Singular Zones in Multidimensional Data Sets

Talk given at
Algebra and Topology: Methods, Computation, and Science

Tomasz Kaczynski

Université de Sherbrooke
and Jagiellonian University

ATMCS4, Münster June 21, 2010

Outline

1. Motivation from Imaging
2. Topological approaches to criticality
3. Ideas from the Conley theory
4. Digital topology
5. Extraction of singular zones in image data
6. Experimentation

Motivation from Imaging

A 2D monochrome *digital image* is a function

$$f : X = [0, M-1] \times [0, N-1] \longrightarrow [0, L-1],$$

where all intervals are in $\mathbb{Z}$ and $L = 2^k$.
A *pixel* is a *space variable* (x, y) and its *gray level* $f(x, y)$.
$M \times N$ *space resolution*, L *dynamical range*, k # of *bits*.

We prefer to view pixel variables (x, y) as squares $[x, x+1] \times [y, y+1] \subset \mathbb{R}^2$ rather than isolated points.

3D images — space variable (x, y, z)
multidimensional images — visualization of multidim data
color digital images — vector-valued functions $f = [f_R, f_G, f_B]$

Image morphology

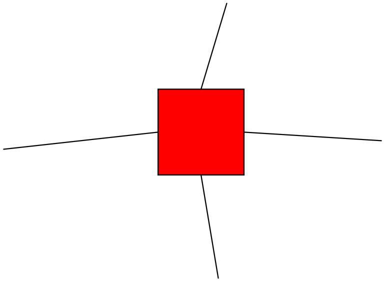
Image understanding and segmentation — construction of

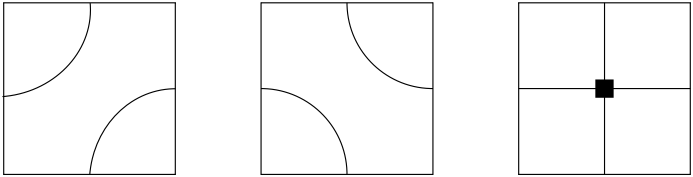
- isolines, contour lines, ...
- ridge lines, ravine lines;
- watersheds;
- $3D+$ analogies: level surfaces, border surfaces, ...

This is based on the knowledge of *critical points*:
summits, pits, saddles.

Saddles — especially important e.g. for isoline construction:
This is where level sets cross.

Do these lines cross?





Topological approaches to criticality

Assume for instance a continuous model:

$X = [-M, M]^d$ *rectangle* in $\mathbb{R}^d$, $f : X \to \mathbb{R}$ continuous.

Digital image context: $f : X \to [0, L-1]$, $f = 0$ close to ∂X (images put into black frame).

For $f \in C^1$, *critical points* are defined by $\nabla f = 0$.
Inflection points are *fragile*: They can be skipped by a discretization of f. Maxima, minima and saddles are *robust*.

We want a definition of *significant* or *robust* critical points.

Well explored idea

Given $\alpha \in \mathbb{R}$, look at the topology of the *sublevel* set

$$f \leq \alpha = \{q \in X \mid f(q) \leq \alpha\}.$$

$c \in \mathbb{R}$ is an *essential critical value* if the topology of the set $f \leq \alpha$ changes as α passes through c. What do we mean by "topology changes at c"?

- Classical criterion: $f \leq \beta$ is not a *deformation retract* of $f \leq \alpha$ for all sufficiently close $\beta < c < \alpha$, where ;
- Computational criterion: *relative homology groups* $H_*(f \leq \alpha; f \leq \beta) \neq 0$, $\beta < c < \alpha$;

Nice! What about *essential critical points* if $\operatorname{card} f^{-1}(c) > 1$?

"Typical" Morse theory assumptions:

Definition (Generic functions)

$f : X \to \mathbb{R}$ is *generic* or is a *Morse function* if

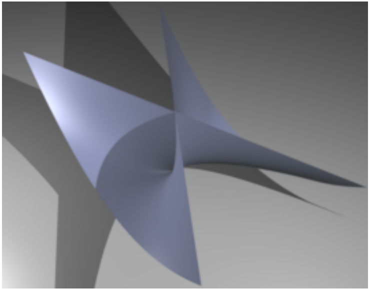
1. f is of class C^2;
2. All critical points of f are non-degenerate;
3. Distinct critical points correspond to distinct critical values.

Arguments making Morse theory suitable for computations

- Continuous function input: a small perturbation can make it generic;
- Discrete data input: a suitable interpolation can make it generic.

Problems

$f(x, y, z) = x^2 - y^2 z = 0$



- Removing degeneracy changes the “landscape”:
  - A visual perception of the generic image is not the same;
  - Isolines or isosurfaces disappear if they are at critical levels, for example this *Whitney umbrella* would be destroyed
- The generality of the topological definition of essential critical values is lost.

Our Goals

Develop

- a topological definition and
- a method for detection and classification

of essential critical points so that

- It works for degenerate critical points and sets of non-isolated critical points;
- It is local, that is, only based on the values of f in a neighborhood of p;
- It can be directly extended to discrete data.

A parallel

Forman's *discrete Morse theory*	versus	our approach
	is alike	
Computing homology up to an isomorphism	versus	Identification of homology generators
Classification Comparison	primary purpose	Extraction of structures for further processing

Weber & al. approach to degeneracy

G.H. Weber, G. Scheuermann, H. Hagen, B. Hamann
2000–2003 visualization papers

An overview

Given f and p, decide a small neighborhood N of p and consider:

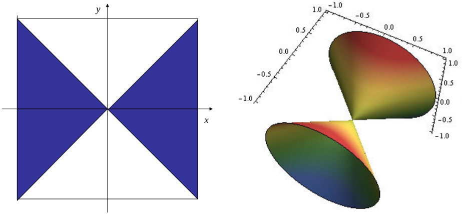
$$N_p = \{x \in U \mid f(x) > f(p)\},$$

$$N_n = \{x \in U \mid f(x) < f(p)\},$$

$$N_z = \{x \in U \setminus \{p\} \mid f(x) = f(p)\}.$$

n_p, n_n, n_z # connected components of N_p, N_n, resp N_z.

$n_p, n_n, n_z \mapsto$ classification of critical types in $\mathbb{R}^2$ and $\mathbb{R}^3$.



2D and 3D saddles

If p is not isolated, then

- Build a maximal connected region C of non regular points;
- Find a *classifying region* N of C;
- Calculate n_p and n_n for N;
- $n_p = n_n = 1 \implies C$ is a *regular region*;
- $n_p = 1$ and $n_n = 0 \implies C$ is a *minimum*;
- $n_p = 0$ and $n_n = 1 \implies C$ is a *maximum*;
- Otherwise, C is a *saddle region*.

Problem

The Weber & al. definitions are

- aimed at functionality in experimentation with discrete data but
- there may be "pathological" cases where finding a correct classifying region N may be impossible or hard. Namely:
- In dimensions $d \geq 3$, the numbers n_p, n_n, n_z may be insufficient to distinguish changes in topology: Homology groups H_k might have to be taken into account;
- Assuming that N is small enough may not be enough: A shape of it may require adjustment to flow lines of ∇f.

Example: Circle of saddles in 3D

Let X be a toroid in $\mathbb{R}^3$ obtained by revolving the disc D,

$$(x-2)^2 + z^2 \leq 1, \quad y = 0$$

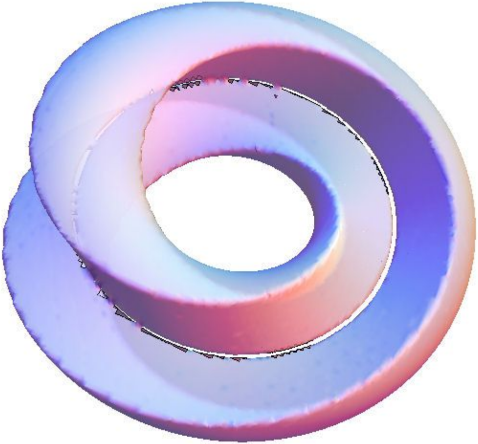
about the z-axis. Use *toric coordinates* (r, θ, ϕ):
(r, θ) polar coordinates for D,
ϕ the angle of rotation about the z-axis.

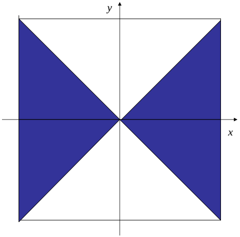
Define $f(r, \theta, \phi)$ by

$$f(r, \theta, \phi) = r^2 \cos^2(\theta - \phi/2) - r^2 \sin^2(\theta - \phi/2).$$

The circle S of radius 2 given by $r = 0$ is a saddle component. Its neighborhoods N form toroids. N_p and N_n are connected, due to winding, so $n_p = n_n = 1$, as for a regular point.

N_p winding about the saddle circle





Ideas from the Conley theory

Flow $(x, t) \mapsto \varphi(x, t)$ generated by $F = -\nabla f$:

$$\frac{\partial \varphi}{\partial x} = F(x).$$

Critical points of f are the equilibrium points of φ,
Non isolated critical points form critical regions: curves, surfaces, plateaus.
We may view them as *isolated invariant sets*:

A compact set N is an *isolating neighborhood* if $\mathrm{Inv}\, N \subset \overset{\circ}{N}$.
Then $S = \mathrm{Inv}\, N$ is an invariant set isolated by N.

Link Weber & al. $\leftrightarrow$ dynamics

Gradient case: N isolating neighborhood $\iff$ f has no critical points on ∂N.

$N_n \leftrightarrow$ the *exit set* L_n of N:
the subset of ∂N where F points *outward*.
$N_p \leftrightarrow$ the *entrance set* L_p of N i.e. F points *inward*.

Theorem ($\iff$ Ważewski Principle)

Suppose that L_n is closed in N. If

$$H_*(N, L_n) \neq 0$$

then N contains a non-empty invariant set S in its interior.

Conley Index $\leftrightarrow$ Morse index

Definition

Under the hypotheses of Ważewski on (N, L_n), the *homological Conley index* of a maximal invariant set S isolated by N is $CH(S) := H_*[N, L_n]$.

Theorem

Let $S = \{p\}$ *a non-degenerate critical point of* f, $m = \lambda(p)$ *be the Morse index of* p. *Then*

$$CH_i(\{p\}) = H_i(S^m, \{*\}) \cong \begin{cases} \mathbb{Z} & \text{if } i = m, \\ 0 & \text{otherwise.} \end{cases}$$

Example p a k-fold saddle in $\mathbb{R}^2 \implies CH_1(\{p\}) = \mathbb{Z}^k$.

Digital topology

The *unitary cubical grid* $\mathcal{K}$ in $\mathbb{R}^d$ consists of *elementary cubes*

$$Q = I_1 \times I_2 \times \cdots I_d$$

where $I_j = [k, k+1]$ or $I_j = \{k\}$ for some $k \in \mathbb{Z}$ (integers).
$\mathcal{K}^d \subset \mathcal{K}$ the set of d-dim cubes.

Given $\mathcal{X} \subset \mathcal{K}$, *carrier* of $\mathcal{X}$ is the *cubical set*

$$X = |\mathcal{X}| = \bigcup \mathcal{X} \subset \mathbb{R}^d.$$

We consider discrete functions

$$f : \mathcal{X}^d \to \mathbb{R}, \text{ where } \mathcal{X}^d \subset \mathcal{K}^d.$$

Goals

Main goal: Classify a d-cube Q by comparing $f(Q)$ with the values of f on the neighboring d-cubes.

First goal: Define *discrete isolating neighborhood* N of Q;
its *exit set* L_n;
and its *entrance set* L_p.

Postulate: $H_*(N, L_p)$ and $H * (N, L_n)$ should be computable.
In particular, N, L_p and L_n should be compact polyhedra.

Digital topology problem: No straightforward extension of the concept of *neighbor* and *boundary* in the digital setting

Digital boundaries

Let $Q \in \mathcal{K}^d$. The combinatorial *wrap* and *outer boundary* of Q are subsets of $\mathcal{K}^d$ defined by

$$wrap\,(Q) = \{P \in \mathcal{K}^d \mid Q \cap P \neq \emptyset\}.$$

and

$$bd(Q) = wrap\,(Q) \setminus Q.$$

Let $Q \in \mathcal{X}^d$ such that $wrap\,(Q) \subset \mathcal{X}^d$. The *upper wrap*, *lower wrap*, and *level wrap* of Q are

$$\overline{wrap}\,(Q) := \{P \in \ bd(Q) \mid f(P) > f(Q)\}\,.$$

$$\underline{wrap}\,(Q) := \{P \in \ bd(Q) \mid f(P) < f(Q)\}\,;$$

$$wrap_{\,z}(Q) := \{P \in \ bd(Q) \mid f(P) = f(Q)\}$$

We analogously define

$$\mathit{wrap}\,(\mathcal{A}) \;\text{ and }\; \mathit{bd}(\mathcal{A})$$

for any $\mathcal{A} \subset \mathcal{X}^d$.

The *upper, lower*, and *level wraps* of $\mathcal{A}$ are defined by local comparisons, e. g.

$$\overline{\mathit{wrap}}\,(\mathcal{A}) = \{P \in \; \mathit{bd}(A) \mid f(P) > f(Q) \text{ for all } Q \in \mathcal{A} \cap \mathit{bd}(P)\}\,.$$

Problem: These sets are disjoint but their supports are not!

2D — distinguishing 4– versus 8–connectivity solved it.
Higher dimensions — not suitable for computing homology.

Rescaling

Factor k rescaling isomorphism $\Lambda^k(x) := kx$.
k'th subdivision of $\mathcal{A} \subset \mathcal{K}^d$ is

$$\operatorname{sd}^k(\mathcal{A}) := \{P \in \Lambda^{1/k}(\mathcal{K}^d) \mid P \subset A\}$$

Factor k scaled wrap and , *outer boundary* of $\mathcal{A} \subset \mathcal{K}^d$ are

$$wrap^k(\mathcal{A}) = \{P \in \Lambda^{1/k}(\mathcal{K}^d) \mid P \cap A \neq \emptyset\},$$

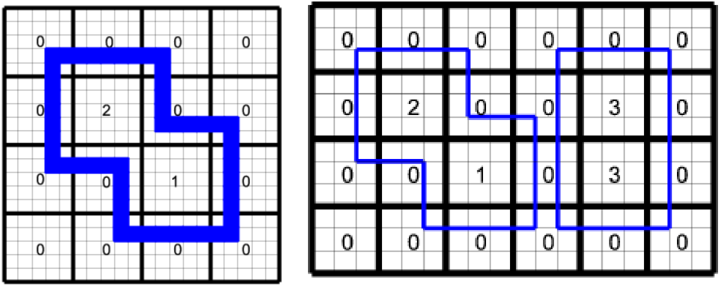
$$bd^k(\mathcal{A}) = wrap^k(\mathcal{A}) \setminus \operatorname{sd}^k(\mathcal{A}).$$

For the *scaled upper wrap*, we take values of f on unit cubes:

$$\overline{wrap}^k(A) := \{P \in bd^k(\mathcal{A}) \mid P \subset |\overline{wrap}(\mathcal{A})|\}.$$

Analogously, we define $\underline{wrap}^k(\mathcal{A})$ and $wrap_z^k(\mathcal{A})$.

The choice $k = 5$



Lemma

Given $\mathcal{A} \subset \mathcal{K}^d$ *and a scaling factor* $k \geq 5$*, the inclusion* $A \hookrightarrow |wrap^k(wrap^k(\mathcal{A}))|$ *induces an isomorphism in homology.*

Extraction of singular zones

Given $\mathcal{A} \in \mathcal{K}^d$, we use the neighborhood

$$N(\mathcal{A}) := |\mathit{wrap}^5(\mathit{wrap}^5(\mathcal{A}))|.$$

For the *exit set*, we take values of f on unit cubes:

$$L_n(\mathcal{A}) := |\underline{\mathit{wrap}}(\mathcal{A})| \cap |\mathit{bd}^5(\mathit{wrap}^5(\mathcal{A}))|.$$

Analogously, we define $L_p(\mathcal{A})$ and $L_z(\mathcal{A})$.
The elementary cube Q is *normal* if

$$H_*(N(\mathcal{A}), L_p(\mathcal{A})) = H_*(N(\mathcal{A}), L_n(\mathcal{A})) = 0,$$

where $\mathcal{A} = \mathit{wrap}(Q)$.
Otherwise, it is *singular*.

First classification algorithm

Algorithm (*Detecting singular cubes)*

```
For each elementary full cube $Q$
build $N, L_p, L_n,$ for wrap$(Q)$
$H := H_*(N, L_p) = 0$ and $H_*(N, L_n) = 0$
    if $H =$ TRUE then $Q$ is normal
        else $Q$ is singular
    endif
```

Consider a singular cube $Q \subset \mathcal{M}^d$. The *classifying component* $\mathcal{C} = \mathcal{C}(Q)$ is defined and sorted by the following algorithm:

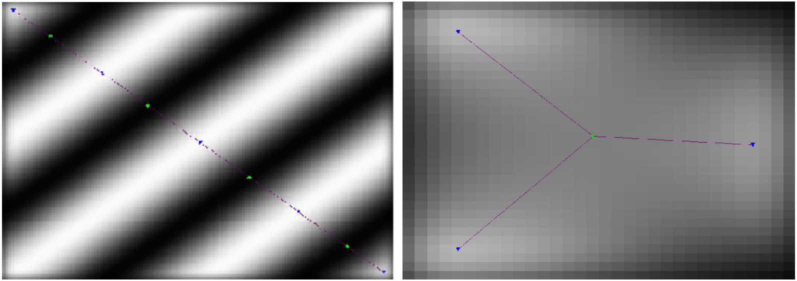
Algorithm (*Sorting components*)

```
For each singular cube $Q$, $\mathcal{C} := wrap(Q)$
while $P \in bd(\mathcal{C}) \cap X$ is singular
    $\mathcal{C} := \mathcal{C} \cup \{P\}$
endwhile
build $N$, $L_p$, $L_n$ for $\mathcal{C}$
 do
    $H := H_*(N, L_p) = 0$ and $H_*(N, L_n) = 0$
        if $H =$ TRUE then $\mathcal{C}$ is a regular component
        else if $L_n = \partial N$ then $\mathcal{C}$ is a maximum component
        else if $L_p = \partial N$ then $\mathcal{C}$ is a minimum component
        else $\mathcal{C}$ is a saddle component
    endif
```

A cube Q is *regular* if $\mathcal{C}(Q)$ is regular and *critical* otherwise.

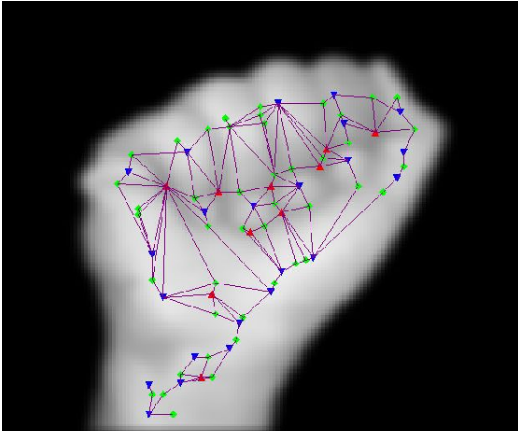
2D experimentation

Allili, Corriveau, Derivière, TK, Trahan [2007]



MCG — Graph edges display connections between critical components.

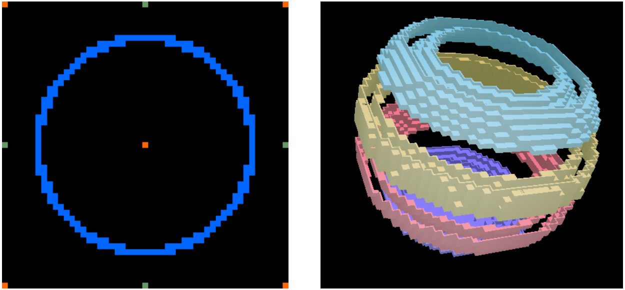
MCG of an image from sign alphabet



Courtesy of Uras and Verri [1994], and Frosini.

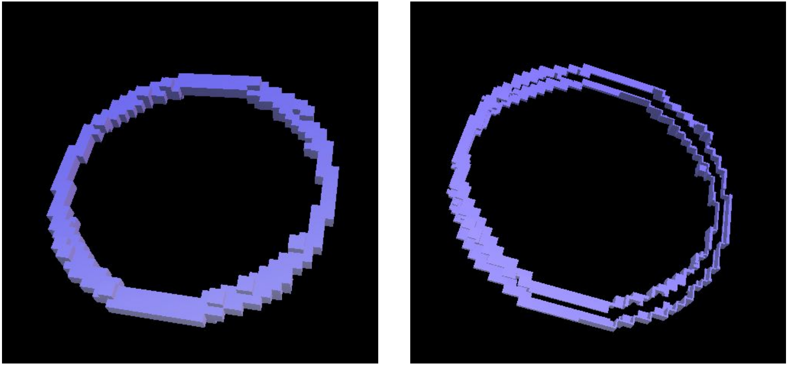
New 2D and 3D experimentation

Allili, Ethier, TK [2010]



Left: critical components of $f(x, y) = (x^2 + y^2 - 5)^2$;
Right: $f(x, y, z) = (x^2 + y^2 + z^2 - 5)^2$, slices through L_p (a union of two spheres).

Twisted Toroid discretisation



Left: N; Right: L_p.

Main directions of future work

- Conceptual: persistence w.r.t. discretisation parameters
  - spatial resolution;
  - dynamical resolution (range thresholding).
- Experimental: testing on larger data files.