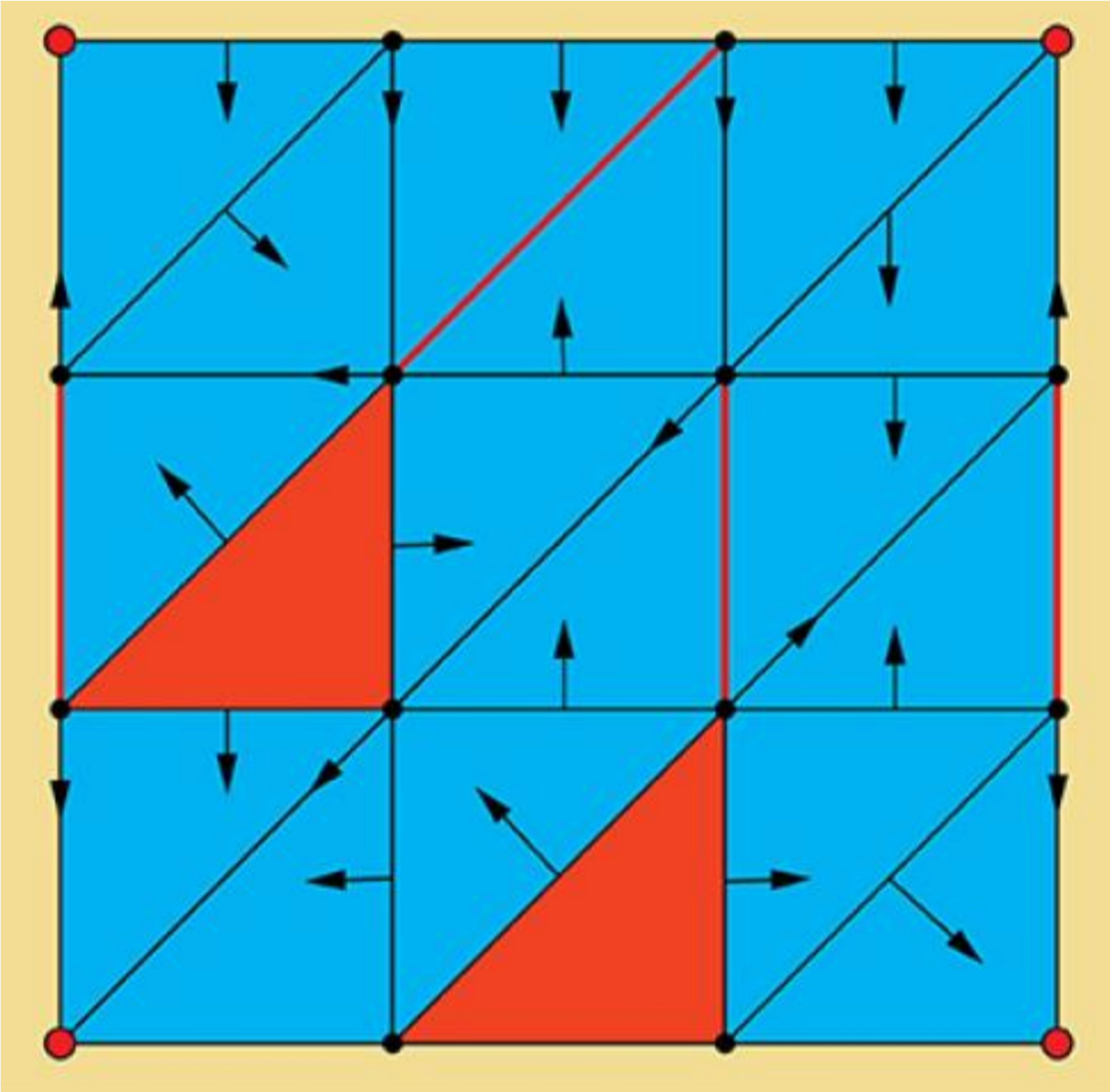
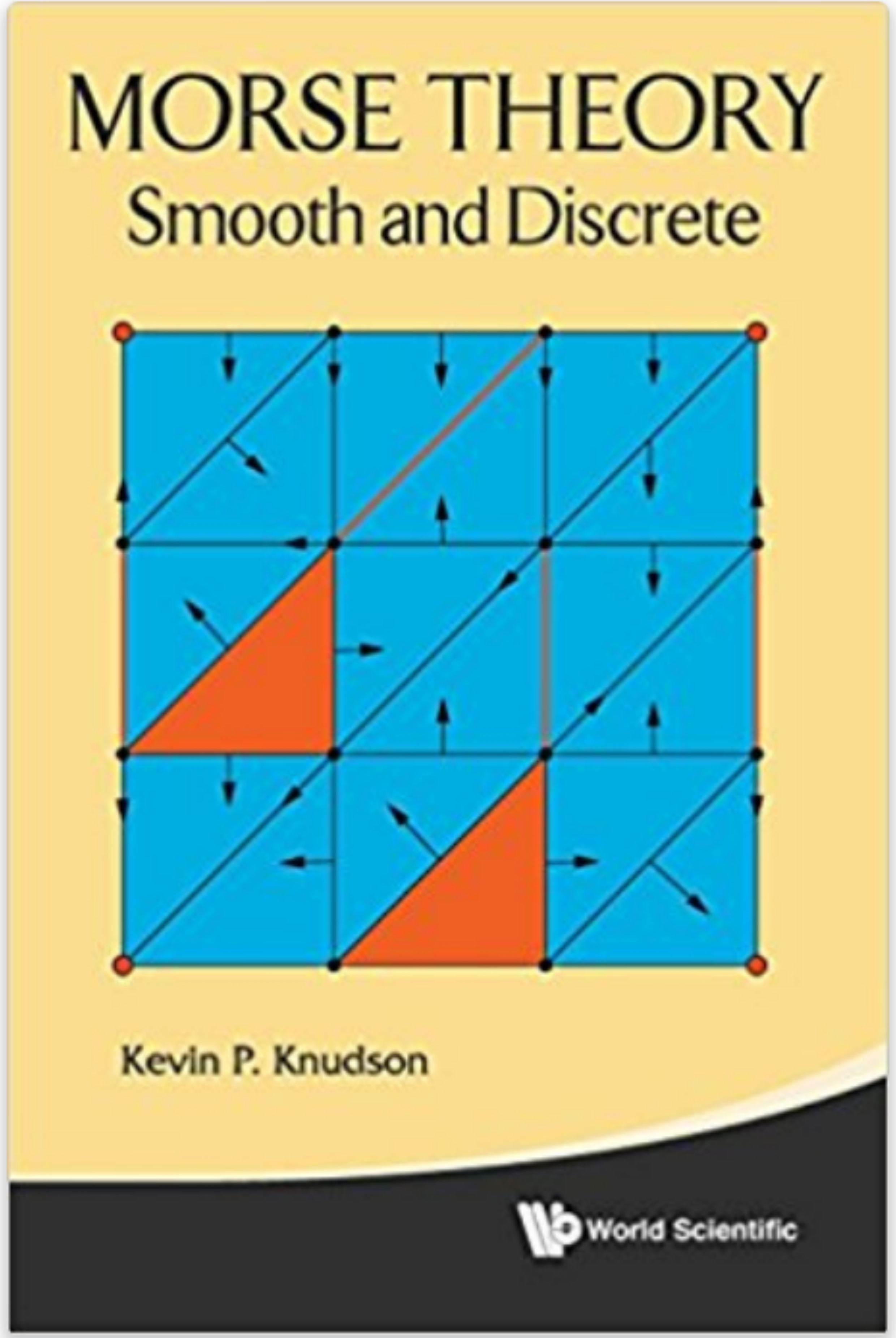


FIGURE 2. A gradient vector field on the torus.

King, Knudson, Mramor 2005




MORSE THEORY
Smooth and Discrete
Kevin P. Knudson
World Scientific

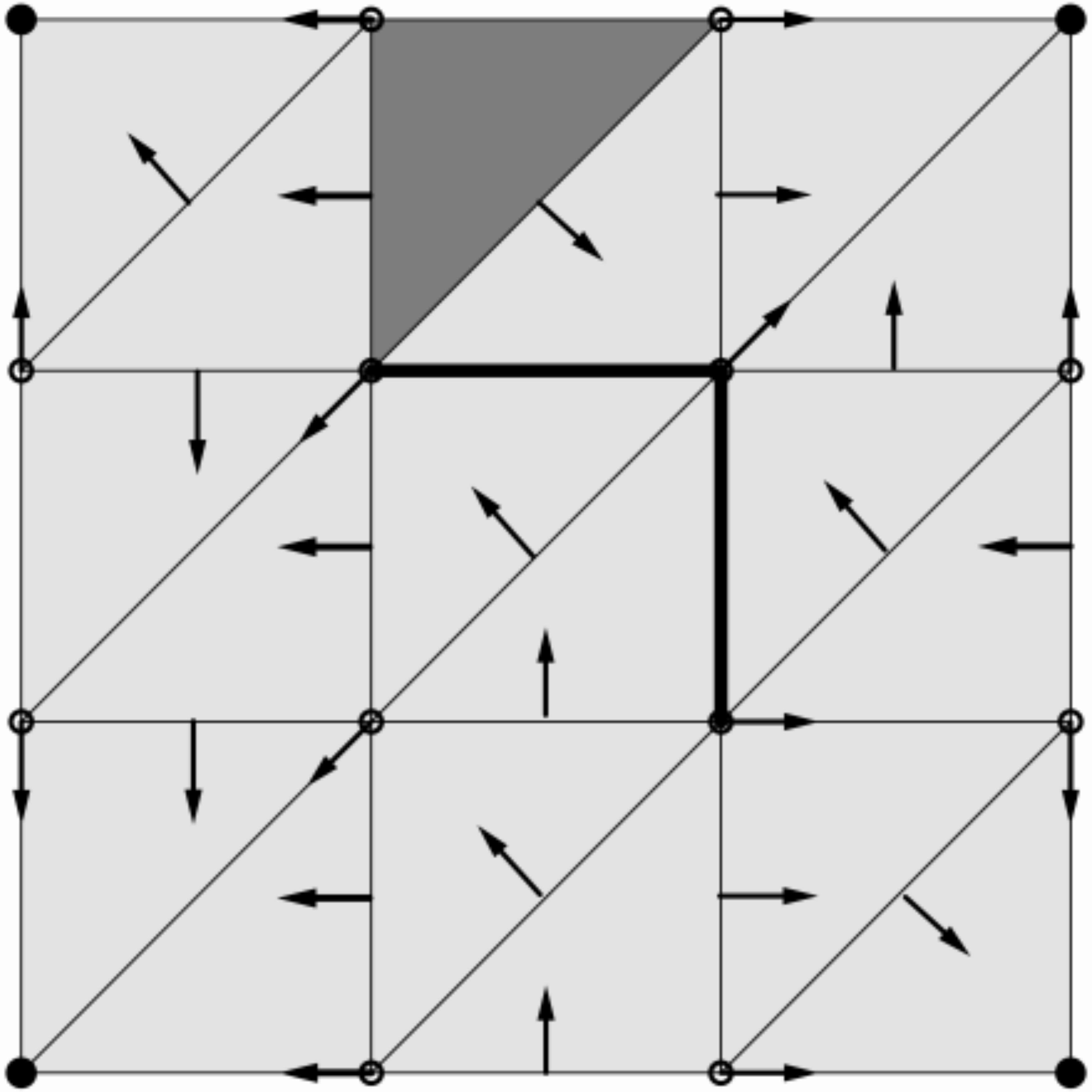


FIGURE 3. A discrete vector field on a triangulated torus. There are four critical cells: the top-center triangle (dark gray), the top and right edges of the center square (indicated with thicker lines), and the vertex obtained by identifying the four corners.

https://arxiv.org/pdf/0808.0051.pdf

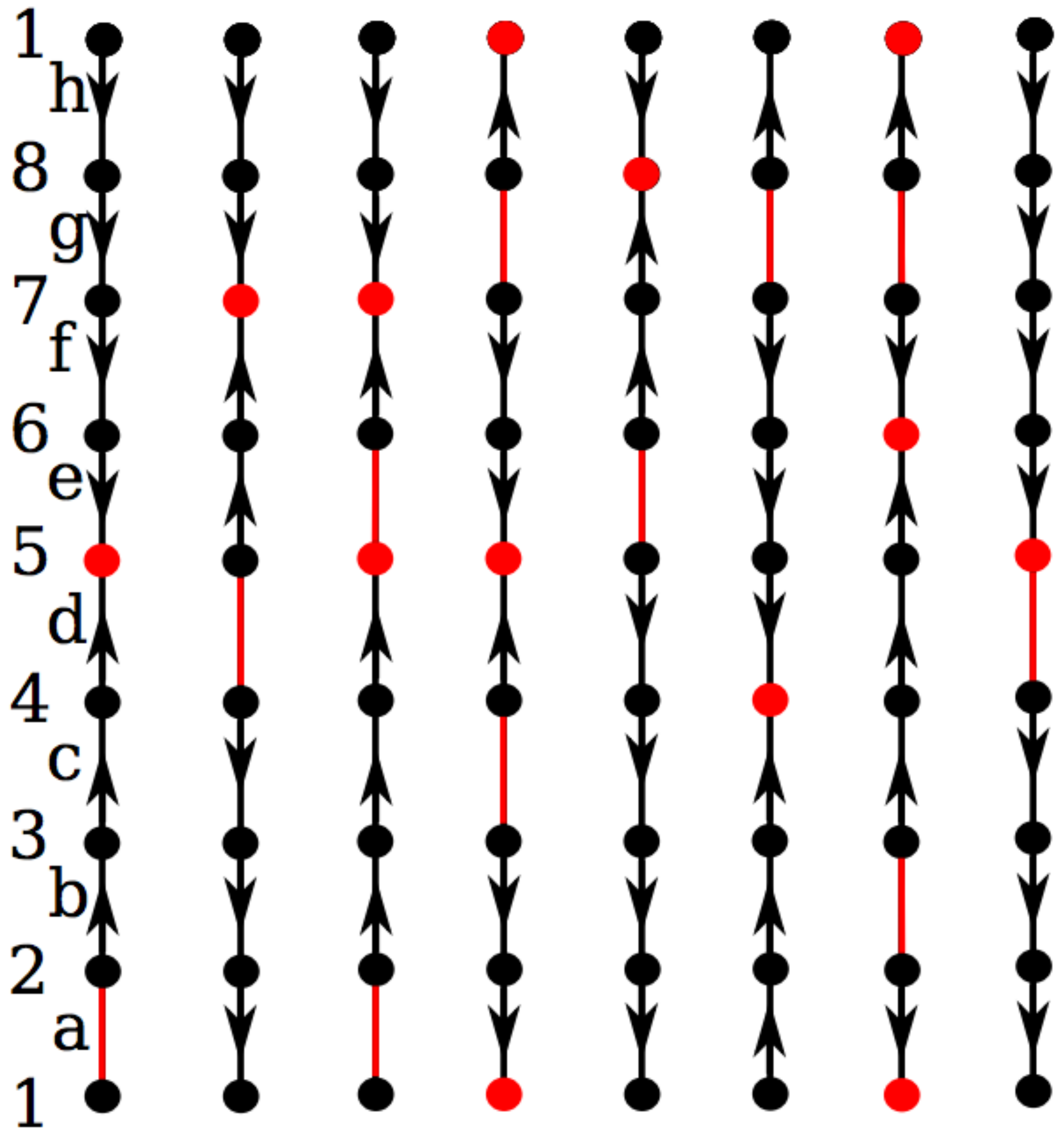


FIGURE 4. A family of discrete gradient fields on the circle. The slices are numbered left to right as $0, 1, \ldots, 7$.

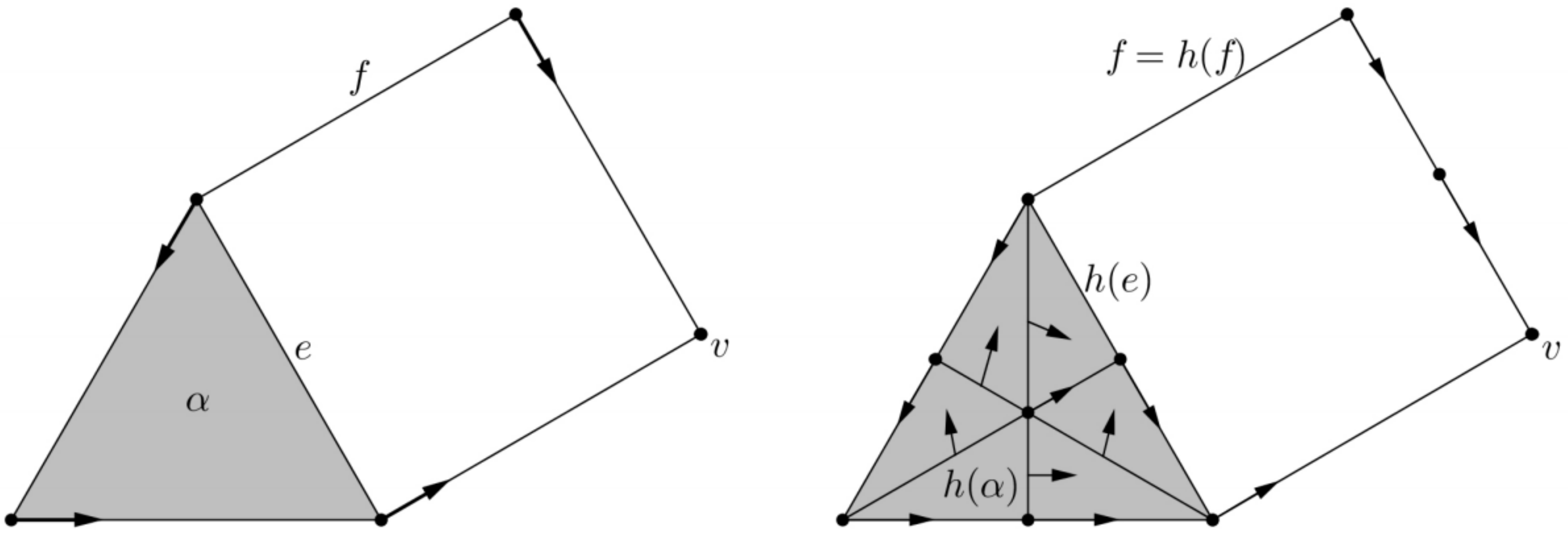
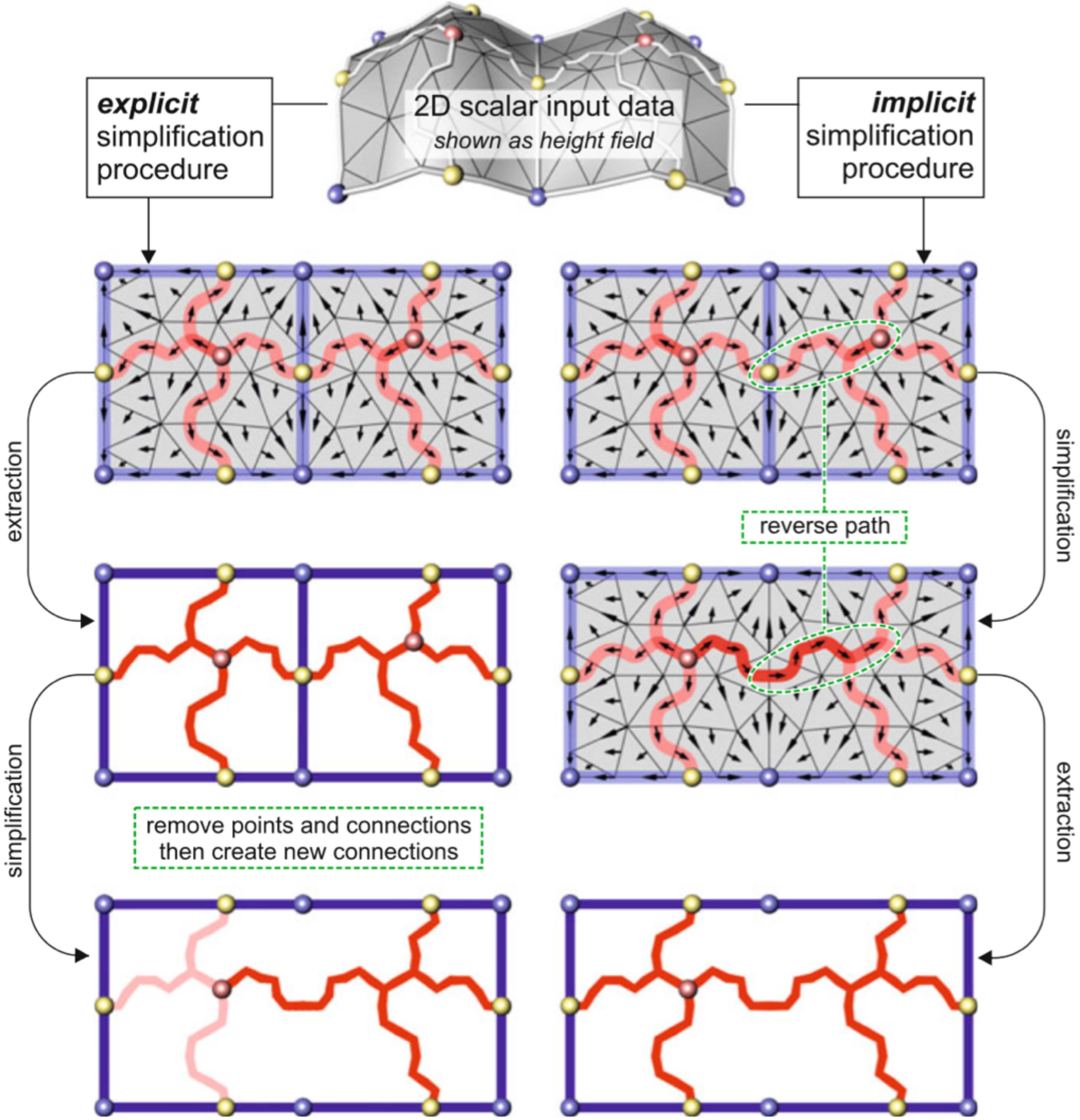


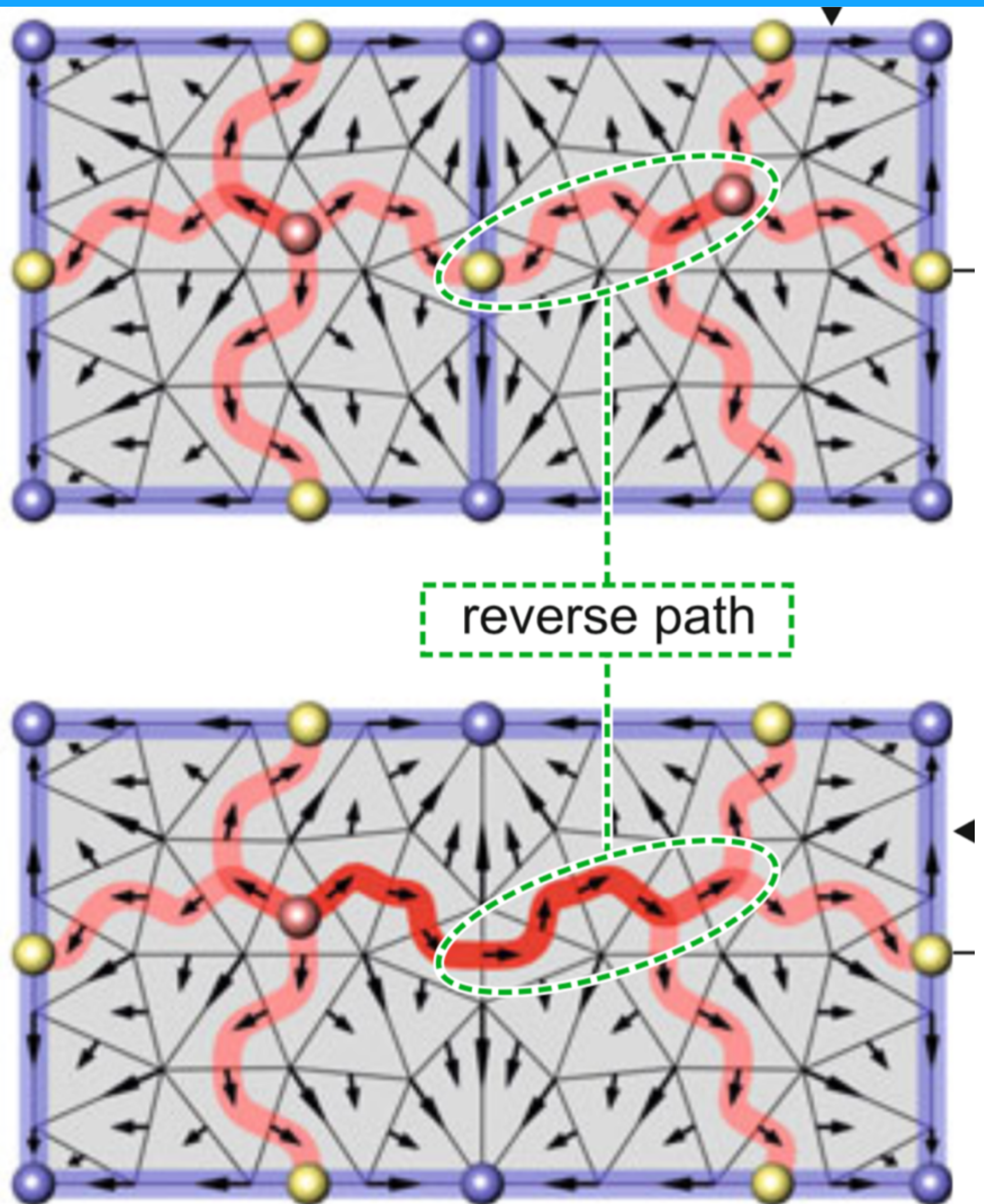
FIGURE 6. Left: A simplicial complex M with a discrete gradient V. The critical cells of V are α, e, f, and v. Right: A refinement N of M and a refinement W of V. The new critical cells $h(\alpha)$ and $h(e)$ are indicated.

DMS Applications

Simplification of MSC



GuntherReininghausSeidel2014


reverse path

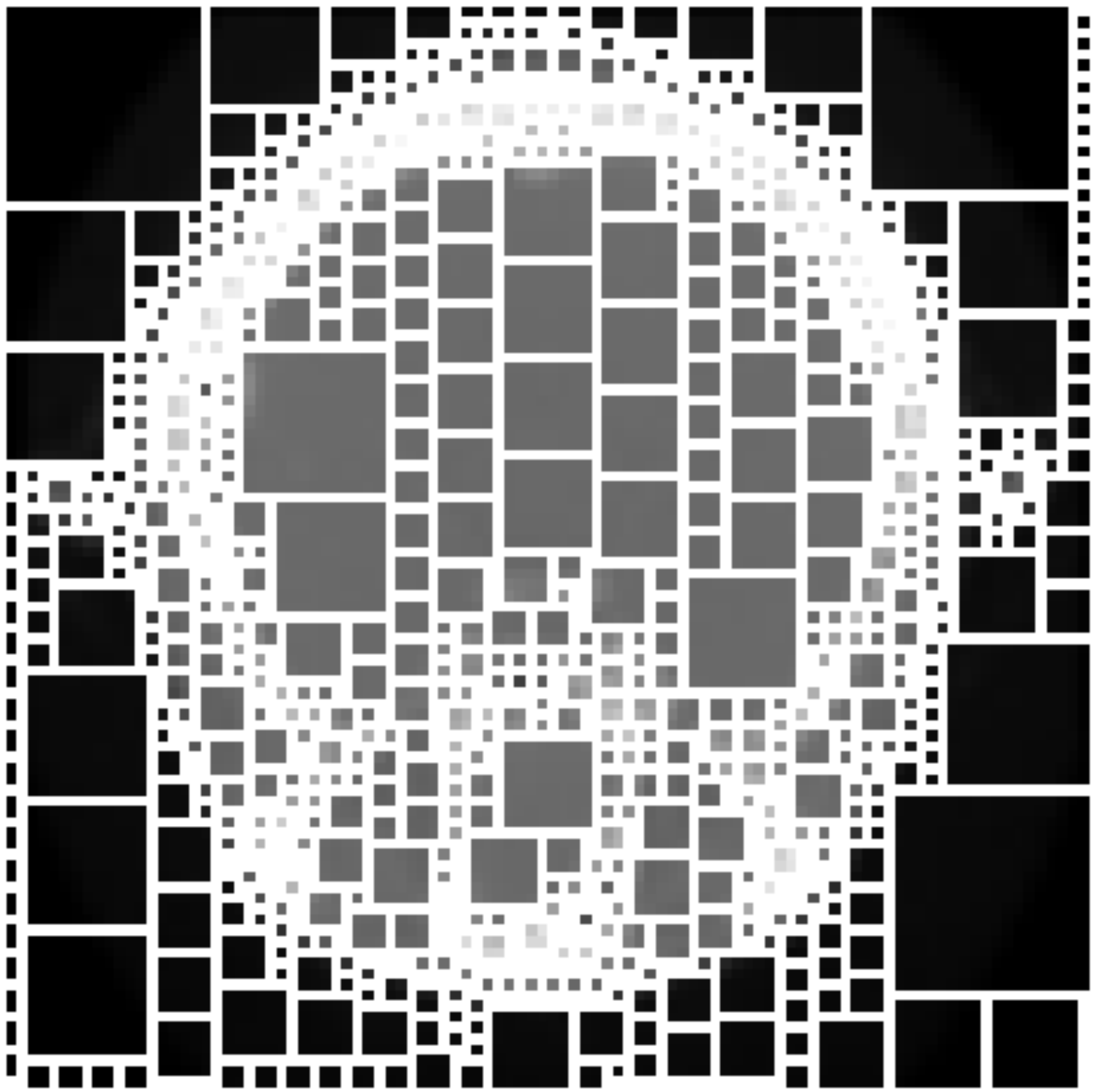


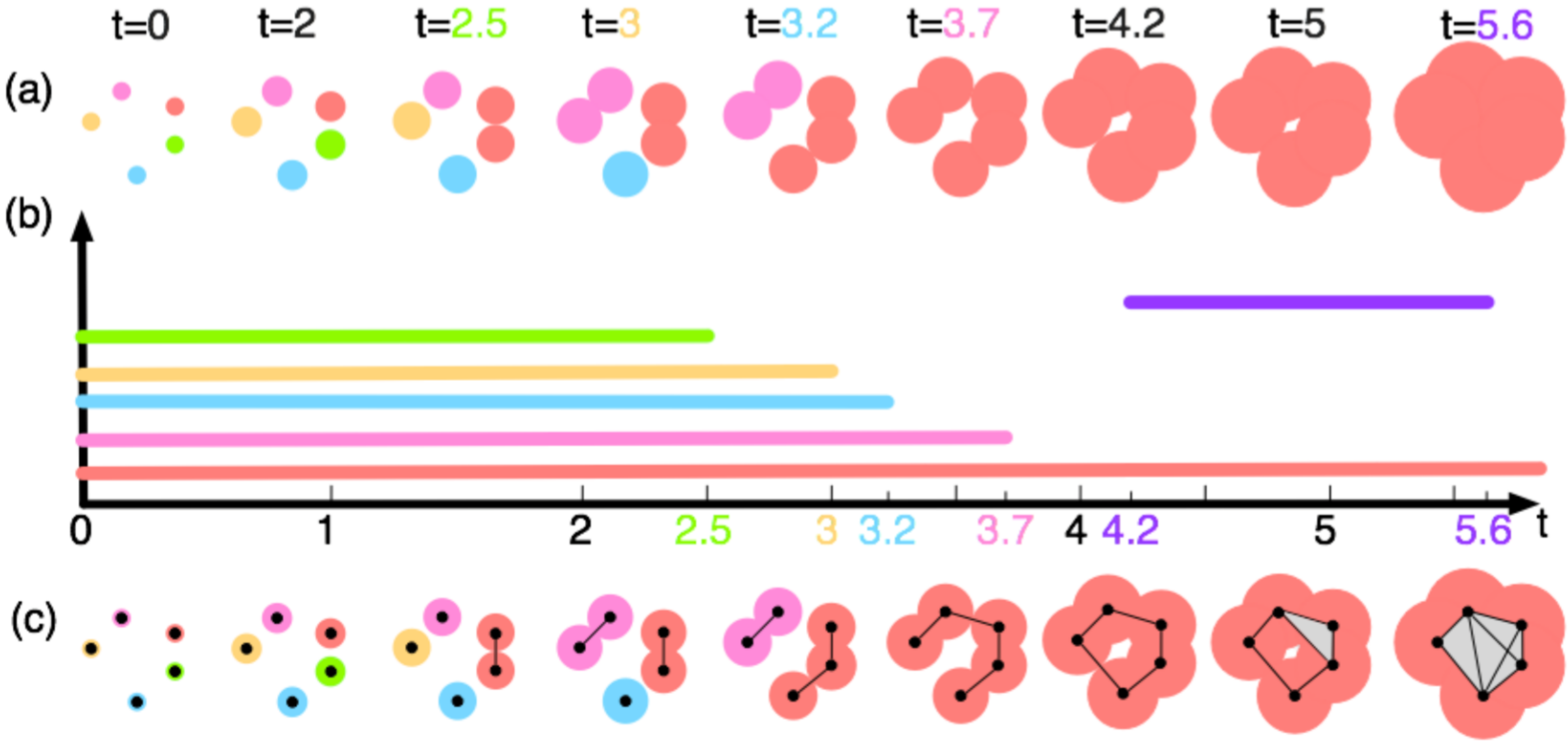
FIGURE 15. Reduced cell decomposition of a head CT scan.

KingKnudsonKosta2016

Structural Inference of High-dimensional Data

A topological view

Review:
Persistent Homology



t=0
t=2
t=2.5
t=3
t=3.2
t=3.7
t=4.2
t=5
t=5.6
(a)
(b)
0
1
2
2.5
3
3.2
3.7
4
4.2
5
5.6
t
(c)

Stability of Persistence Diagrams

Homological features encoded as barcodes or persistent diagrams

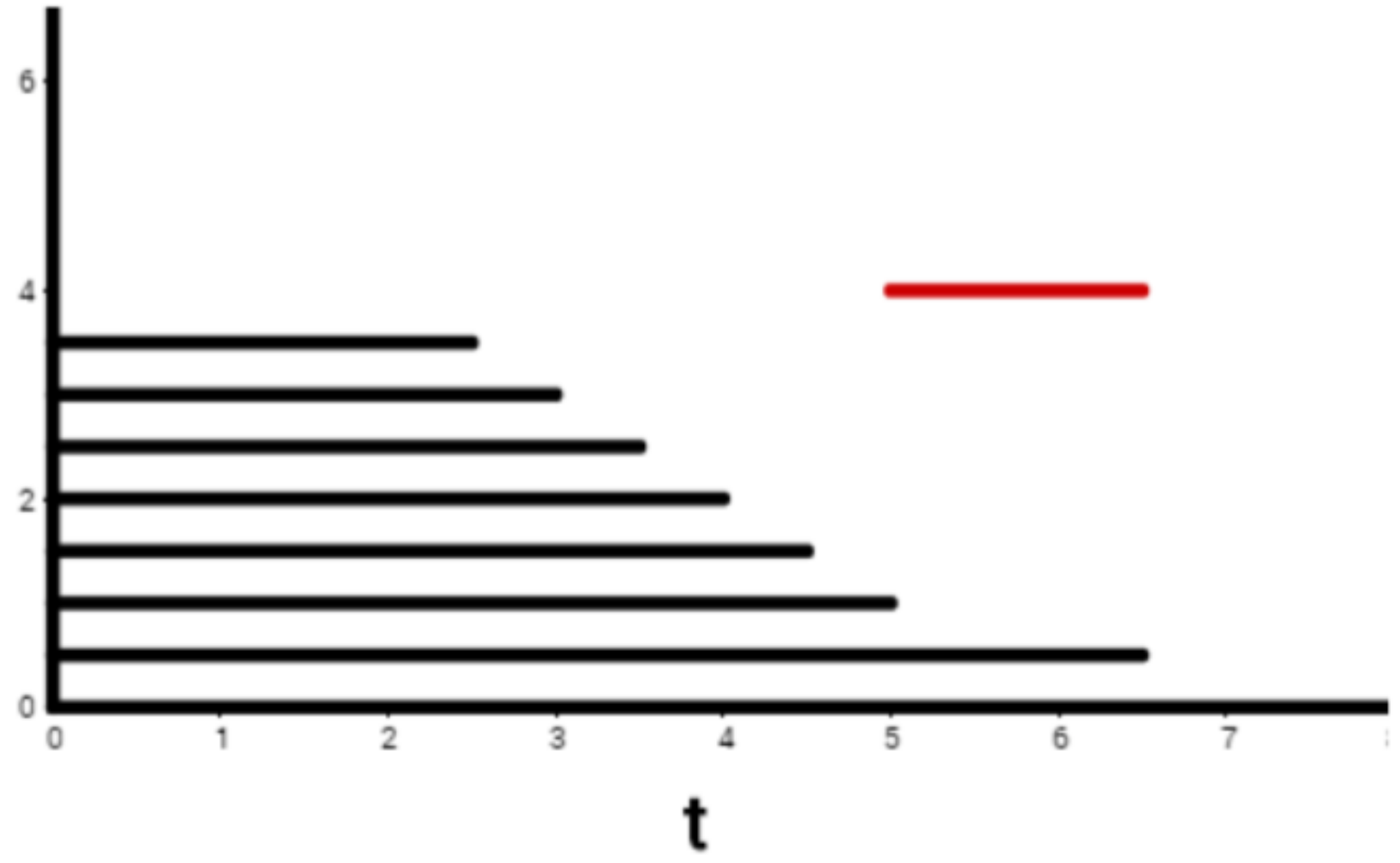


Figure: Barcode

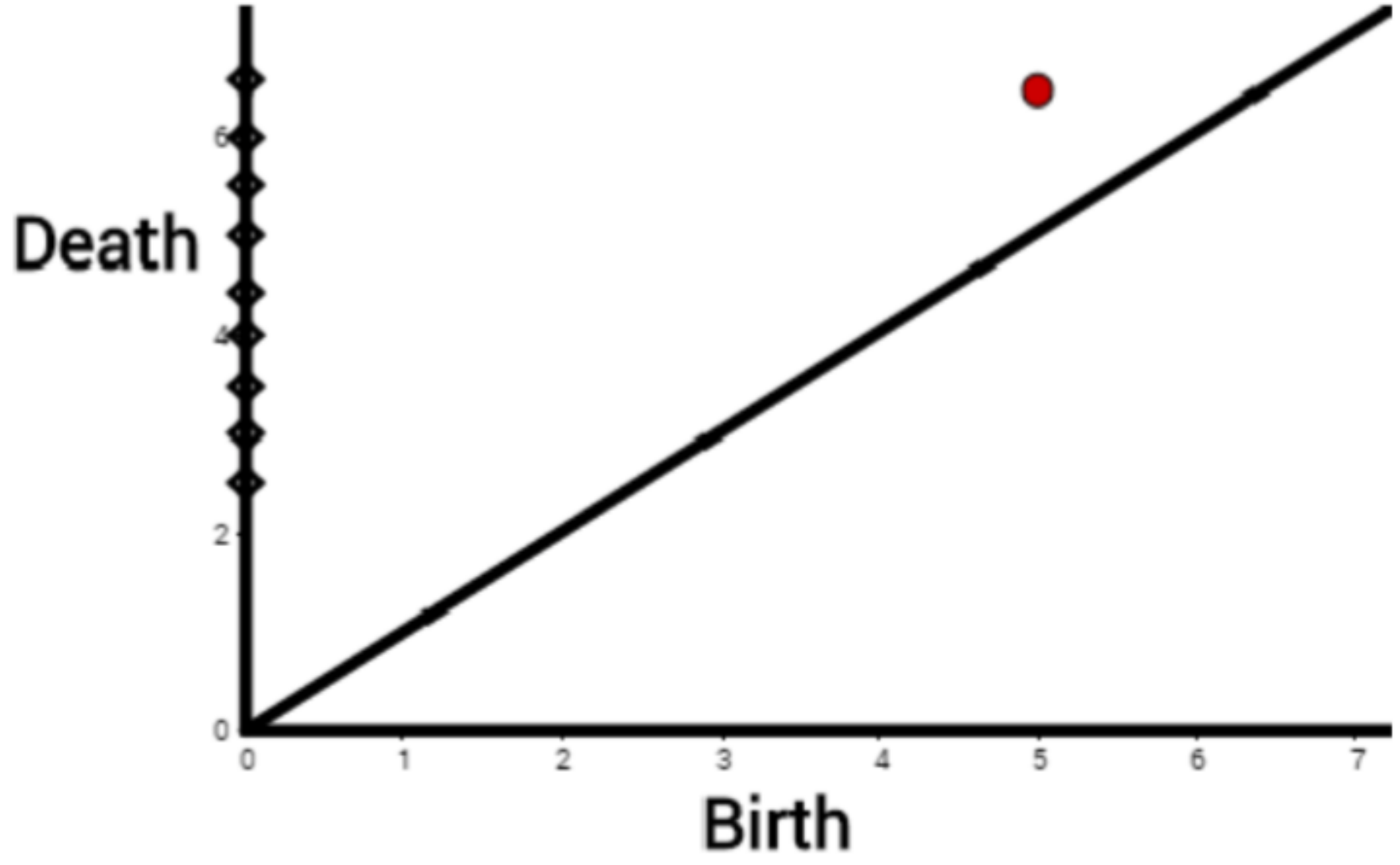


Figure: Persistence Diagram

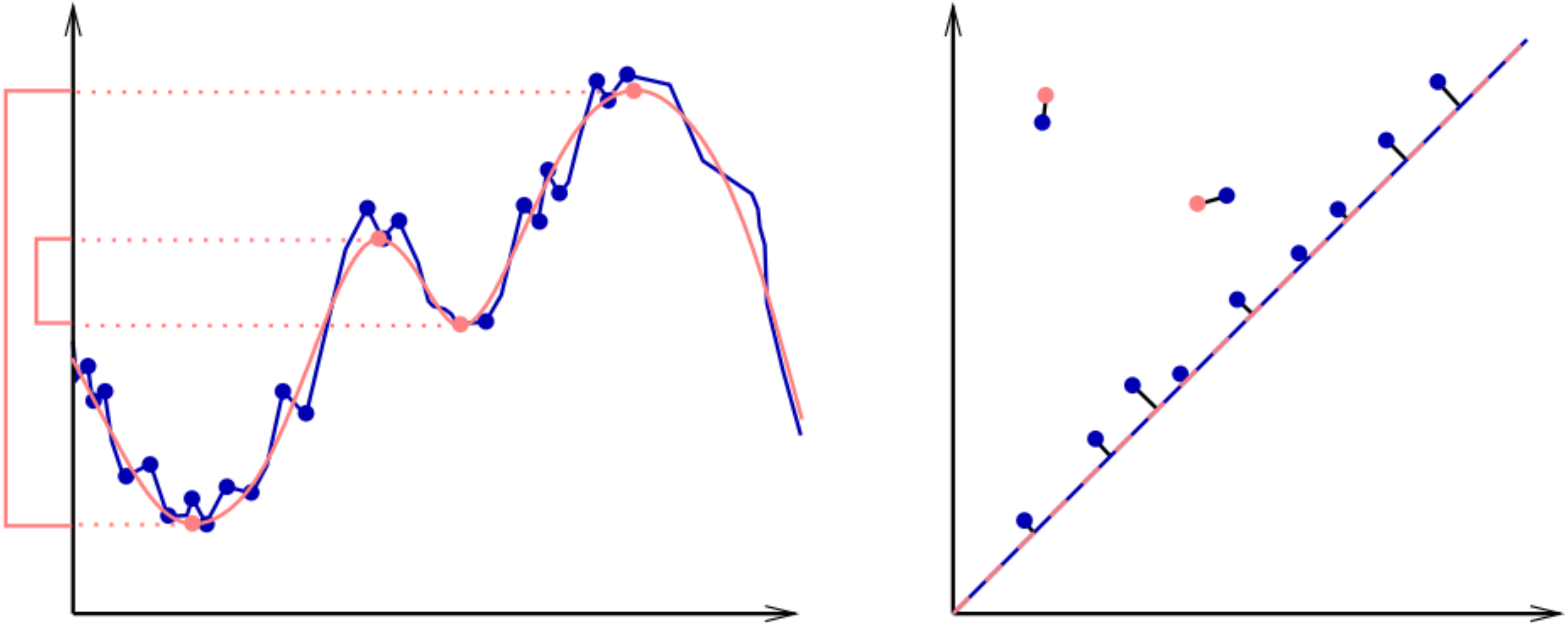


Figure 2: Left: two close functions, one with many and the other with just four critical values. Right: the persistence diagrams of the two functions, and the bijection between them.

https://users.cs.duke.edu/~edels/Papers/2007-J-01-StabilityPersistenceDiagrams.pdf

Bottleneck Distance

We need some definitions. For points $p = (p_1, p_2)$ and $q = (q_1, q_2)$ in $\bar{\mathbb{R}}^2$, let $\|p - q\|_\infty$ be the maximum of $|p_1 - q_1|$ and $|p_2 - q_2|$. Similarly for functions f and g, let $\|f - g\|_\infty = \sup_x |f(x) - g(x)|$. Let X and Y be multisets of points.

DEFINITION. The *Hausdorff distance* and the *bottleneck distance* between X and Y are

$$d_H(X, Y) = \max\{\sup_x \inf_y \|x - y\|_\infty, \sup_y \inf_x \|y - x\|_\infty\}$$

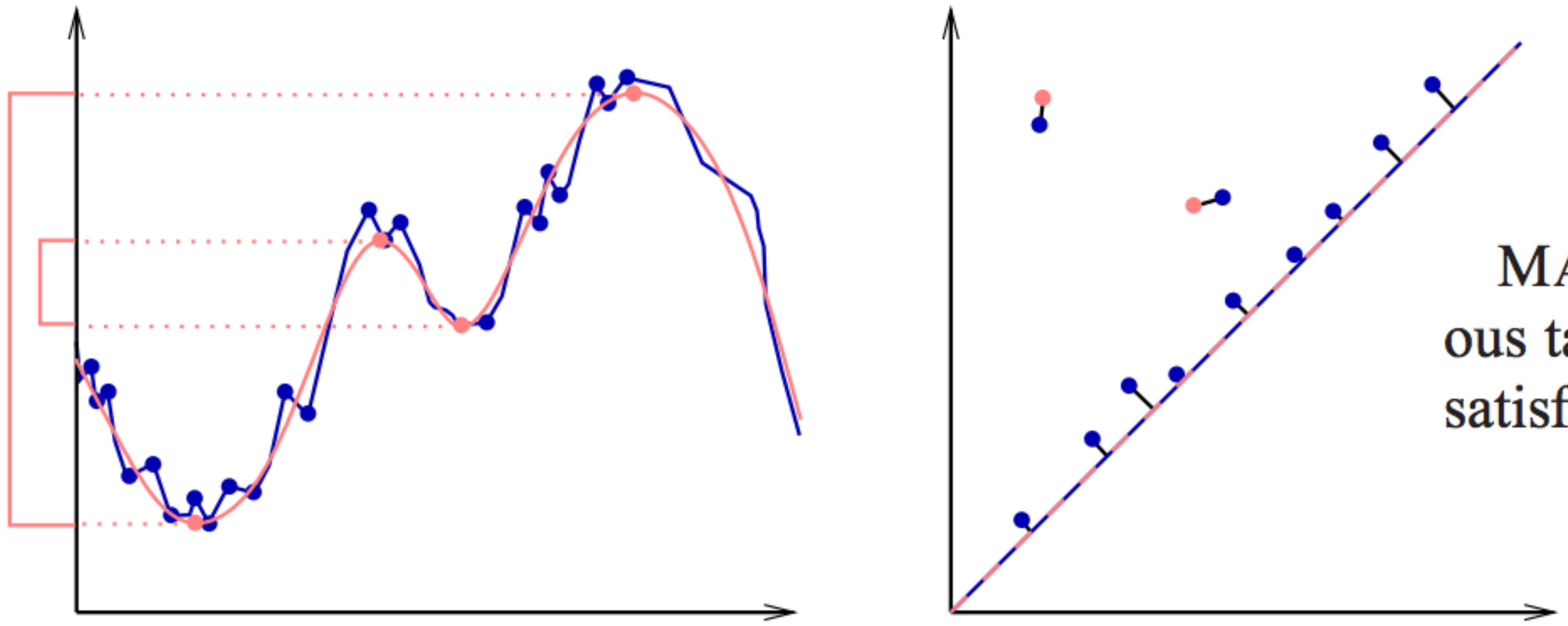
$$d_B(X, Y) = \inf_\gamma \sup_x \|x - \gamma(x)\|_\infty,$$

where $x \in X$ and $y \in Y$ range over all points and γ ranges over all bijections from X to Y. Here we interpret each point with multiplicity k as k individual points and the bijection is between the resulting sets.

https://users.cs.duke.edu/~edels/Papers/2007-J-01-StabilityPersistenceDiagrams.pdf

Bottleneck Distance Stability

$$d_B(X,Y) = \inf_{\gamma} \sup_{x} \|x - \gamma(x)\|_\infty$$

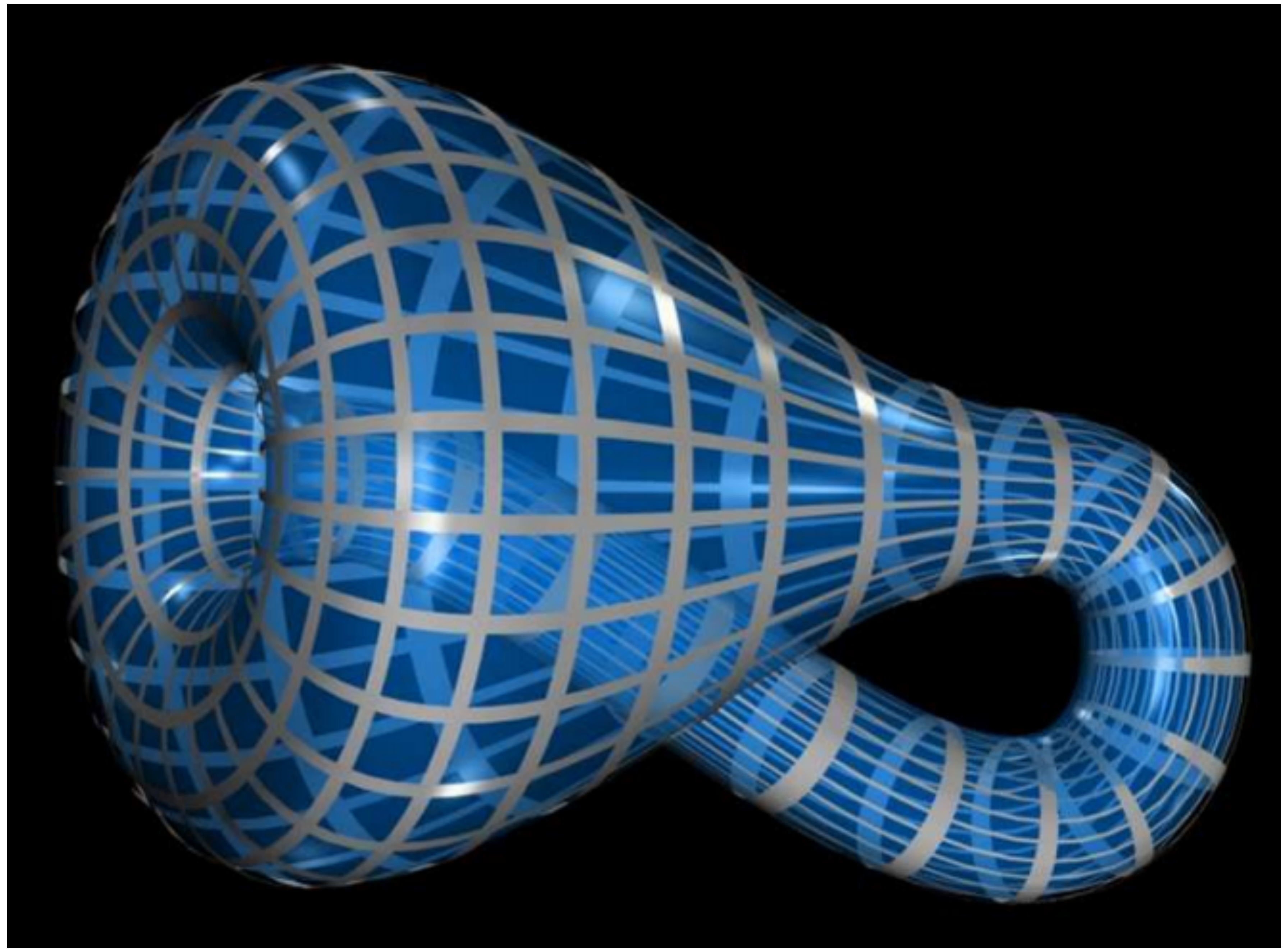


MAIN THEOREM. Let $\mathbb{X}$ be a triangulable space with continuous tame functions $f, g : \mathbb{X} \to \mathbb{R}$. Then the persistence diagrams satisfy $d_B(\mathrm{D}(f), \mathrm{D}(g)) \leq \|f - g\|_\infty$.

$$d_B(\mathrm{D}(f), \mathrm{D}(g)) \leq \|f - g\|_\infty$$

Cohen-SteinerEdelsbrunnerHarer2005

https://users.cs.duke.edu/~edels/Papers/2007-J-01-StabilityPersistenceDiagrams.pdf

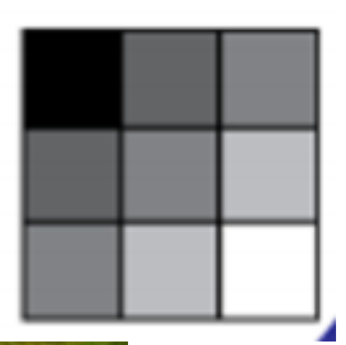


Space of Natural Images

[CarlssonnIshkhanovSilva2008]

http://webee.technion.ac.il/Sites/People/adler/Boris-topology.pdf



Taking 3 by 3 patches


Highlight

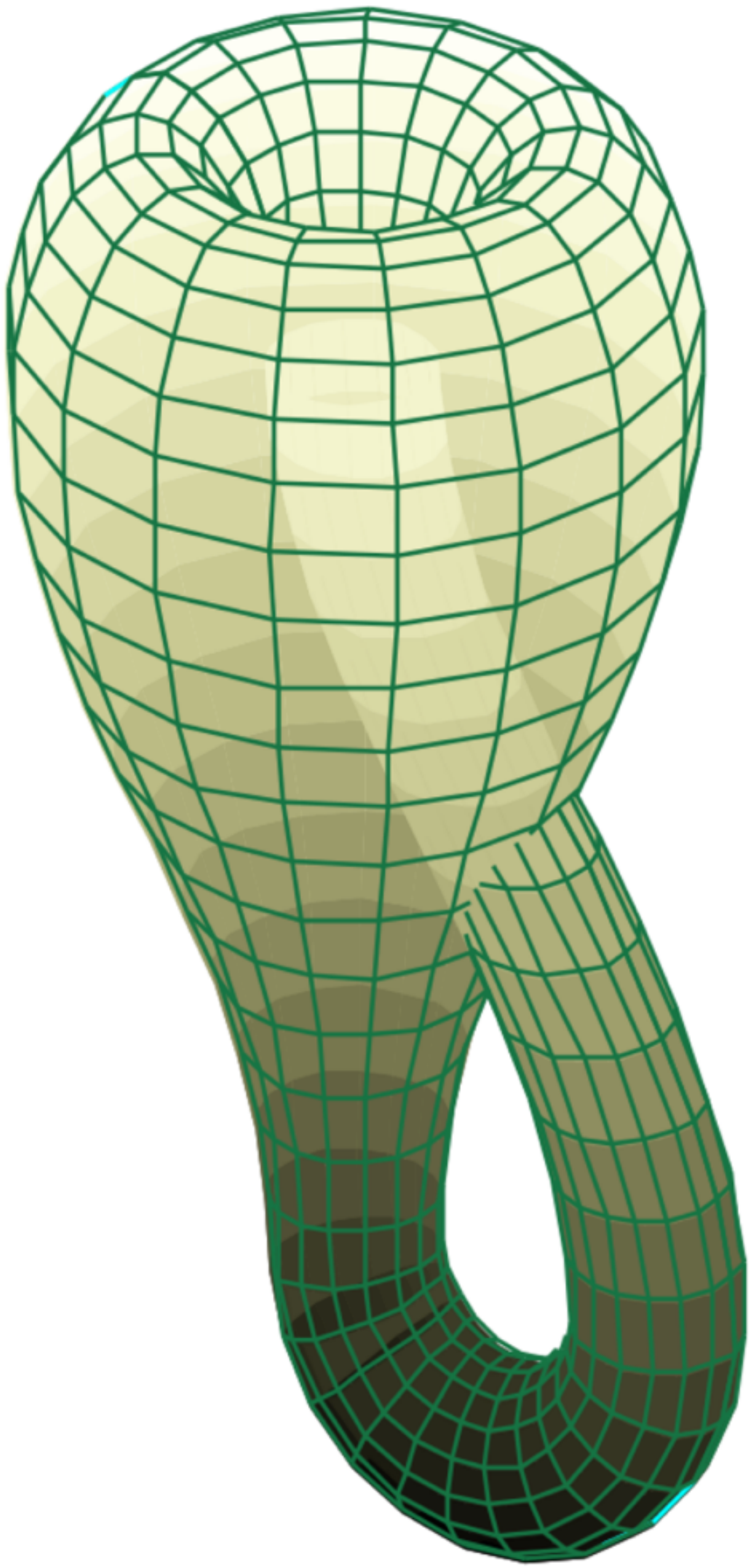
In this paper we continued the qualitative (topological) approach to the study of the space of 3 by 3 patches coming from natural images initiated in [8]. Perhaps the key advantage of this approach is that it allows one to find highly non-linear yet extremely important subsets within the data which otherwise would be very hard to discover using more common statistical techniques.

a collection of $4 \cdot 10^6$ 3 by 3 patches

http://webee.technion.ac.il/Sites/People/adler/Boris-topology.pdf [CarlssonnIshkhanovSilva2008]

High-level techniques

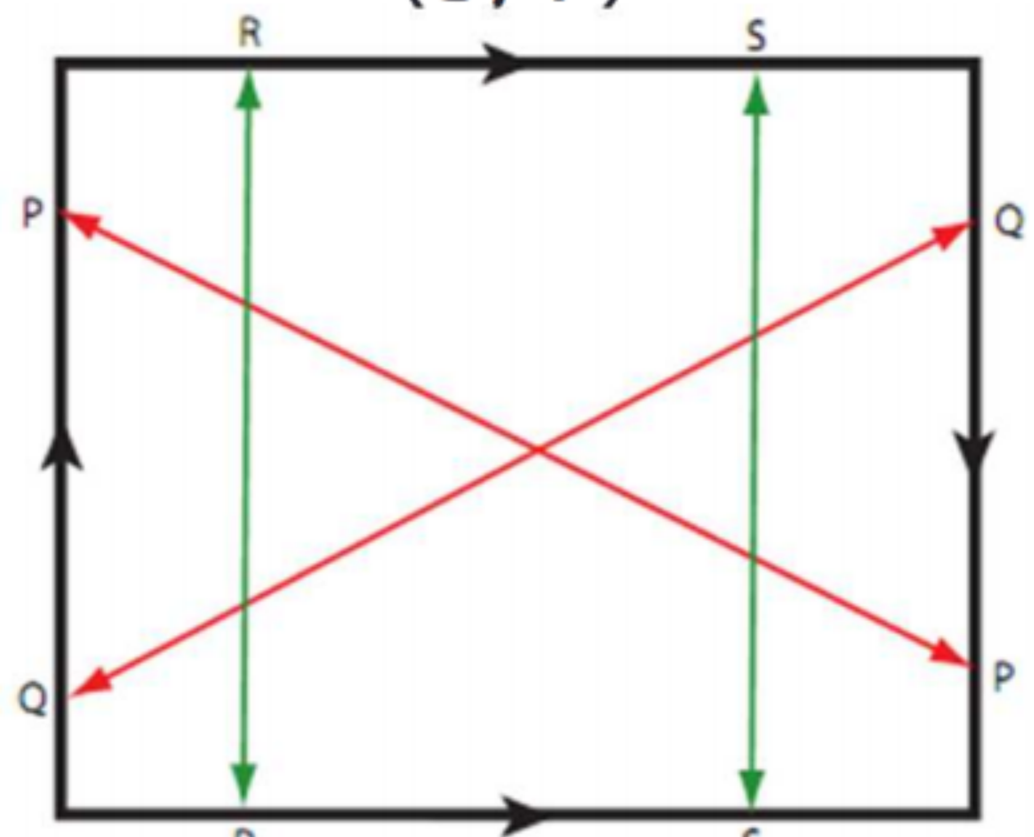
- Persistent homology
- The space of patches: preprocessing
- Klein bottle



Klein Bottle

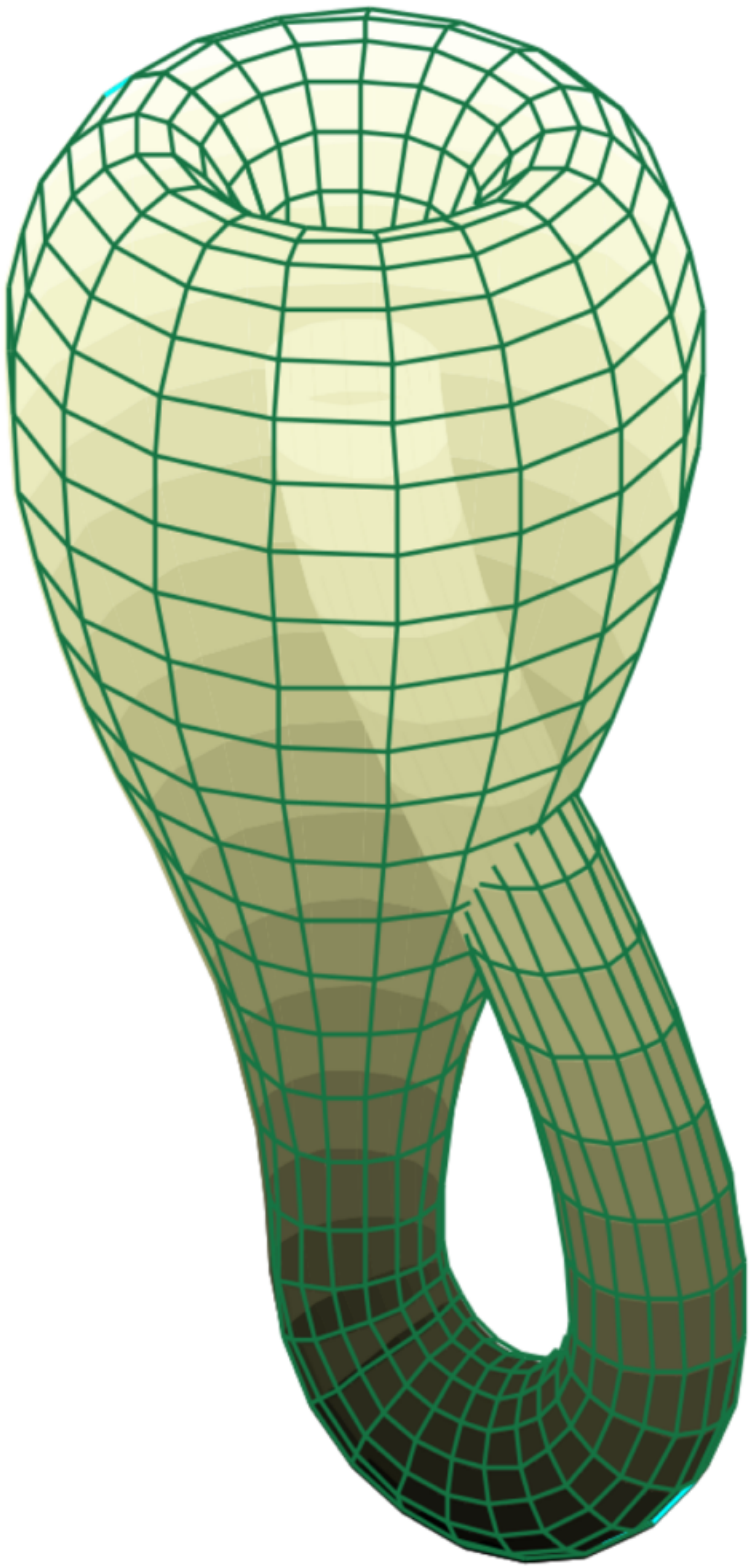
- Created by identifying points of a unit square:

$$(-1, t) \Leftrightarrow (1, -t)$$
$$(s, -1) \Leftrightarrow (s, 1)$$



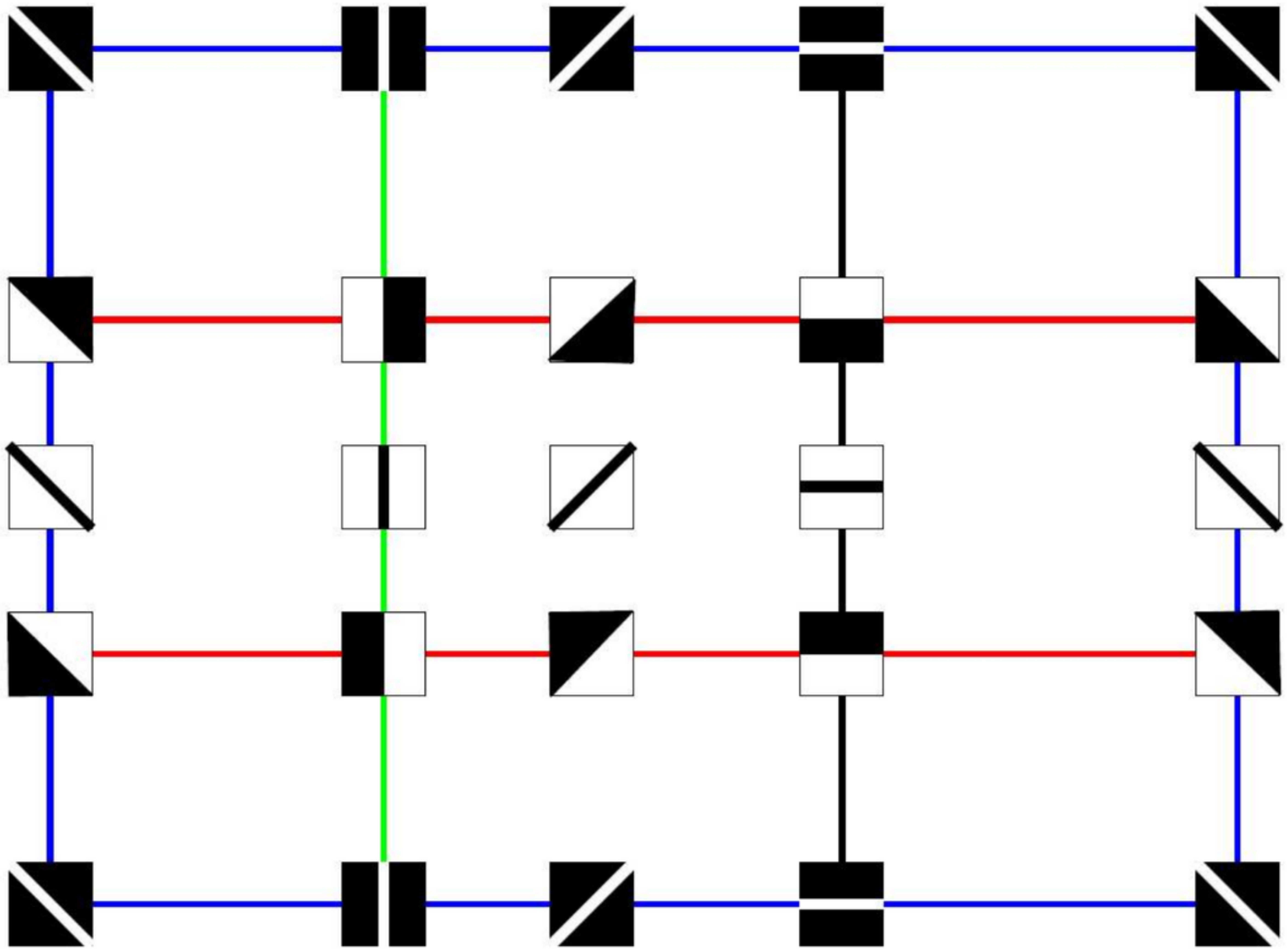
https://en.wikipedia.org/wiki/Klein_bottle http://webee.technion.ac.il/Sites/People/adler/Boris-topology.pdf

Klein Bottle



https://en.wikipedia.org/wiki/Klein_bottle

https://www.youtube.com/watch?v=AAsICMPwGPY



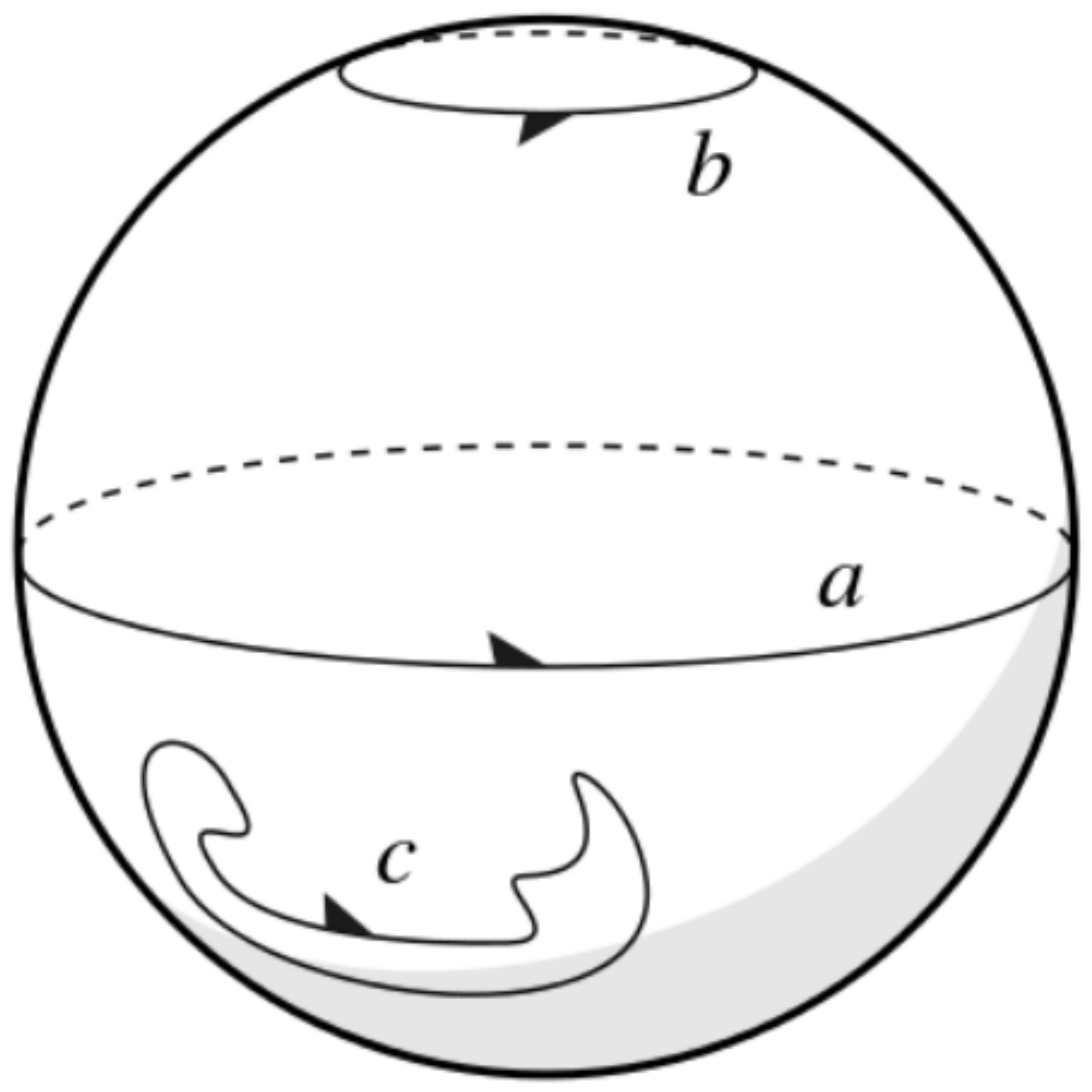
3 by 3 patches parametrized by the Klein bottle

[CarlssonnIshkhanovSilva2008]

How do we infer the Klein bottle?

- Via persistent homology…
- The (persistent) homology of the space of (many) images matches that of a Klein bottle under Z2 homology.

Back to homology

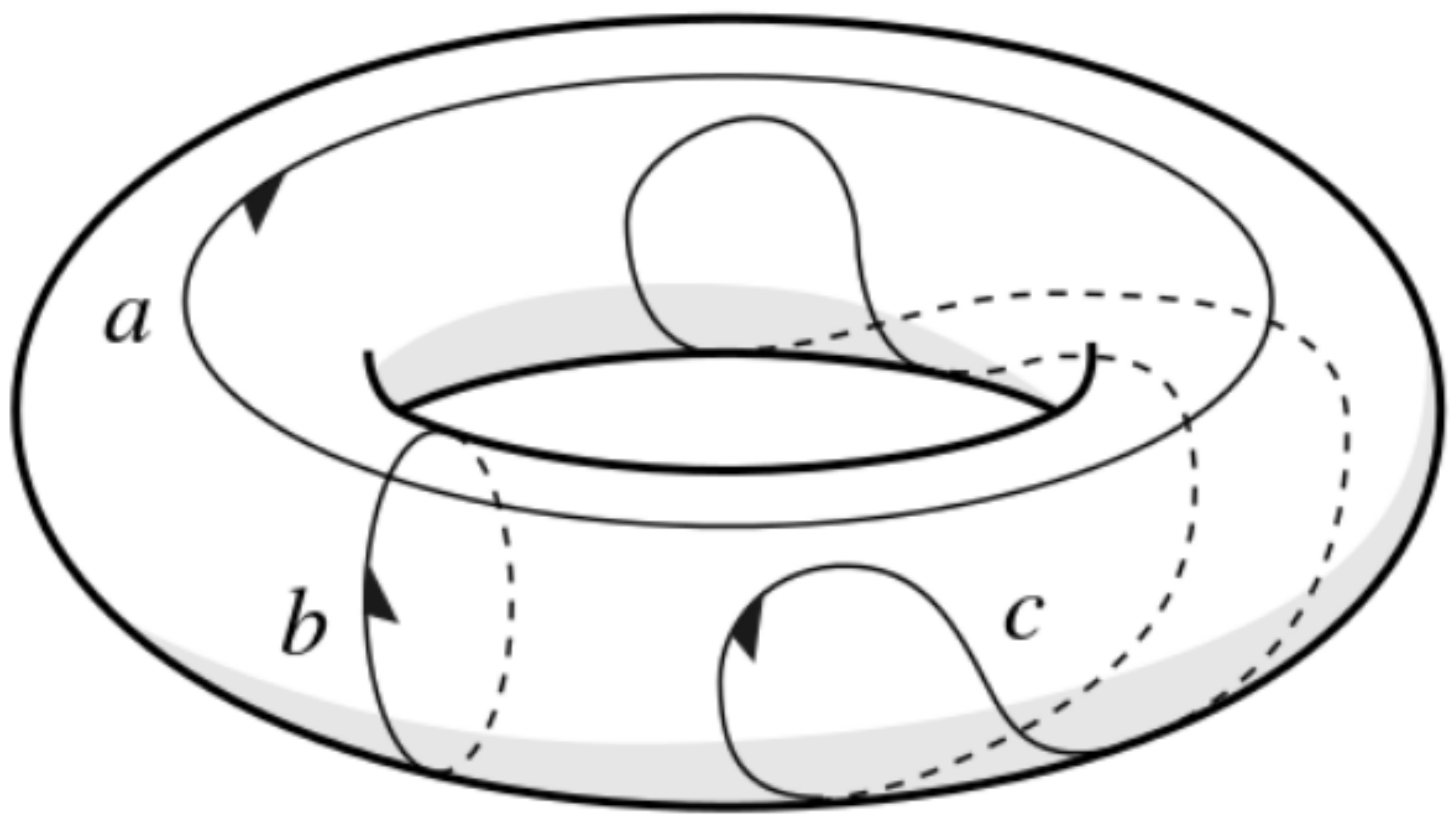


- Any cycle on the sphere can shrink to a point

b_0	b_1	b_2
1	0	1

https://en.wikipedia.org/wiki/Homology_(mathematics)

Back to homology

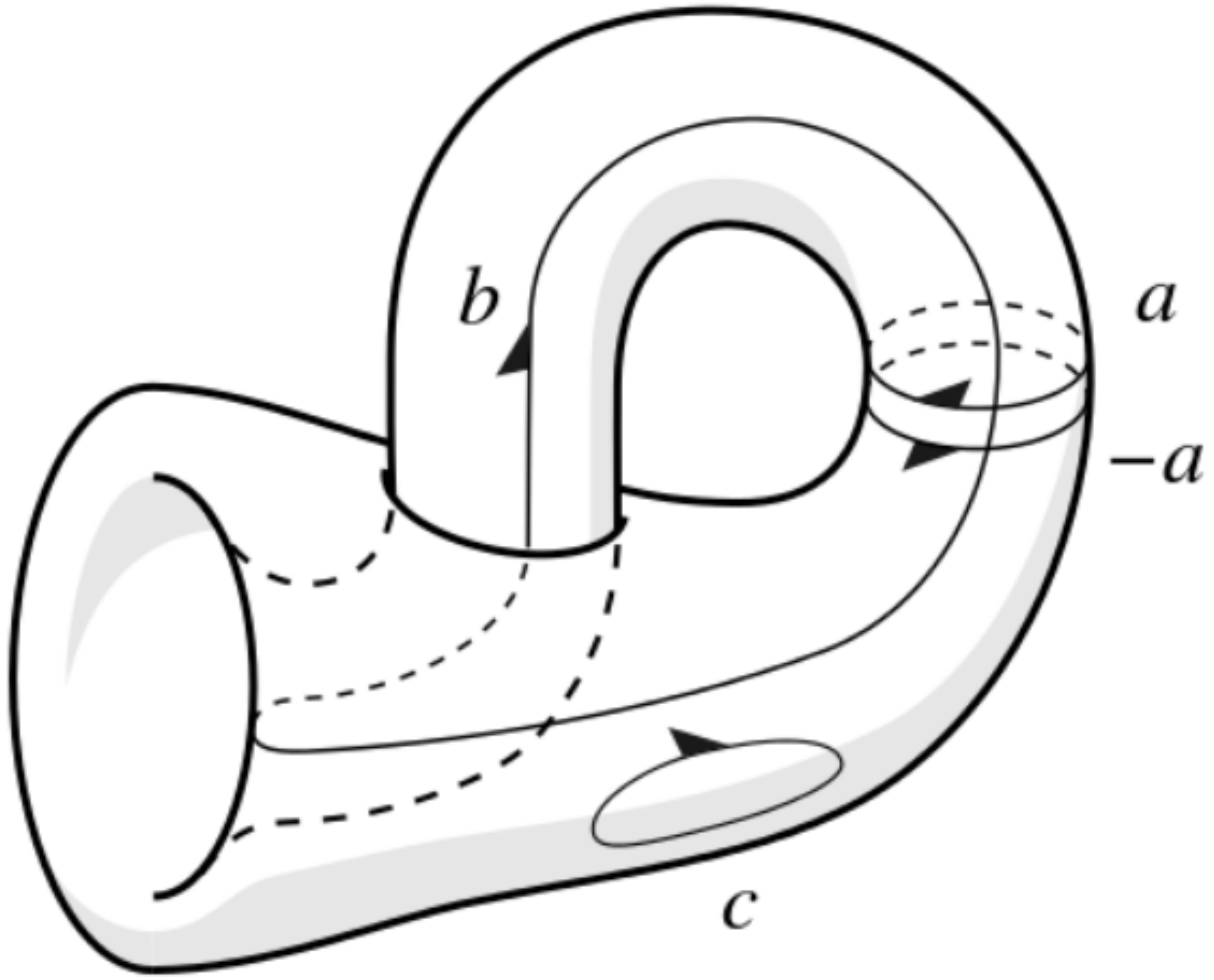


- Not all cycles on the torus can shrink to a point

b_0	b_1	b_2
1	2	1

https://en.wikipedia.org/wiki/Homology_(mathematics)

Back to homology



- Not all cycles on the Klein bottle can shrink to a point

b_0	b_1	b_2
1	1	0

- Under Z2 homology, the rank is identical to that of a torus

https://en.wikipedia.org/wiki/Homology_(mathematics)

Persistent homology

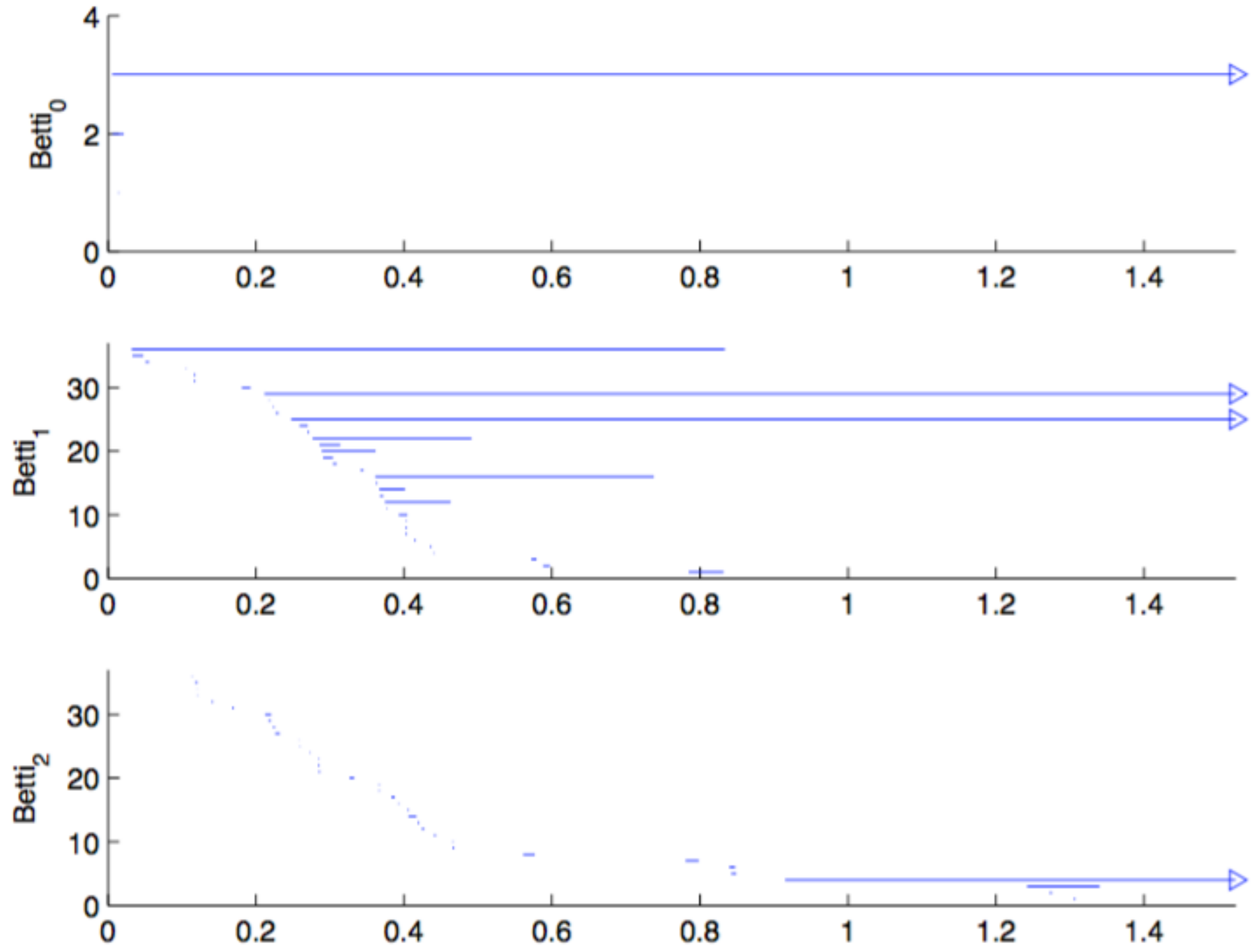


Figure 9: $PLEX$ results for $X(100, 10)$ in $\mathcal{M}$

Application

- Image compression: replace the high-contrast patches of the given image by the points on the surface of the embedded Klein bottle that best approximate them.

http://webee.technion.ac.il/Sites/People/adler/Boris-topology.pdf

Application: Texture Representation

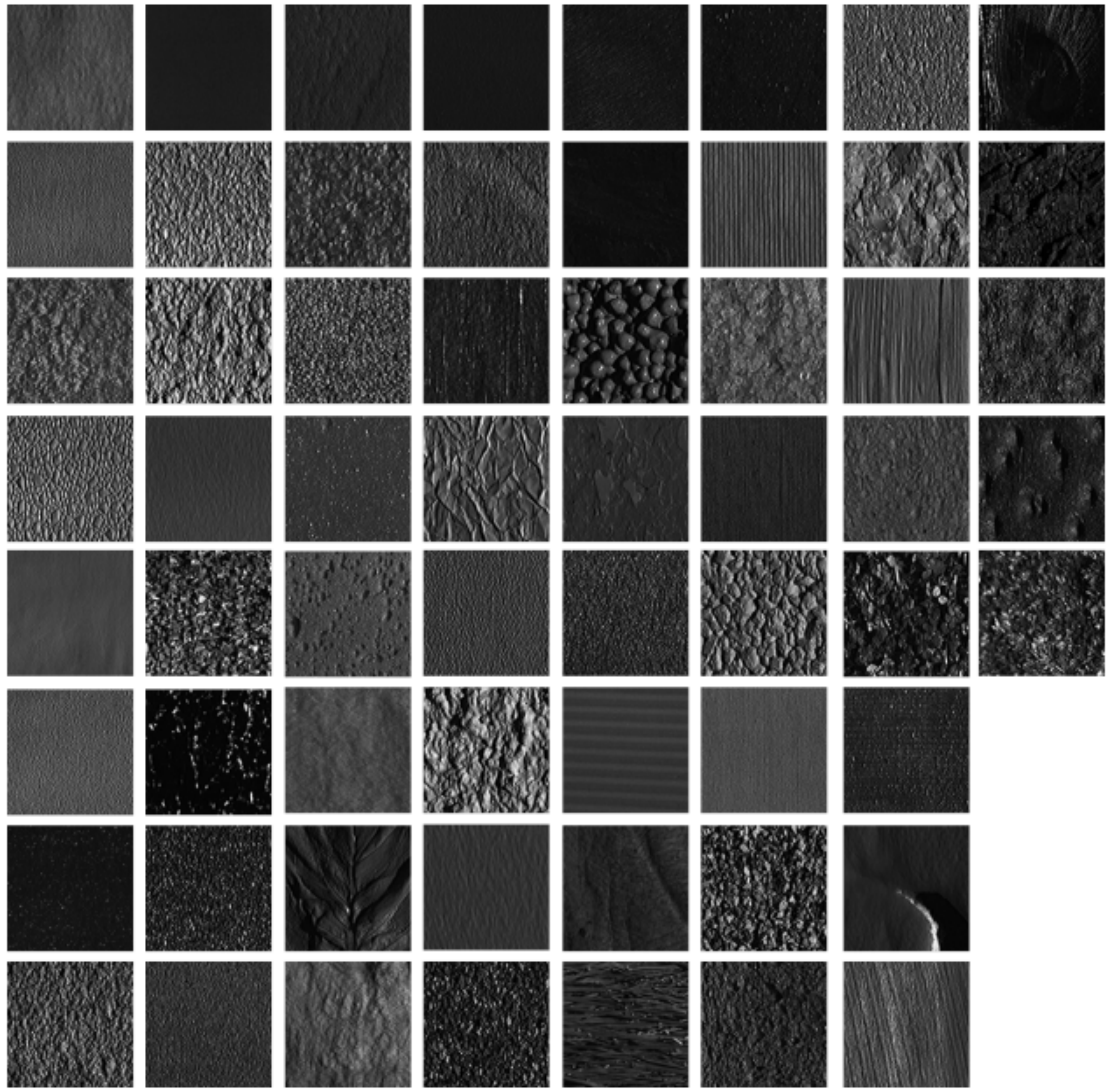


Figure 12: The 61 materials in the CUReT data set

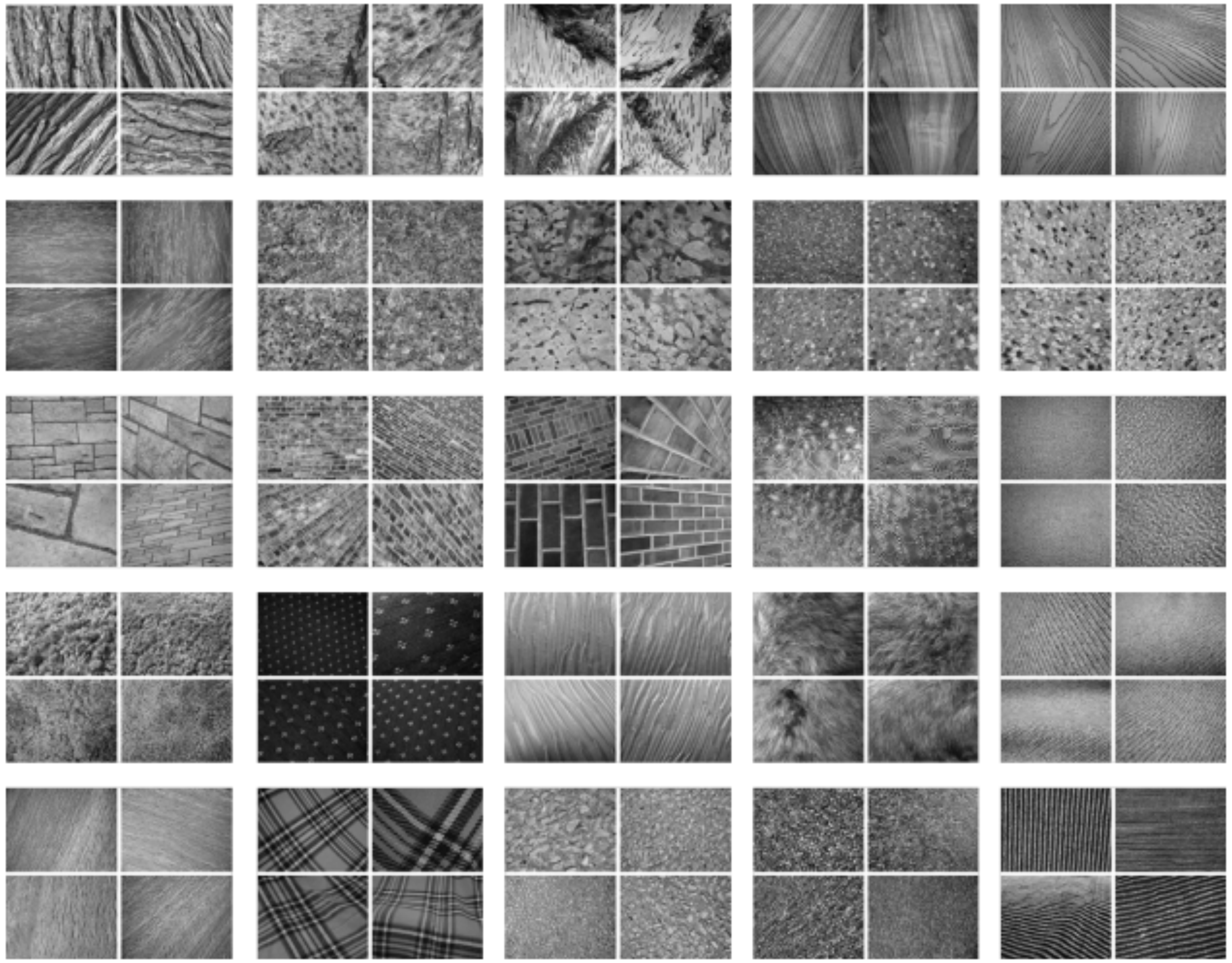
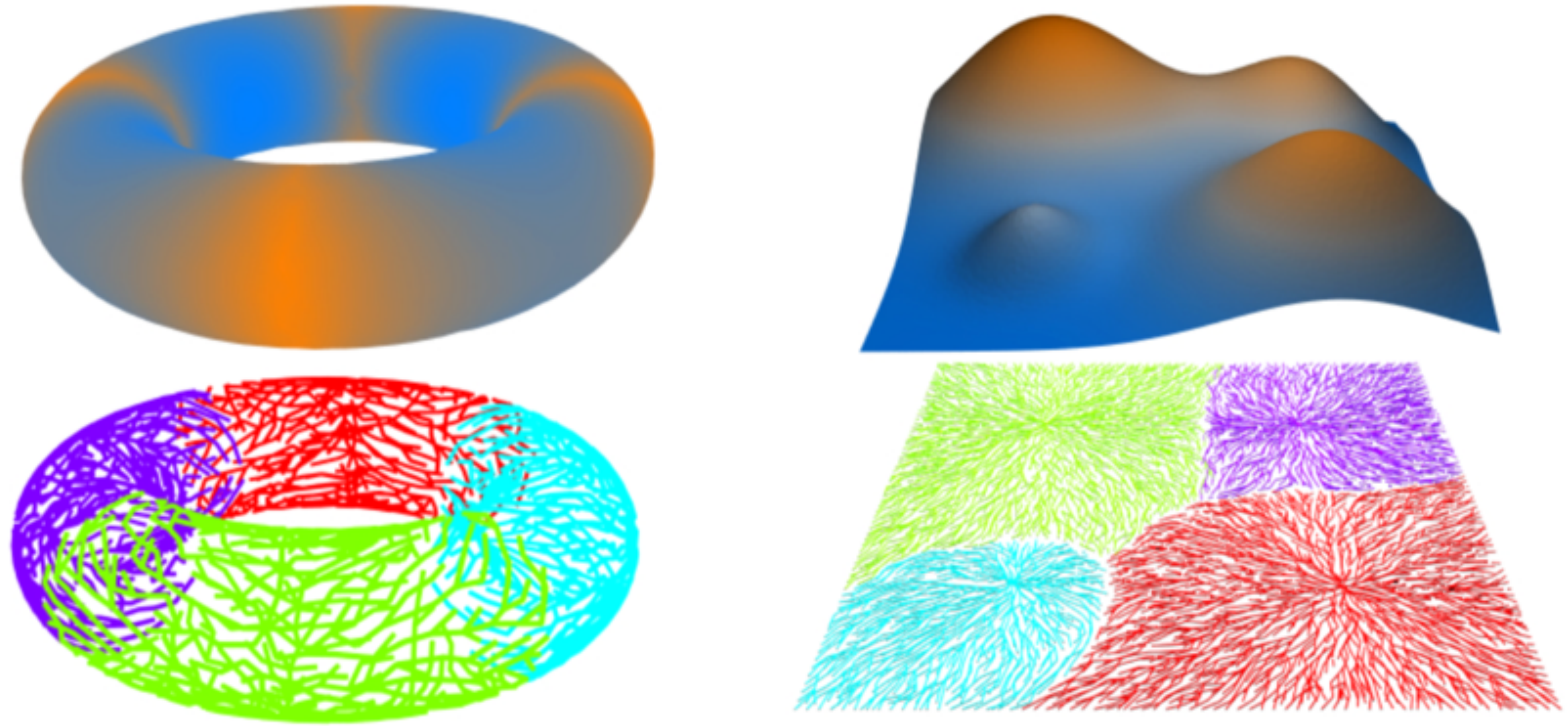


Figure 14: Four exemplars from each of the twenty five classes in the UIUC Texture data set.

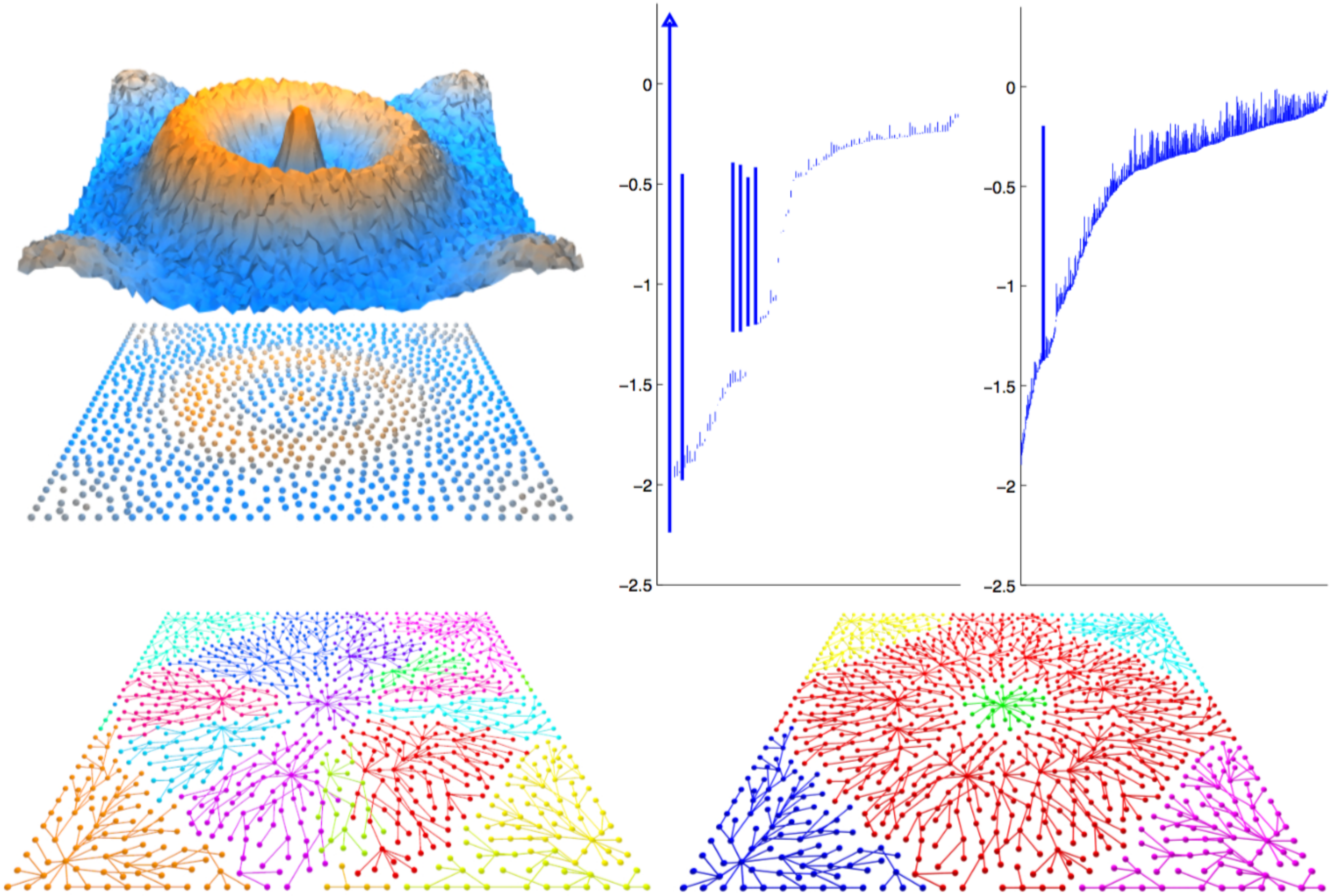
https://fds.duke.edu/db/attachment/2638



Analysis of Scalar Fields over Point Cloud

[ChazalGuibasOudot2008]

http://ailab.ijs.si/primoz_skraba/papers/sf_SODA.pdf



[ChazalGuibasOudot2008]

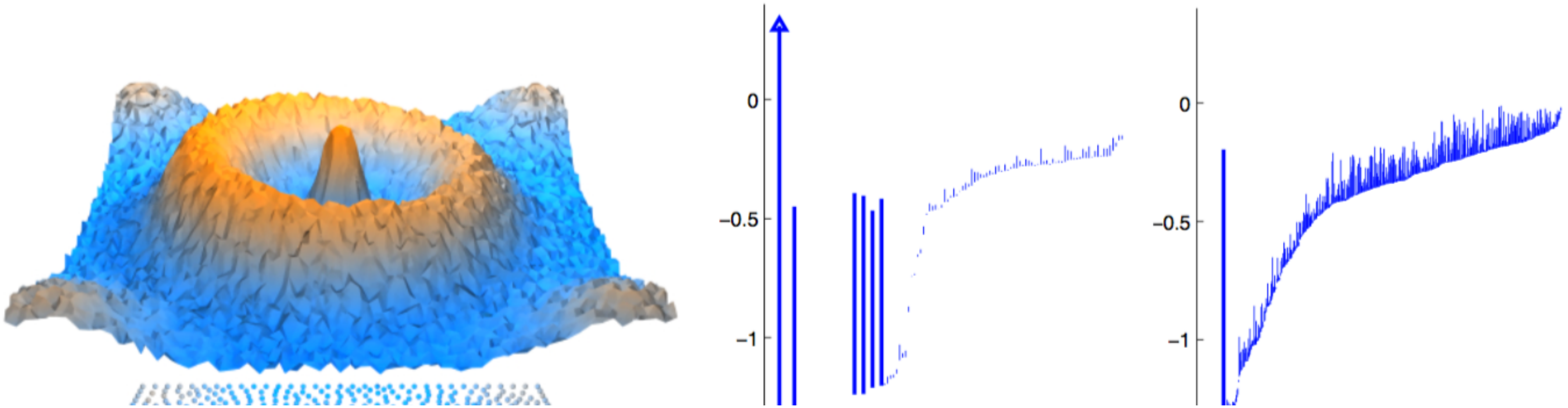
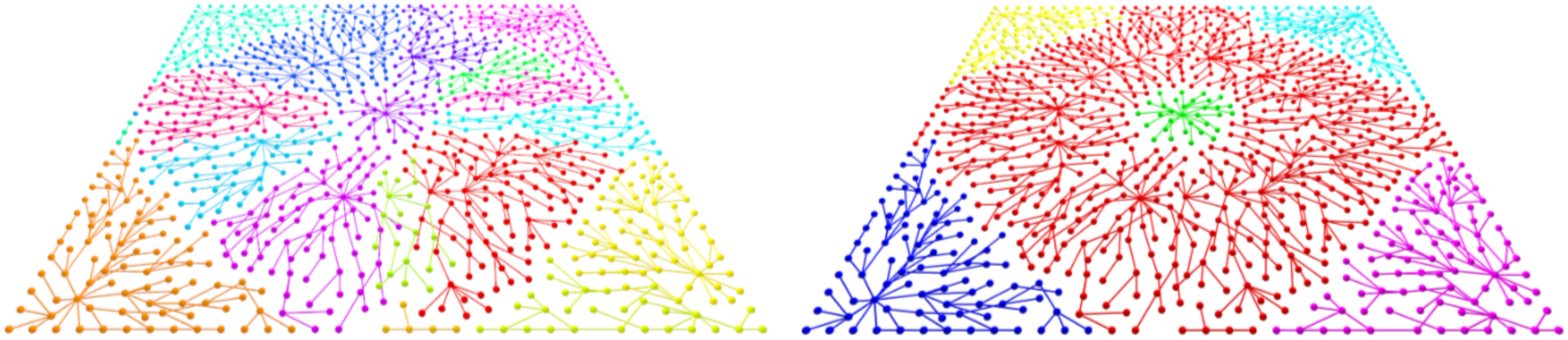


Figure 1: *Top row, left: a noisy scalar field f defined over a sampled planar square domain $\mathbb{X}$; center and right: approximations of the 0- and 1-dimensional persistence barcodes of $(-f)$ generated by our method from the values of f at the sample points and from their approximate pairwise geodesic distances in $\mathbb{X}$. The six long intervals in the 0-dimensional barcode correspond to the six prominent peaks of f (including the top of the crater), while the long interval in the 1-dimensional barcode reveals the ring shape of the basin of attraction of the top of the crater. Bottom row: approximate basins of attraction of the peaks of f, before (left) and after (right) merging non-persistent clusters, thus revealing the intuitive structure of f.*

[ChazalGuibasOudot2008]

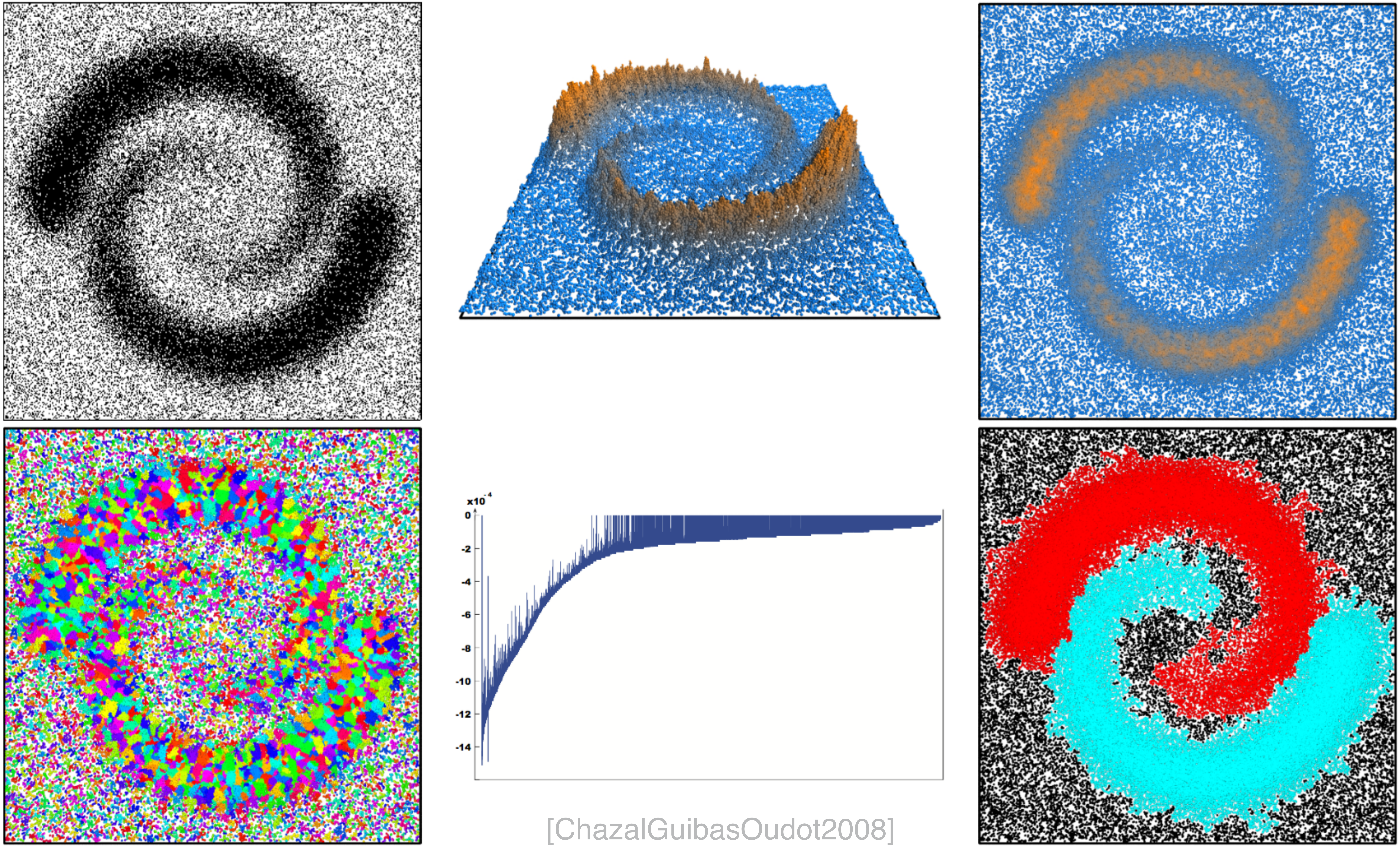
Figure 1: *Top row, left: a noisy scalar field f defined over a sampled planar square domain $\mathbb{X}$; center and right: approximations of the 0- and 1-dimensional persistence barcodes of $(-f)$ generated by our method from the values of f at the sample points and from their approximate pairwise geodesic distances in $\mathbb{X}$. The six long intervals in the 0-dimensional barcode correspond to the six prominent peaks of f (including the top of the crater), while the long interval in the 1-dimensional barcode reveals the ring shape of the basin of attraction of the top of the crater. Bottom row: approximate basins of attraction of the peaks of f, before (left) and after (right) merging non-persistent clusters, thus revealing the intuitive structure of f.*



[ChazalGuibasOudot2008]

High-level techniques

- Persistent homology
- Merge low persistent clusters while maintain high persistent clusters



[ChazalGuibasOudot2008]

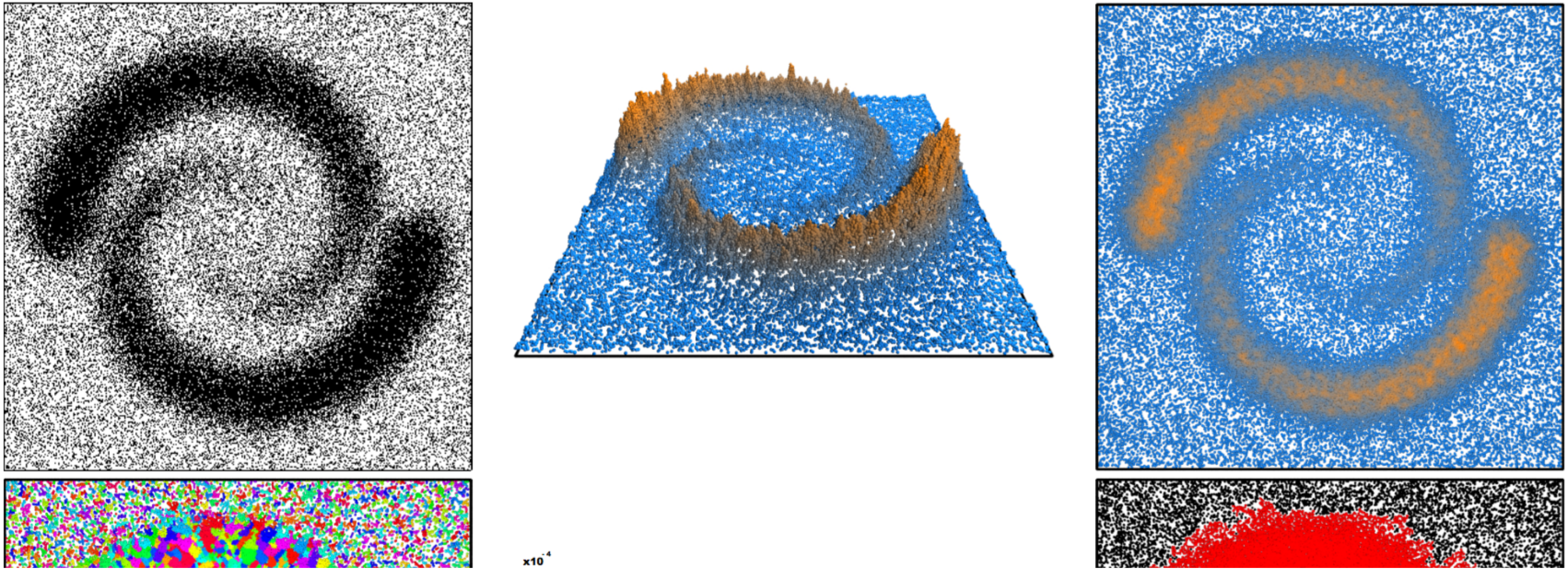


Figure 2: *A result in clustering. The top row shows the input provided to the algorithm of Section 4.2: the data points (left), or rather their pairwise Euclidean distances, and the estimated density function f (center and right). The 3-d view of f illustrates how noisy this function can be in practice, thereby emphasizing the importance of our robustness result (Theorem 3.2). The bottom row shows the estimated basins of attraction of the peaks of f, before (left) and after (right) merging non-persistent clusters. The 0-dimensional persistence barcode of $(-f)$ (center) contains two prominent intervals corresponding to the two main clusters. Since the estimated density is everywhere non-negative, the barcode has been thresholded at 0. Thus, intervals reaching 0 correspond to independent connected components in the Rips graph. Among those, the ones that appear lately are treated as noise and their basins of attraction shown in black, since their corresponding density peaks are low.*

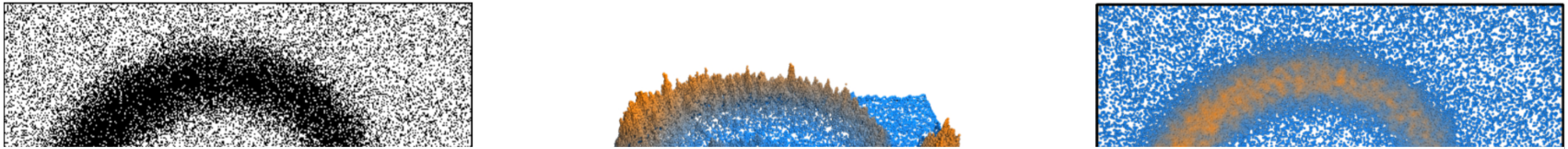
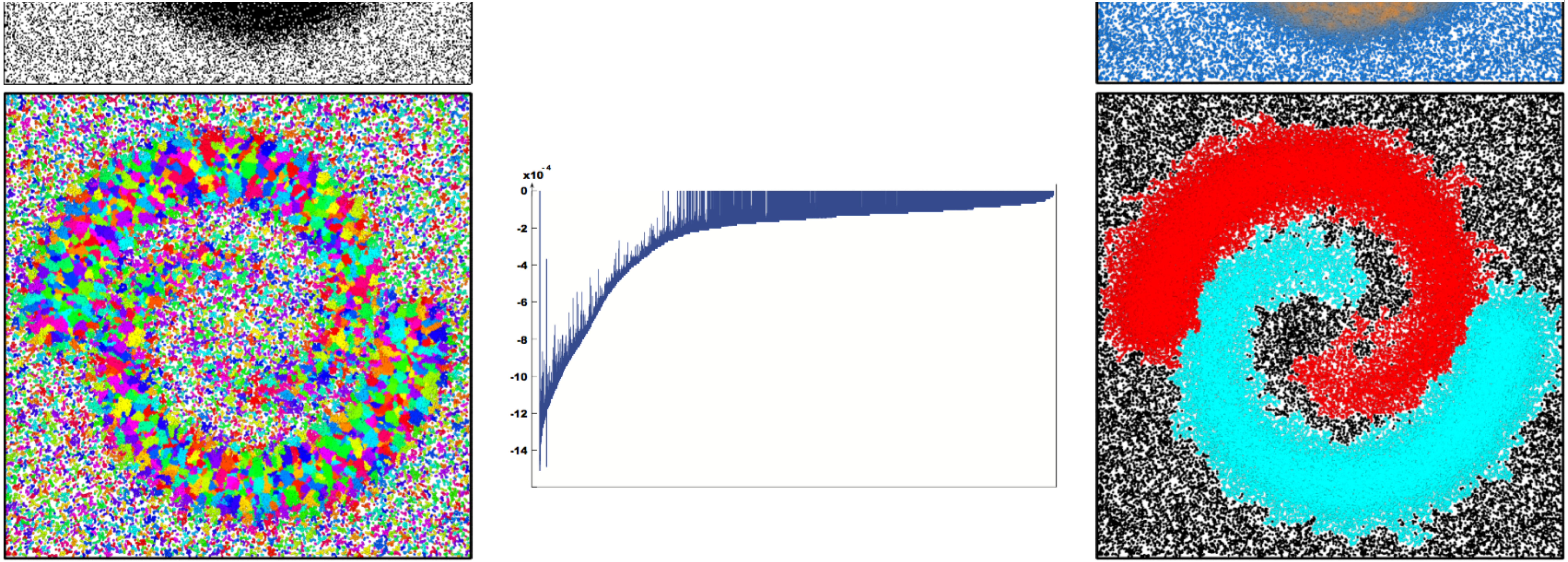


Figure 2: *A result in clustering. The top row shows the input provided to the algorithm of Section 4.2: the data points (left), or rather their pairwise Euclidean distances, and the estimated density function f (center and right). The 3-d view of f illustrates how noisy this function can be in practice, thereby emphasizing the importance of our robustness result (Theorem 3.2). The bottom row shows the estimated basins of attraction of the peaks of f, before (left) and after (right) merging non-persistent clusters. The 0-dimensional persistence barcode of $(-f)$ (center) contains two prominent intervals corresponding to the two main clusters. Since the estimated density is everywhere non-negative, the barcode has been thresholded at 0. Thus, intervals reaching 0 correspond to independent connected components in the Rips graph. Among those, the ones that appear lately are treated as noise and their basins of attraction shown in black, since their corresponding density peaks are low.*



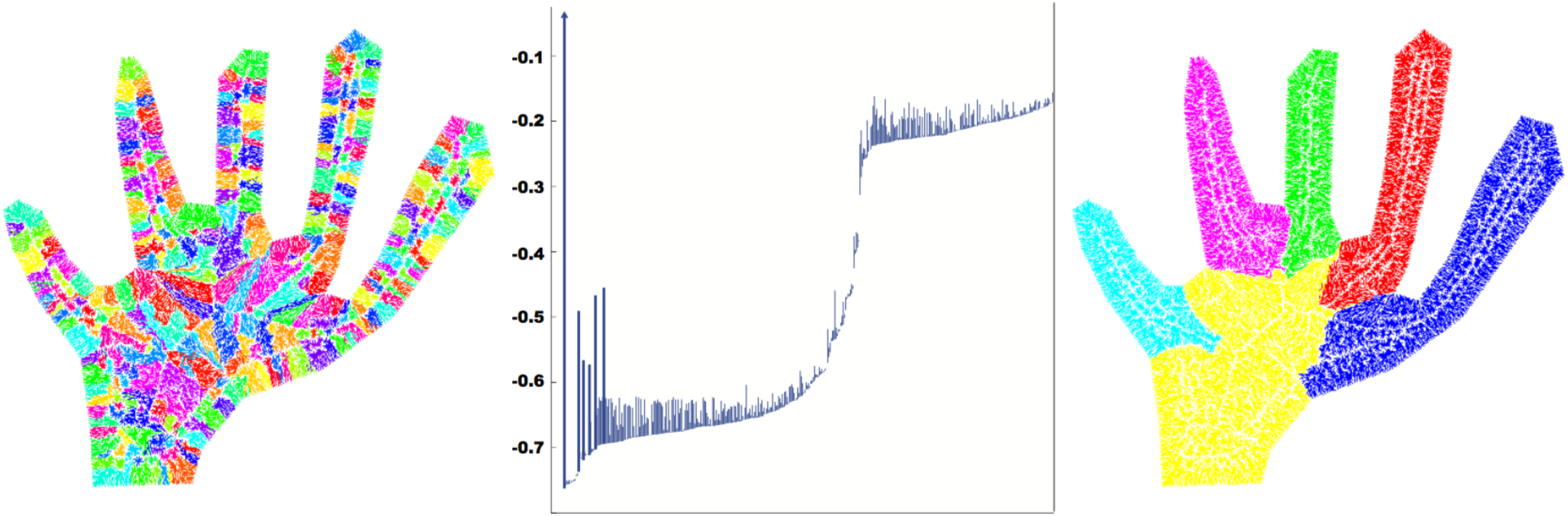


Figure 3: *Segmentation result on a sampled hand-shaped 2-D domain. The segmentation function is the (normalized) diameter of the set of nearest boundary points. The barcode shows six long intervals, corresponding to the palm of the hand and to the five fingers. The results before and after merging non-persistence clusters are shown respectively to the left and to the right of the barcode.*

[ChazalGuibasOudot2008]

Thanks!

Any questions?

You can find me at: **beiwang@sci.utah.edu**

CREDITS

Special thanks to all people who made and share these awesome resources for free:

- Presentation template designed by Slidesmash
- Photographs by unsplash.com and pexels.com
- Vector Icons by Matthew Skiles

Presentation Design

This presentation uses the following typographies and colors:

Free Fonts used:

http://www.1001fonts.com/oswald-font.html

https://www.fontsquirrel.com/fonts/open-sans

Colors used