Introduction to Topological Data Analysis

Julien Tierny

▶ **To cite this version:**

Julien Tierny. Introduction to Topological Data Analysis. Doctoral. France. 2017. cel-01581941

HAL Id: cel-01581941
https://hal.science/cel-01581941

Submitted on 5 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sorbonne Universités, UPMC Univ Paris 06

Laboratoire d'Informatique de Paris 6

Julien Tierny

Introduction to Topological Data Analysis

Sorbonne Universités, UPMC Univ Paris 06
Laboratoire d'Informatique de Paris 6
UMR UPMC/CNRS 7606 – Tour 26
4 Place Jussieu – 75252 Paris Cedex 05 – France

Notations

$\mathbb{X}$	Topological space
$\partial\mathbb{X}$	Boundary of a topological space
$\mathbb{M}$	Manifold
$\mathbb{R}^d$	Euclidean space of dimension d
$\sigma,\ \tau$	d-simplex, face of a d-simplex
$v,\ e,\ t,\ T$	Vertex, edge, triangle and tetrahedron
$Lk(\sigma),\ St(\sigma)$	Link and star of a simplex
$Lk_d(\sigma),\ St_d(\sigma)$	d-simplices of the link and the star of a simplex
$\mathcal{K}$	Simplicial complex
$\mathcal{T}$	Triangulation
$\mathcal{M}$	Piecewise linear manifold
β_i	i-th Betti number
χ	Euler characteristic
α_i	i^{th} barycentric coordinates of a point p relatively to a simplex σ
$f : \mathcal{T} \to \mathbb{R}$	Piecewise linear scalar field
∇f	Gradient of a PL scalar field f
$Lk^-(\sigma),\ Lk^+(\sigma)$	Lower and upper link of σ relatively to f
$o(v)$	Memory position offset of the vertex v
$\mathcal{L}^-(i),\ \mathcal{L}^+(i)$	Sub- and sur-level set of the isovalue i relatively to f
$\mathcal{D}(f)$	Persistence diagram of f
$\mathcal{C}(f)$	Persistence curve of f
$\mathcal{R}(f)$	Reeb graph of f
$l(\mathcal{R}(f))$	Number of loops of $\mathcal{R}(f)$
$\mathcal{T}(f)$	Contour tree of f
$\mathcal{J}(f),\ \mathcal{S}(f)$	Join and split trees of f
$\mathcal{MS}(f)$	Morse-Smale complex of f

1 Introduction to Topological Data Analysis

Contents

This document introduces the key concepts and algorithms of Topological Data Analysis in low dimensions. First, the input data representation is formalized. Second, some of the core concepts of Topological Data Analysis are presented, including critical points, notions of Persistent Homology, Reeb graphs and Morse-Smale Complexes. Finally, a brief review of the state-of-the-art algorithms is presented.

For the reader's convenience, the most important definitions and properties are highlighted with boxes.

For further readings, I refer the reader to the excellent introduction to Computational Topology by Edelsbrunner and Harer (EH09).

"The most applicable of all is a good theory."
Herbert Edelsbrunner.

1.1 Data representation

In scientific visualization, scalar data is in general defined on an input geometrical object (hereafter named *"Domain"*). It is represented by a finite set of sample values, continuously extended in space to the entirety of the domain thanks to an interpolant. In the following, I first formalize a generic domain representation. Next, I formalize a representation of the scalar data on this object (hereafter termed *"Range"*).

1.1.1 Domain representation

In the following, I formalize a generic domain representation. This notion is introduced constructively. At the end of the subsection, I further describe topological notions relative to this domain representation that will be used in the remainder of the manuscript.

Preliminary notions

Definition 1 (Topology) *A topology on a set* $\mathbb{X}$ *is a collection* T *of subsets of* $\mathbb{X}$ *having the following properties:*

- *The sets* $\emptyset$ *and* $\mathbb{X}$ *are in* T*;*
- *The union of any sub-collection of* T *is in* T*;*
- *The intersection of a finite sub-collection of* T *is in* T*.*

Definition 2 (Topological space) *A set* $\mathbb{X}$ *for which a topology* T *is defined is called a* topological space*.*

For example, the space of real numbers $\mathbb{R}$ is a topological space.

Definition 3 (Open set) *A subset* $\mathbb{A} \subset \mathbb{X}$ *of the topological space* $\mathbb{X}$ *is an* open *set of* $\mathbb{X}$ *if it belongs to* T*.*

Definition 4 (Closed set) *A subset* $\mathbb{B} \subset \mathbb{X}$ *of the topological space* $\mathbb{X}$ *is a* closed *set of* $\mathbb{X}$ *if its complement* $\mathbb{X} - \mathbb{B}$ *is open.*

Intuitively, open sets are subsets of topological spaces which do not contain their boundaries. For example, considering the space of real numbers $\mathbb{R}$, $(-\infty, 0) \cup (1, +\infty)$ and $[0, 1]$ are complements and respectively open and closed sets.

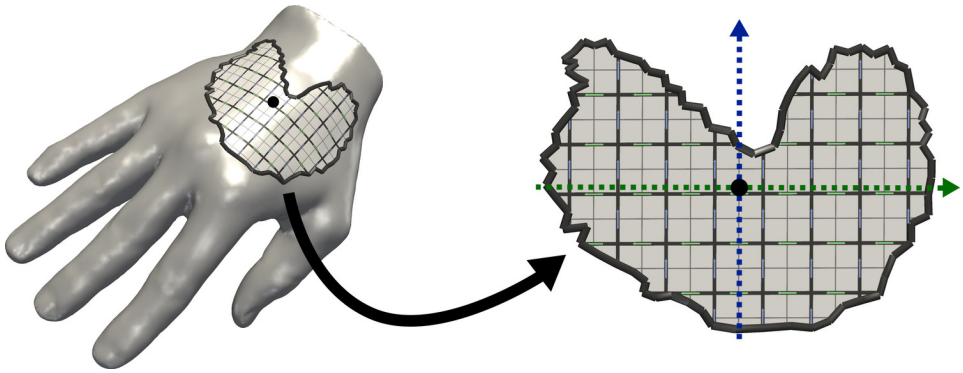

Figure 1.1 – *Example of* 2*-manifold: any point of the surface (left, black dot) has an open neighborhood (textured chart) that is homeomorphic to an open Euclidean* 2*-ball (that can be* unfolded *to the plane, right).*

Property 1 (Open sets)

- *The set* $\emptyset$ *is open;*
- *The union of any number of open sets is open;*
- *The intersection of a finite number of open sets is open.*

These properties follow from the definition of topology.

Definition 5 (Covering) *A collection of subsets of a topological space* $\mathbb{X}$ *is a* covering *of* $\mathbb{X}$ *if the union of all its elements is equal to* $\mathbb{X}$*.*

Definition 6 (Compact topological space) *A topological space* $\mathbb{X}$ *is* compact *if every open covering of it contains a finite sub-collection that is also a covering of* $\mathbb{X}$*.*

Definition 7 (Function) *A* function $f : \mathbb{A} \to \mathbb{B}$ *associates each element of the topological space* $\mathbb{A}$ *with a unique element of the topological space* $\mathbb{B}$*.*

Definition 8 (Injection) *A function* $f : \mathbb{A} \to \mathbb{B}$ *is an* injection *if for each pair* $a_1, a_2 \in \mathbb{A}$ *such that* $a_1 \neq a_2$*,* $f(a_1) \neq f(a_2)$*.* f *is said to be* one-to-one*.*

Definition 9 (Bijection) *A function* $f : \mathbb{A} \to \mathbb{B}$ *is a* bijection *if for each element* $b \in \mathbb{B}$ *there is exactly one element* $a \in \mathbb{A}$ *such that* $f(a) = b$*.* f *is said to be bijective. It is also said to be* one-to-one *(injective) and* onto *(surjective).*

Definition 10 (Continuous function) *A function* $f : \mathbb{A} \to \mathbb{B}$ *is* continuous *if for each open subset* $\mathbb{C} \in \mathbb{B}$*, the set* $f^{-1}(\mathbb{C})$ *is an open subset of* $\mathbb{A}$*.*

Definition 11 (Homeomorphic spaces) *Two topological spaces* $\mathbb{A}$ *and* $\mathbb{B}$ *are* homeomorphic *if and only if there exists a continuous bijection* $f : \mathbb{A} \to \mathbb{B}$ *with a continuous inverse* $f^{-1} : \mathbb{B} \to \mathbb{A}$*.* f *is a* homeomorphism*.*

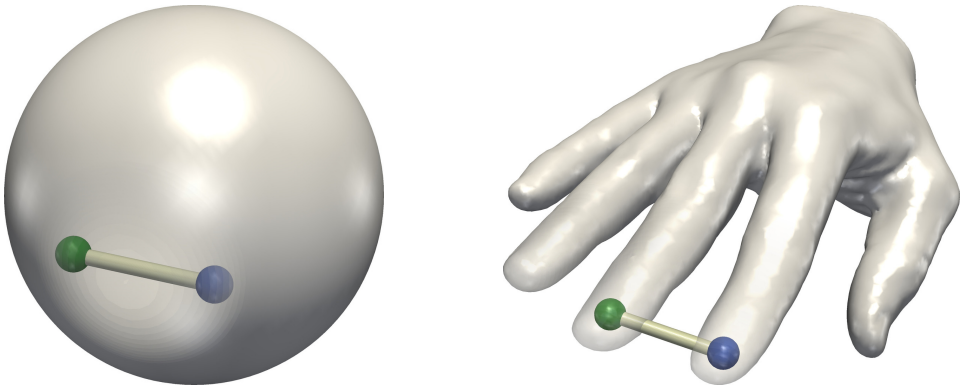

Figure 1.2 – *Examples of convex (left) and non-convex (right) 3-manifolds (volumes). On the left, any two points (green and blue spheres) can be linked by a line segment that belongs to the volume (white cylinder). This is not the case for the right volume.*

Definition 12

(Manifold) *A topological space* $\mathbb{M}$ *is a* d-manifold *if every element* $m \in \mathbb{M}$ *has an open neighborhood* $\mathbb{N}$ *homeomorphic to an open Euclidean* d*-ball.*

An intuitive description of a d-manifold is that of a curved space, which has locally the structure of an Euclidean space of dimension d, but which has a more complicated global structure (Euclidean spaces are therefore special cases of manifolds). Figure 1.1 illustrates this with the example of a 2-manifold (surface).

Domain formalization

In the following we formally introduce our domain representation as well as representations for connectivity information.

Definition 13 (Convex set) *A set* $\mathbb{C}$ *of an Euclidean space* $\mathbb{R}^n$ *of dimension* n *is* convex *if for any two points* x *and* y *of* $\mathbb{C}$ *and all* $t \in [0,1]$ *the point* $(1-t)x + ty$ *also belongs to* $\mathbb{C}$*.*

Intuitively, a convex set is a set such that any two points of the set can be linked by a line segment that belongs to the set, as illustrated with 3-manifolds (volumes) in Figure 1.2.

Definition 14 (Convex hull) *The* convex hull *of a set points* $\mathcal{P}$ *of an Euclidean space* $\mathbb{R}^n$ *is the unique minimal convex set containing all points of* $\mathcal{P}$*.*

Definition 15 (Simplex) *A* d-simplex *is the convex hull* σ *of* $d+1$ *affinely independent points of an Euclidean space* $\mathbb{R}^n$*, with* $0 \leq d \leq n$*.* d *is the dimension of* σ*.*

Definition 16 (Vertex) *A* vertex v *is a* 0*-simplex of* $\mathbb{R}^3$*.*

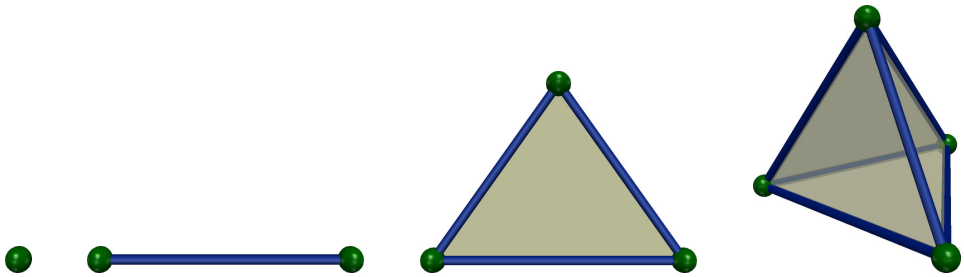

Figure 1.3 – *Illustrations of* 0 *(green),* 1 *(blue),* 2 *(white) and* 3*-simplices (transparent), from left to right, along with their faces.*

Definition 17 (Edge) *An* edge *e is a* 1*-simplex of* $\mathbb{R}^3$.

Definition 18 (Triangle) *A* triangle *t is a* 2*-simplex of* $\mathbb{R}^3$.

Definition 19 (Tetrahedron) *A* tetrahedron *T is a* 3*-simplex of* $\mathbb{R}^3$.

Definition 20 (Face) *A* face *τ of a d-simplex σ is the simplex defined by a non-empty subset of the $d+1$ points of σ, and is noted $\tau \leq \sigma$. We will note τ_i a face of dimension i.*

In summary, a d-simplex is the smallest combinatorial construction that can represent a neighborhood of a d-dimensional Euclidean space. As illustrated in Figure 1.3, it is composed of *faces*, that are themselves $(d-1)$, $(d-2)$, ..., and 0-simplices.

Definition 21 (Simplicial complex) *A* simplicial complex *$\mathcal{K}$ is a finite collection of non-empty simplices $\{\sigma_i\}$, such that every face τ of a simplex σ_i is also in $\mathcal{K}$, and any two simplices σ_i and σ_j intersect in a common face or not at all.*

Definition 22 (Star) *The* star *of a simplex σ of a simplicial complex $\mathcal{K}$ is the set of simplices of $\mathcal{K}$ that contain σ: $St(\sigma) = \{\tau \in \mathcal{K}, \sigma \leq \tau\}$. We will note $St_d(\sigma)$ the set of d-simplices of $St(\sigma)$.*

Definition 23

(Link) *The* link *of σ is the set of faces of the simplices of $St(\sigma)$ that are disjoint from σ: $Lk(\sigma) = \{\tau \leq \Sigma, \Sigma \in St(\sigma), \tau \cap \sigma = \emptyset\}$. We will note $Lk_d(\sigma)$ the set of d-simplices of $Lk(\sigma)$.*

In other words, the star of a simplex σ is the set of simplices having σ as a face, as illustrated Figure 1.4 (top). The notion of link is illustrated at the bottom of Figure 1.4.

Definition 24 (Underlying space) *The* underlying space *of a simplicial complex $\mathcal{K}$ is the union of its simplices $|\mathcal{K}| = \cup_{\sigma \in \mathcal{K}} \sigma$.*

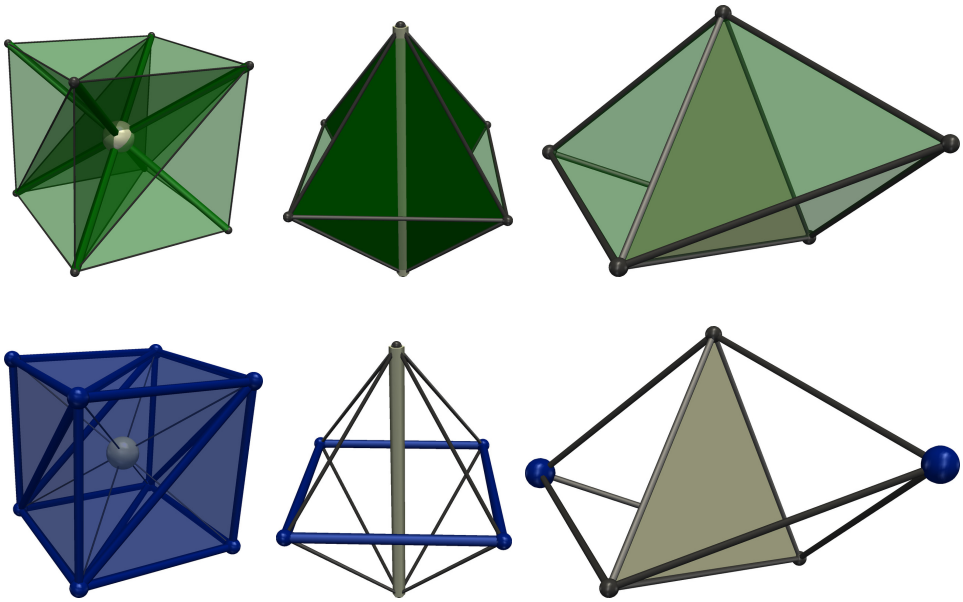

Figure 1.4 – *Illustrations of stars (green, top) and links (blue, bottom) for* 0*,* 1 *and* 2*-simplices (white, from left to right) of a* 3*-dimensional simplicial complex.*

Definition 25

(Triangulation) *The* triangulation $\mathcal{T}$ *of a topological space* $\mathbb{X}$ *is a simplicial complex* $\mathcal{K}$ *whose underlying space* $|\mathcal{K}|$ *is homeomorphic to* $\mathbb{X}$*.*

The notion of triangulation has been preferred here to other competing representations for its practical genericity: any mesh representation (regular grid, unstructured grid, etc.) can be easily converted into a triangulation by subdividing each of its d-cells into valid d-simplices (having only $(d+1)$ linearly independent points), as illustrated in Figure 1.5 for the case of a regular grid. Also, note that for regular grids, the resulting triangulation can be implicitly encoded (i.e. adjacency relations can be retrieved on demand, without storage, thanks to the recurring subdivision pattern of the regular grid). Moreover, as detailed in the next subsection, triangulations can be accompanied with well-behaved interpolants, which facilitate reasoning and computation with scalar data.

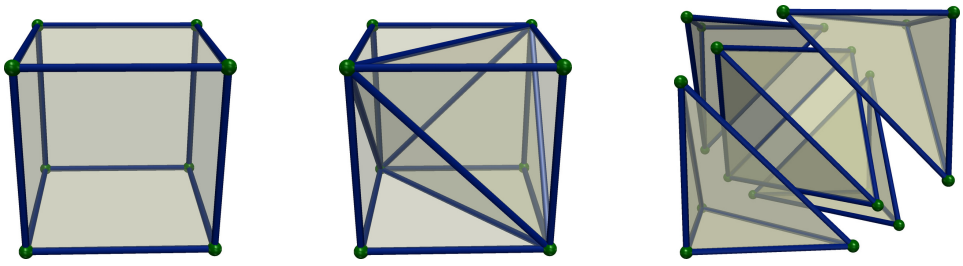

Figure 1.5 – *A* 3*-dimensional regular grid (left) can be easily converted into a triangulation by subdividing each of its voxels independently into* 5 *tetrahedra (center, right: exploded view). This subdivision can be implicitly encoded.*

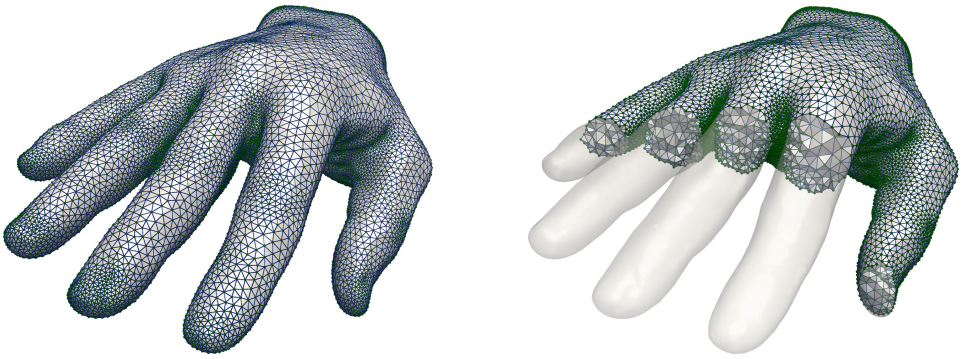

Figure 1.6 – *Example of PL 3-manifold (left, right: clipped view).*

As discussed further down this manuscript, for reasoning and robustness purposes, the following, more restrictive, notion is often preferred over triangulations.

Definition 26

(Piecewise Linear Manifold) *The triangulation of a manifold* $\mathbb{M}$ *is called a* piecewise linear manifold *and is noted* $\mathcal{M}$.

Therefore, a piecewise linear (PL) manifold is a combinatorial representation of a manifold that derives from the notion of triangulation, as illustrated in Figure 1.6. It can be efficiently represented in memory by storing for each dimension d, the list of d-simplices as well as their stars and links. In the remainder of this manuscript, we will consider PL-manifolds as our generic domain representations.

Topological invariants

In the following, I describe a few *topological invariants*: entities that do not change under continuous transformations of the domain (variations in point positions but no variation in connectivity). These notions are instrumental in Topological Data Analysis.

Definition 27 (Path) *A homeomorphism* $p : (a, b) \to \mathbb{C}$ *from an open interval* $(a, b) \subseteq \mathbb{R}$ *to a subset* $\mathbb{C}$ *of a topological space* $\mathbb{X}$ *is called a* path *on* $\mathbb{X}$ *between* $p(a)$ *and* $p(b)$.

Definition 28 (Connected topological space) *A topological space* $\mathbb{X}$ *is* connected *if for any two points of* $\mathbb{X}$ *there exists a path between them on* $\mathbb{X}$. *be*

Definition 29 (Connected components) *The maximally connected subsets of a topological space* $\mathbb{X}$ *are called its* connected components.

Definition 30 (Homotopy) *A* homotopy *between two continuous functions* f *and* g *is a continuous function* $H : \mathbb{X} \times [0, 1] \to \mathbb{Y}$ *from the product of a topological space*

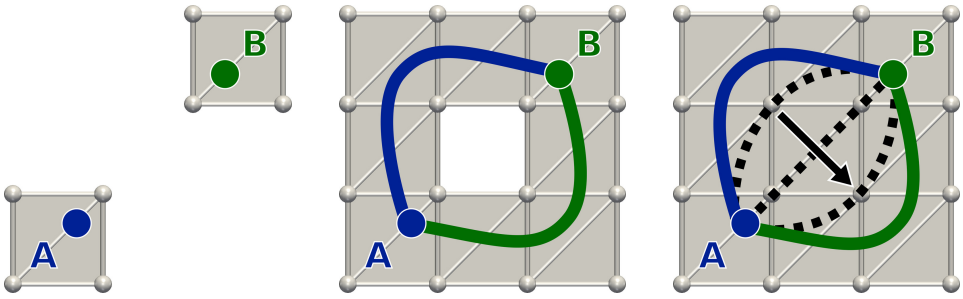

Figure 1.7 – *Examples of disconnected, connected and simply connected domains (from left to right).*

$\mathbb{X}$ *with the closed unit interval to a topological space* $\mathbb{Y}$ *such that for each point* $x \in \mathbb{X}$, $H(x,0) = f(x)$ *and* $H(x,1) = g(x)$. *If there exists a homotopy between them,* f *and* g *are said to be* homotopic.

While homeomorphism deals with the matching between neighborhoods, homotopies additionally require that a continuous transformation exist between them, by considering neighborhoods as images of functions (the notion of homotopy is then refined to that of isotopy). Here, the second parameter of an homotopy can be seen as time in this continuous transformation process. For instance, a circle and a knot are homeomorphic but are not homotopic since the knot needs to be cut and stitched back to be turned into a circle, which is not a continuous transformation.

Definition 31

(Simply connected topological space) *A topological space* $\mathbb{X}$ *is* simply connected *if it is connected and if for any two points of* $\mathbb{X}$, *any two paths between them on* $\mathbb{X}$ *are homotopic.*

As illustrated in Figure 1.7, a domain is not simply connected if for any two points, any pair of paths between them cannot be continuously transformed into one another (black paths in Figure 1.7, right).

Definition 32 (Boundary) *The* boundary *of a topological space* $\mathbb{X}$, *noted* $\partial\mathbb{X}$, *is the complement in* $\mathbb{X}$ *of the subspace of* $\mathbb{X}$, *called the* interior *of* $\mathbb{X}$, *composed of all the elements* $x \in \mathbb{X}$ *such that* x *has an open neighborhood* $\mathbb{N}$.

Definition 33 (Boundary component) *A* boundary component *of a topological space* $\mathbb{X}$ *is a connected component of its boundary* $\partial\mathbb{X}$.

Definition 34 (p-chain) *A* p-chain *of a triangulation* $\mathcal{T}$ *of a topological space* $\mathbb{X}$ *is a formal sum (with modulo* 2 *coefficients) of* p*-simplices of* $\mathcal{T}$.

Definition 35 (p-cycle) *A* p-cycle *of a triangulation* $\mathcal{T}$ *of a topological space* $\mathbb{X}$ *is a* p*-chain with empty boundary.*

Definition 36 (Group of p-cycles) *The* group of p-cycles *of a triangulation* $\mathcal{T}$ *of a topological space* $\mathbb{X}$ *is the group of all* p*-cycles of* $\mathcal{T}$*, noted* $Z_p(\mathcal{T})$*, which forms a sub-group of all* p*-chains of* $\mathcal{T}$*.*

Definition 37 (p-boundary) *A* p-boundary *of a triangulation* $\mathcal{T}$ *of a topological space* $\mathbb{X}$ *is the boundary of a* $(p+1)$*-chain.*

Property 2 (p-boundary) *A* p*-boundary is a* p*-cycle.*

Definition 38 (Group of p-boundaries) *The* group of p-boundaries *of a triangulation* $\mathcal{T}$ *of a topological space* $\mathbb{X}$ *is the group of all* p*-boundaries of* $\mathcal{T}$*, noted* $B_p(\mathcal{T})$*, which forms a sub-group of all* p*-cycles of* $\mathcal{T}$*.*

Definition 39 (Homology group) *The* p^{th} homology group *of a triangulation* $\mathcal{T}$ *of a topological space* $\mathbb{X}$ *is its* p^{th} *cycle group modulo its* p^{th} *boundary group:* $H_p(\mathcal{T}) = Z_p(\mathcal{T})/B_p(\mathcal{T})$*.*

Intuitively, two p-cycles are said to be equivalent, or homologous, if they can be continuously transformed into each other (through formal sums with modulo 2 coefficients) without being collapsible to a point. Then, one can further group p-cycles into *classes* of equivalent p-cycles. Each class can be represented by a unique representative p-cycle that is called *generator* (and that is homologous to any other p-cycle of the class), as illustrated in Figure 1.8 with a green 1-cycle (center) and a green 2-cycle (right). Enumerating the number of generators of a homology group enables to introduce intuitive topological invariants called Betti numbers.

Definition 40

(Betti number) *The* p^{th} Betti number *of a triangulation* $\mathcal{T}$ *of a topological space* $\mathbb{X}$ *is the rank of its* p^{th} *homology group:* $\beta_p(\mathcal{T}) = rank(H_p(\mathcal{T}))$*.*

In low dimensions, Betti numbers have a very concrete interpretation. For instance, for PL 3-manifolds, β_0 corresponds to the number of connected components, β_1 to the number of handles and β_2 to the number of voids, as illustrated in Figure 1.8 (β_3 is equal to 0 for PL 3-manifolds with boundary, i.e. that can be embedded in $\mathbb{R}^3$).

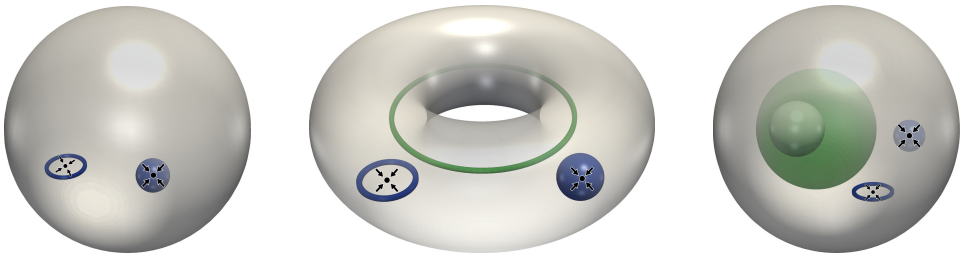

Figure 1.8 – *Examples of PL 3-manifolds with varying Betti numbers. From left to right: a 3-ball, a solid torus, a 3-ball with a void. From left to right,* $(\beta_0, \beta_1, \beta_2)$ *is equal to* $(1,0,0)$*,* $(1,1,0)$*, and* $(1,0,1)$*). Generators are displayed in green, while examples of non-generator p-cycles are displayed in blue.*

Definition 41

(Euler characteristic) *The* Euler characteristic *of a triangulation* $\mathcal{T}$ *of a topological space* $\mathbb{X}$ *of dimension* d*, noted* $\chi(\mathcal{T})$*, is the alternating sum of its Betti numbers:* $\chi(\mathcal{T}) = \sum_{i=0}^{i=d}(-1)^i\beta_i(\mathcal{T})$.

Property 3 (Euler characteristic) *The* Euler characteristic *of a triangulation* $\mathcal{T}$ *of a topological space* $\mathbb{X}$ *of dimension* d *is also equal to the alternating sum of the number of its i-simplices:* $\chi(\mathcal{T}) = \sum_{i=0}^{i=d}(-1)^i|\sigma_i|$.

1.1.2 Range representation

In the following, I formalize a range representation based on the previously introduced domain representation. Additionally, I will introduce a few related geometrical constructions that will be instrumental to Topological Data Analysis.

Piecewise linear scalar fields

Definition 42

(Barycentric coordinates) *Let* p *be a point of* $\mathbb{R}^n$ *and* σ *a* d*-simplex. Let* α_0*,* α_1*,* $\ldots$*,* α_d *be a set of real coefficients such that* $p = \sum_{i=0}^{i=d}\alpha_i\tau_0^i$ *(where* τ_0^i *is the* i^{th} *zero dimensional face of* σ*) and such that* $\sum_{i=0}^{i=d}\alpha_i = 1$*. Such coefficients are called the* barycentric coordinates *of* p *relatively to* σ.

Property 4 (Barycentric coordinates) *The barycentric coordinates of* p *relative to* σ *are unique.*

Property 5 (Barycentric coordinates) *If and only if there exists an* i *for which* $\alpha_i \notin [0,1]$*, then* p *does not belong to* σ*, otherwise it does.*

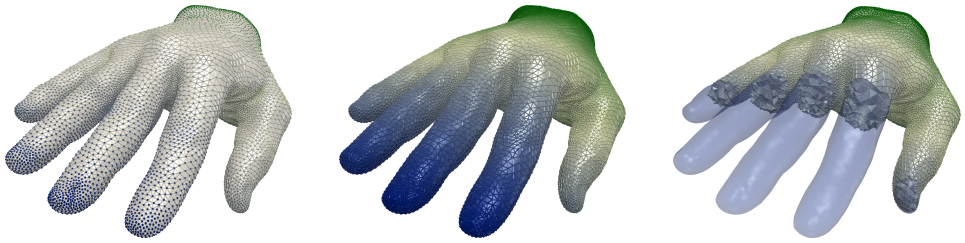

Figure 1.9 – *Example of PL scalar field f defined on a PL 3-manifold $\mathcal{M}$. From left to right: restriction $\hat{f}$ of f on the 0-simplices of $\mathcal{M}$, f (the color coding denotes the linear interpolation within each simplex), clipped view of f.*

Definition 43

(Piecewise Linear Scalar Field) *Let $\hat{f}$ be a function that maps the 0-simplices of a triangulation $\mathcal{T}$ to $\mathbb{R}$. Let $f : \mathcal{T} \to \mathbb{R}$ be the function linearly interpolated from $\hat{f}$ such that for any point p of a d-simplex σ of $\mathcal{T}$, we have: $f(p) = \sum_{i=0}^{i=d} \alpha_i \hat{f}(\tau_0^i)$ (where τ_0^i is the i^{th} zero dimensional face of σ). f is called a* piecewise linear (PL) scalar field.

Piecewise linear scalar fields will be our default representation for scalar data. Typically, the input data will then be given in the form of a triangulation with scalar values attached to its vertices ($\hat{f}$). The linear interpolation provided by the barycentric coordinates can be efficiently computed on demand (on the CPU or the GPU, as illustrated in Figure 1.9) and has several nice properties that makes it well suited for combinatorial reasonings.

Property 6 *(Gradient of a Piecewise Linear Scalar Field) The* gradient ∇f *of a PL scalar field $f : \mathcal{T} \to \mathbb{R}$ is a curl free vector field that is piecewise constant (constant within each d-simplex of $\mathcal{T}$).*

This property has several implications that will be discussed in the following subsections.

Definition 44 (Lower Link) *The* lower link $Lk^-(\sigma)$ *(respectively the* upper link $Lk^+(\sigma)$*) of a d-simplex σ relatively to a PL scalar field f is the subset of the link $Lk(\sigma)$ such that each of its zero dimensional faces has a strictly lower (respectively higher) f value than those of σ.*

Given the above definition, it is often useful to disambiguate configurations of equality in f values between vertices (thus equalitiy configurations in $\hat{f}$). Therefore, $\hat{f}$ is often slightly perturbed with a mechanism inspired by simulation of simplicity (EM90) to turn $\hat{f}$ into an injective function. This can be achieved in the following way, by adding to $\hat{f}$ a second function $\hat{g}$

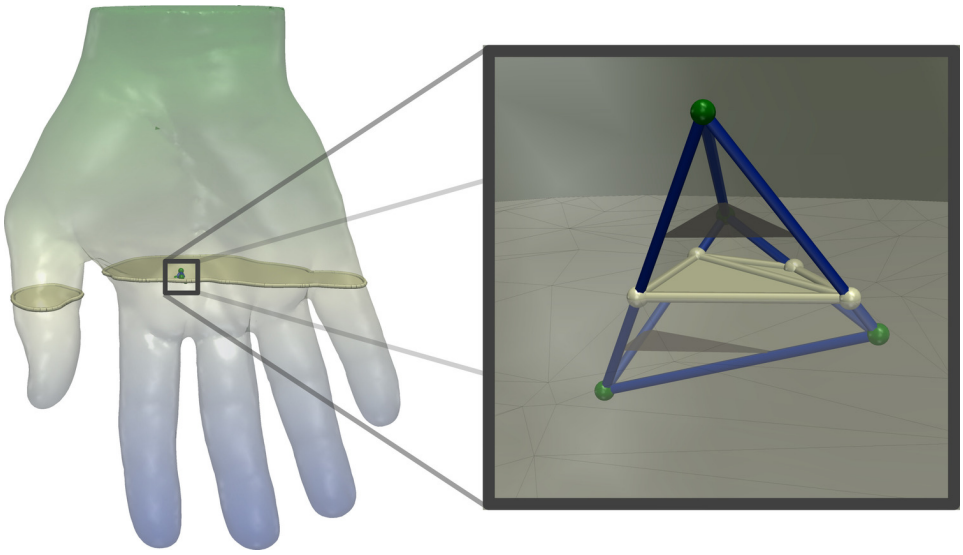

Figure 1.10 – *Example of level set (isosurface, left) of a PL scalar field defined on a PL 3-manifold. Right: restriction of the isosurface to a 3-simplex.*

that is injective. Let $o(v)$ denote the position integer offset of the vertex v in memory. $o(v)$ is injective. Then, to turn $\hat{f}$ into an injective function, one needs to add to it $\epsilon o(v)$ where ϵ is an arbitrarily small real value. As the original simulation of simplicity, this mechanism can be implemented numerically (by choosing the smallest possible value for ϵ depending on the machine precision) or preferably symbolically by re-implementing the necessary predicates. For instance, to decide if a vertex v_0 is lower than a vertex v_1, one needs to test $\hat{f}(v_0) < \hat{f}(v_1)$ and, in case of equality, test $o(v_0) < o(v_1)$ to disambiguate. In the following, we will therefore consider that $\hat{f}$ is always injective in virtue of this mechanism. Therefore, no d-simplex of $\mathcal{T}$ collapses to a point of $\mathbb{R}$ through f for any non-zero d.

Related geometrical constructions

Based on our representation for scalar data on geometrical domains, I will now introduce a few geometrical constructions that will be instrumental in Topological Data Analysis.

Definition 45 (Sub-level set) *The* sub-level set $\mathcal{L}^-(i)$ *(respectively the* sur-level set $\mathcal{L}^+(i)$*) of an isovalue $i \in \mathbb{R}$ relatively to a PL scalar field $f : \mathcal{M} \to \mathbb{R}$ is the set of points:* $\{p \in \mathcal{M} \mid f(p) \leq i\}$ *(respectively* $\{p \in \mathcal{M} \mid f(p) \geq i\}$*).*

Definition 46

(Level set) *The* level-set $f^{-1}(i)$ *of an isovalue* $i \in \mathbb{R}$ *relatively to a PL scalar field* $f : \mathcal{M} \to \mathbb{R}$ *is the pre-image of* i *onto* $\mathcal{M}$ *through* f*:* $f^{-1}(i) = \{p \in \mathcal{M} \mid f(p) = i\}$.

Property 7 (Level set) *The level set* $f^{-1}(i)$ *of a regular isovalue* $i \in \mathbb{R}$ *relatively to a PL scalar field* $f : \mathcal{M} \to \mathbb{R}$ *defined on a PL* d*-manifold* $\mathcal{M}$ *is a* $(d-1)$*-manifold.*

Property 8 (Level set) *Let* $f : \mathcal{T} \to \mathbb{R}$ *be a PL scalar field and* σ *be a* d*-simplex of* $\mathcal{T}$*. For any isovalue* $i \in f(\sigma)$*, the restriction of* $f^{-1}(i)$ *within* σ *belongs to an Euclidean subspace of* $\mathbb{R}^d$ *of dimension* $(d-1)$.

This latter property directly follows from the property 6 on the gradient of PL scalar fields, which is piecewise constant (a level set is everywhere orthogonal to the gradient). It follows that the level sets of PL scalar fields defined on PL manifolds can be encoded as PL manifolds, as illustrated with the white PL 2-manifold in Figure 1.10 (right).

Property 9 (Level set) *Let* $f : \mathcal{T} \to \mathbb{R}$ *be a PL scalar field and* σ *be a* d*-simplex of* $\mathcal{T}$*. For any two isovalues* $i \neq j$ *belonging to* $f(\sigma)$*, the restrictions of* $f^{-1}(i)$ *and of* $f^{-1}(j)$ *within* σ *are parallel.*

This property also follows from the property 6 on the gradient of PL scalar fields and is illustrated in Figure 1.10 (right, dark gray isosurfaces), which shows an *isosurface* restricted to a 3-simplex (i.e. a level set of a PL scalar field defined on a PL 3-manifold). Such strong properties (planarity and parallelism) enable to derive robust and easy-to-implement algorithms for level set extraction (called *"Marching Tetrahedra"* for PL 3-manifolds, and *"Marching Triangles"* for PL 2-manifolds).

Definition 47 (Contour) *Let* $f^{-1}(i)$ *be the level set of an isovalue* i *relatively to a PL scalar field* $f : \mathcal{T} \to \mathbb{R}$*. Each connected component of* $f^{-1}(i)$ *is called a contour.*

Definition 48

(Integral line) *Let* $f : \mathcal{M} \to \mathbb{R}$ *be a PL scalar field defined on a PL manifold* $\mathcal{M}$*. An* integral line *is a path* $p : \mathbb{R} \to \mathcal{C} \subset \mathcal{M}$ *such that* $\frac{\partial}{\partial t} p(t) = \nabla f(p(t))$*.* $lim_{t \to -\infty} p(t)$ *and* $lim_{t \to \infty} p(t)$ *are called the* origin *and the* destination *of the integral line respectively.*

In other words, an integral line is a path which is everywhere tangential to the gradient. In virtue of property 6 on the gradient of PL scalar fields, it follows that integral lines can be encoded as PL 1-manifolds, as illustrated with the white PL 1-manifold in Figure 1.11 (right).

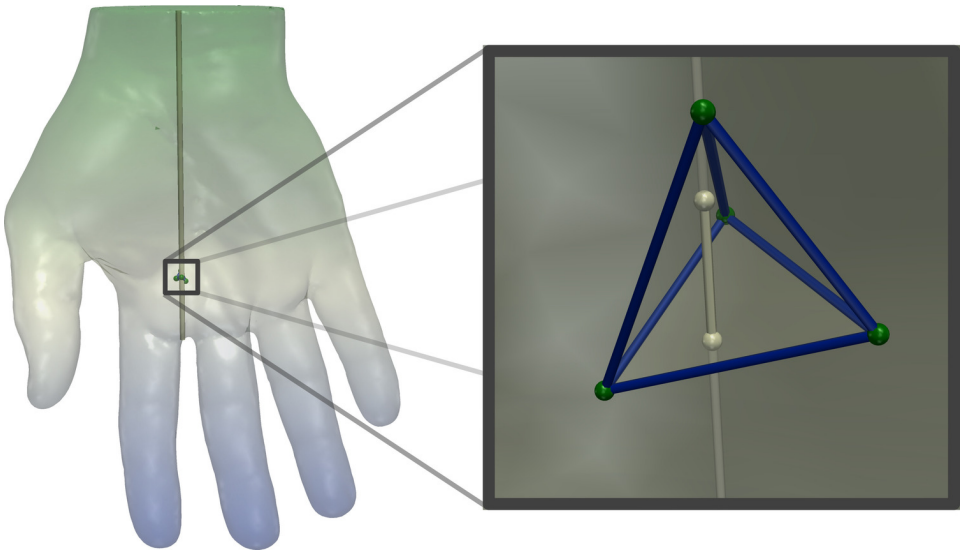

Figure 1.11 – *Examples of integral line (left) of a PL scalar field defined on a PL* 3*-manifold. Right: restriction of the integral line to a* 3*-simplex.*

1.2 Topological abstractions

Level sets (and especially contours) and integral lines are fundamental geometrical objects in Scientific Visualization for the segmentation of regions of interests (burning flames in combustion, interaction pockets in chemistry, etc.) or the extraction of filament structures (galaxy backbones in cosmology, covalent interactions in chemistry, etc.).

Intuitively, the key idea behind Topological Data Analysis is to segment the data into regions where these geometrical objects are *homogeneous* from a topological perspective, and to summarize these homogeneity relationships into a *topological abstraction*. Such a segmentation strategy enables to access these features more efficiently and to classify them according to application dependent metrics for further processing.

In the following, I introduce such topological abstractions for feature extraction, segmentation and classification purposes.

1.2.1 Critical points

In the smooth setting, critical points are points of a manifold where the gradient of a smooth scalar field vanishes. Unfortunately, this notion does not directly translate into the PL setting since the gradient of a PL scalar field is piecewise constant. This requires to use an alternate definition, which interestingly involves topological and combinatorial reasonings.

Morse theory (Mil63) relates the study of the topology of manifolds to the study of a specific group of smooth scalar fields defined on them

(called *Morse* functions). One of the key results of Morse theory is the following property: the Betti numbers of the sub-level sets of a Morse function only change in the vicinity of a critical point. In other words, the topology of the sub-level sets only evolves when crossing a critical point. This observation is at the basis of a formalization of critical points in the PL setting.

Definition 49

(Critical point) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL scalar field defined on a PL manifold $\mathcal{M}$. A vertex v of $\mathcal{M}$ is a* regular point *if and only if both $Lk^-(v)$ and $Lk^+(v)$ are simply connected. Otherwise, v is a* critical point *of f and $f(v)$ is called a* critical isovalue *(as opposed to* regular isovalues*).*

Definition 50 (Critical contour) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL scalar field defined on a PL manifold $\mathcal{M}$. A contour of f which contains one of its critical points is called a* critical contour*.*

In virtue of these definitions and the property 6 on their gradient, PL scalar fields have many nice properties regarding their critical points.

Property 10 (Critical points) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL scalar field defined on a compact PL manifold $\mathcal{M}$. The set of critical points of f, noted C_f, contains only isolated critical points and its cardinality $|C_f|$ is finite.*

These properties follow from the fact that the gradient of a PL scalar field is piecewise constant: $\hat{f}$ is assumed to be injective, thus any d-simplex with $d \neq 0$ is mapped to a non null gradient vector. Therefore, critical points are isolated and can only occur on vertices. This makes their number finite for compact PL manifolds. This property is essential for a combinatorial reasoning on critical points. Also, note that the above definitions are independent of the dimension of $\mathcal{M}$.

Definition 51 (Extremum) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL scalar field defined on a PL manifold $\mathcal{M}$. A critical point v is a* minimum *(respectively a* maximum*) of f if and only if $Lk^-(v)$ (respectively $Lk^+(v)$) is empty.*

Definition 52 (Saddle) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL scalar field defined on a PL manifold $\mathcal{M}$. A critical point v is a* saddle *if and only if it is neither a minimum nor a maximum of f.*

Figure 1.12 illustrates the notion of critical points on a toy example. The evolution in the topology of the level sets can be observed in the right insets, which illustrate vertex stars (the smallest combinatorial neighbor-

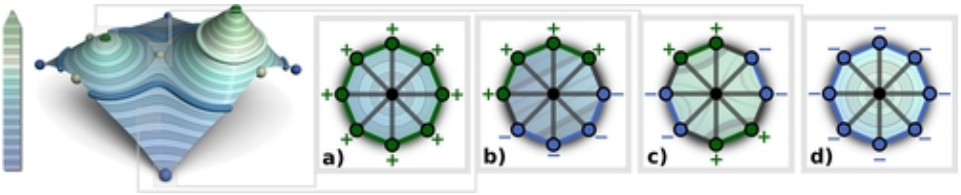

Figure 1.12 – *Scalar field on a terrain (left). A level set is shown in blue, a contour is shown in white. Vertices can be classified according to the the connectivity of their lower (blue) and upper (green) links. From left to right: a minimum (a, blue spheres on the left), a regular point (b), a simple saddle (c, white spheres on the left) and a maximum (d, green spheres on the left).*

hood of a vertex in a triangulation). For a minimum (respectively a maximum) $\beta_0(f^{-1}(i))$ increases (respectively decreases) by one in the vicinity of the extremum. For a regular point, this number does not evolve when crossing a regular point. When crossing a saddle, the number of connected components of the restriction of $f^{-1}(i)$ to the star of the vertex first decreases by one exactly at the saddle and then increases by one right above it.

Critical points are usually classified according to their index. For PL scalar fields defined on PL 2-manifolds, minima have index 0, saddles index 1 and maxima index 2. As the dimension of the domain increases, the number of types of critical points also increases. For PL scalar fields defined on PL 3-manifolds, minima have index 0, 1-saddles (saddles that locally merge level sets) have index 1, 2-saddles (saddles that locally split level sets) have index 2 and maxima index 3. In the following, we will note C_f^i the set of critical points of f of index i.

Definition 53 (Saddle multiplicity) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL scalar field defined on a PL manifold $\mathcal{M}$ and let v be a saddle of f. Let k be the maximum value between $\beta_0(Lk^-(v))$ and $\beta_0(Lk^+(v))$. The* multiplicity *of a saddle is equal to $(k-1)$. A saddle of multiplicity 1 is called a* simple saddle. *It is called a* multi-saddle *otherwise, or $(k-1)$-fold saddle, or alternatively a* degenerate critical point.

Definition 54

(PL Morse scalar field) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL scalar field defined on a PL manifold $\mathcal{M}$. f is a* PL Morse scalar field *if and only if* (i) *all its critical points have distinct f values and* (ii) *f has no degenerate critical point.*

In practice, any PL scalar field can be easily perturbed into a PL Morse scalar field. The first condition can be easily satisfied by forcing $\hat{f}$ to be injective as described in the previous subsection. The second condition can be satisfied by a process called multi-saddle unfolding (EH09), that

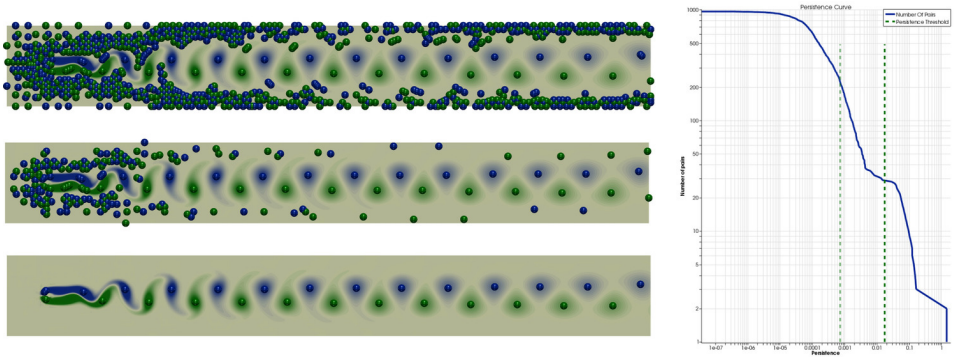

Figure 1.13 – *Minima (blue) and maxima (green) of the orthogonal curl component of a flow simulation of the von Kármán street (flow turbulence behind an obstacle, here at the left of the domain). Right: persistence curve of the field. Selecting extrema involved in pairs more persistent than an increasing threshold (vertical lines, right) yields a hierarchy of critical point sets (left). Here, the light green vertical line (right) corresponds to the middle level (left) while the dark green line (right) corresponds to the bottom level (left). In practice, a flat plateau in the persistence curve (right) often indicates a separation between noise and features.*

locally re-triangulates the star of a $(k-1)$-fold saddle into $(k-1)$ simple saddles.

Interestingly, PL Morse scalar fields inherit from most of the properties of their smooth counter-parts. In particular, the Morse-Euler relation, as first shown by Banchoff (Ban70), still holds for PL Morse scalar fields.

Property 11

(Morse-Euler relation) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL Morse scalar field defined on a closed PL manifold $\mathcal{M}$ of dimension d. Then the* Morse-Euler relation *holds:* $\chi(\mathcal{M}) = \sum_{i=0}^{i=d} (-1)^i |C_f^i|$

This property nicely summarizes the relation between the critical points of a PL Morse scalar field and the topology of its domain. Moreover, it also illustrates the global consistency of the local critical point classification definition (definition 49).

In practice, critical points often directly translate into points of interest application wise. For instance, in 2D vector fields obtained in computational fluid dynamics, extrema of the curl of the field indicate the locations of vortices, a high-level notion that has important implications in the efficiency evaluation of a flow (Figure 1.13, bottom).

1.2.2 Notions of persistent homology

As described in the previous subsection, critical points of PL scalar fields can be extracted with a robust, localized yet globally consistent, combina-

torial and inexpensive classification (definition 49). However, in practice, this classification strategy will identify, among others, critical points corresponding to slight function undulations coming from the noise in the data generation process (acquisition noise, numerical noise in simulations), as illustrated in Figure 1.13 (top). Therefore, to make critical point extraction reliable and useful in practice, one needs to derive a mechanism to further classify critical points into noise or signal, given some application dependent metric. This is the purpose of Persistent Homology.

Definition 55 (Filtration) *Let $f : \mathcal{K} \to \mathbb{R}$ be an injective scalar field defined on a simplicial complex $\mathcal{K}$ such that $f(\tau) < f(\sigma)$ for each face τ of each simplex σ. Let n be the number of simplices of $\mathcal{K}$ and let $\mathcal{L}^-(i)$ be the sub-level set of f by the i^{th} value in the sorted set of simplex values. The nested sequence of subcomplexes $\mathcal{L}^-(0) \subset \mathcal{L}^-(1) \subset \cdots \subset \mathcal{L}^-(n-1) = \mathcal{K}$ is called the* filtration *of f.*

The general notion of filtration is preferred here to the more specific notion of lower star filtration (specifically adapted to PL scalar fields) as this general introduction will be useful in the next subsections.

Definition 56 (Homomorphism) *A* homomorphism *is a map between groups that commutes with the group operation.*

For instance, the group operation for the group of p-chains is the formal sum of p-simplices (see definition 34).

The filtration of a scalar field f induces a sequence of homomorphisms between the homology groups of the subcomplexes of $\mathcal{K}$:

$$H_p(\mathcal{L}^-(0)) \to H_p(\mathcal{L}^-(1)) \to \cdots \to H_p(\mathcal{L}^-(n-1)) = H_p(\mathcal{K}) \qquad (1.1)$$

Definition 57 (Persistent homology group) *The p^{th}* persistent homology groups *are the images of the homomorphisms induced by inclusion, noted $H_p^{i,j}$, for $0 \leq i \leq j \leq n-1$. The corresponding p^{th}* persistent Betti numbers *are the ranks of these groups, $\beta_p^{i,j} = rank(H_p^{i,j})$.*

Figure 1.14 provides a visual interpretation of the notion of 0^{th} persistent Betti number, which characterizes connected components. Given two nested sub-complexes $\mathcal{L}^-(i)$ and $\mathcal{L}^-(j)$, with $i < j$, the sub-complex induced by inclusion with regard to the 0^{th} homology group is noted $\mathcal{L}^-(i,j)$: it is defined by the connected components of $\mathcal{L}^-(j)$ which have non empty intersections with these of $\mathcal{L}^-(i)$ (which *includes* them). The 0^{th} homology group of $\mathcal{L}^-(i,j)$ is composed of the classes of the 0^{th} ho-

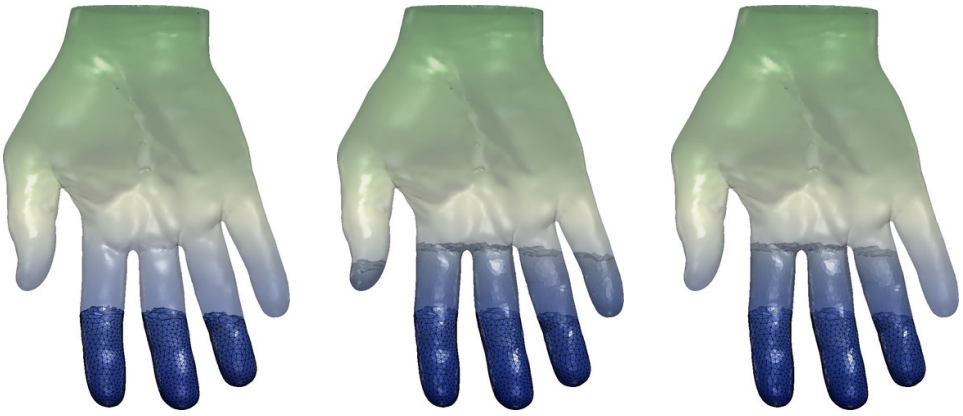

Figure 1.14 – *Sub-complexes induced by the filtration of a PL scalar field defined on a PL 3-manifold (dark blue: $\mathcal{L}^-(i)$, light blue: $\mathcal{L}^-(j)$). From left to right: $\beta_0(\mathcal{L}^-(i)) = 3$, $\beta_0(\mathcal{L}^-(j)) = 4$, $\beta_0(\mathcal{L}^-(i,j)) = 2$.*

mology group that already existed at the i^{th} isovalue and which still exist at the j^{th} isovalue.

In this example, only 2 of the 4 connected components of $\mathcal{L}^-(j)$ include connected components of $\mathcal{L}^-(i)$. Therefore, among the 3 connected components of $\mathcal{L}^-(i)$, only 2 of them are *persistent* in the interval $[i, j]$.

Therefore, persistent homology provides a mechanism to characterize the importance of topological features (here connected components) with regard to a specific measure (here a PL scalar field), at multiple scales (here the interval $[i, j]$). This interpretation can be generalized to other topological features (for PL 3-manifolds cycles and voids) by extending it to other Betti numbers.

The previous example illustrated the case where a class of the 0^{th} homology group (representing a connected component) disappeared in between the i^{th} and j^{th} isovalues: $\beta_0(\mathcal{L}^-(i)) = 3$, $\beta_0(\mathcal{L}^-(i,j)) = 2$. One can further track the precise isovalue where classes appear, disappear or merge with others. Such events correspond to a change in the Betti numbers of the sub-level set of the scalar field. As discussed in the previous subsection, these changes occur at critical points. Therefore, the notions of *birth* and *death* of classes of persistent homology groups can be associated with pairs of critical points of the input scalar field, and the absolute value of their f value difference is called the *persistence* of the pair.

In particular, the merge of two classes represent a *death* event. In such a case, the least persistent class (the youngest of the two) is chosen as the dying class. This choice is often called the *Elder's rule* (EH09). Once this is established, one can pair without ambiguity all the critical points of a scalar field. Note that this observation could already be foreseen with the

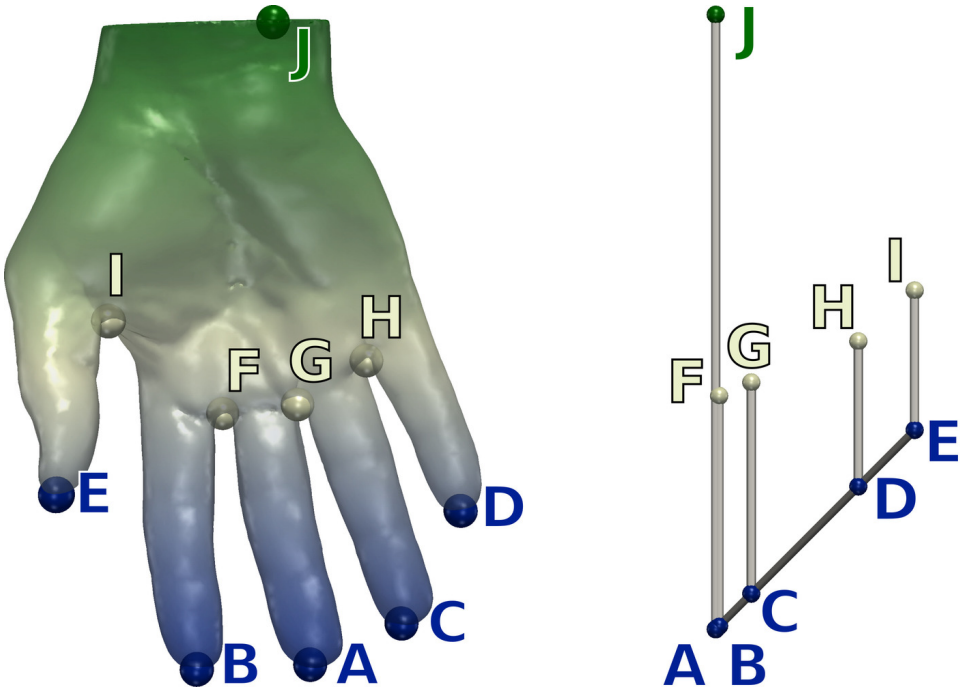

Figure 1.15 – *Critical points of a PL scalar field f defined on a PL 3-manifold (left) and its persistence diagram $\mathcal{D}(f)$ (right). In the diagram, each pair of critical points is represented by a white bar and its persistence is given by the height of the bar.*

Morse-Euler relation. For instance, the removal of an index 2 critical point (a maximum) of a PL Morse scalar field defined a PL 2-manifold implies the removal of a paired index 1 critical point (a saddle) in order to keep the Euler characteristic constant (property 11).

Therefore, it is possible to enumerate all the classes of all p^{th} homology group by enumerating the critical point pairs identified with the above strategy. This list of critical point pairs can be concisely encoded with a topological abstraction called the *Persistence Diagram* (Figure 1.15), noted $\mathcal{D}(f)$. This diagram is a one-dimensional simplicial complex that embeds each pair in $\mathbb{R}^2$ by using its *birth* value as a first component and its *birth* and *death* values as second components. The persistence diagram comes with several interesting properties. In particular, the stability theorem (CSEH05) states that given two PL scalar fields f and g defined on a common domain, the bottleneck distance between their persistence diagrams is bounded by the difference between the two functions with regard to the infinity norm: $d_B(\mathcal{D}(f), \mathcal{D}(g)) \leq ||f - g||_\infty$. Intuitively, this means that given a slight perturbation of a scalar field, its persistence diagram will only slightly vary. This stability result further motivates the usage of the persistence diagram as a stable topological abstraction of a scalar field (used for instance in function comparison).

In practice however, an alternate representation is often preferred to isolate critical point pairs corresponding to important features from these corresponding to noise. The *Persistence Curve*, noted $\mathcal{C}(f)$, is a diagram that plots the number of critical point pairs with persistence higher than a threshold ϵ, as a function of this threshold ϵ. When displayed in logarithmic scale, such curves often exhibit a flat plateau separating features with very low persistence from these of higher persistence (see Figure 1.13). In practice, such a plateau is instrumental to manually identify a relevant persistence threshold for the user-driven selection of the most important critical points of the field, as illustrated in Figure 1.13 (bottom). Also, note that extracting the critical point pairs more persistent than a threshold ϵ for increasing values of ϵ yields a hierarchy of sets of critical points, that enables to interactively explore them at multiple scales of importance, as showcased throughout Figure 1.13.

1.2.3 Reeb graph

The persistence curve and the persistence diagrams provide concise representations of the critical point pairs of a PL scalar field, along with their persistence. However, they do not provide any information related to the adjacency relations of these pairs on the domain. This is the purpose of more advanced topological abstractions, such as the Reeb graph (Ree46).

Definition 58

(Reeb graph) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL Morse scalar field defined on a compact PL manifold $\mathcal{M}$. Let $f^{-1}(f(p))_p$ be the contour of f containing the point $p \in \mathcal{M}$. The Reeb graph $\mathcal{R}(f)$ is a one-dimensional simplicial complex defined as the quotient space on $\mathcal{M} \times \mathbb{R}$ by the equivalence relation $(p_1, f(p_1)) \sim (p_2, f(p_2))$, which holds if:*

$$\begin{cases} f(p_1) = f(p_2) \\ p_2 \in \left(f^{-1}(f(p_1))\right)_{p_1} \end{cases}$$

The Reeb graph can also be defined alternatively as the *contour retract* of $\mathcal{M}$ under f (a continuous map that retracts each contour to a single point, such that its image is a subset of its domain and its restriction to its image is the identity). Note that f can be decomposed into $f = \psi \circ \phi$, where $\phi : \mathcal{M} \to \mathcal{R}(f)$ is the contour retraction and $\psi : \mathcal{R}(f) \to \mathbb{R}$ is a continuous function that maps points of $\mathcal{R}(f)$ to their f value in $\mathbb{R}$.

Intuitively, as suggested by the previous definition, the Reeb graph

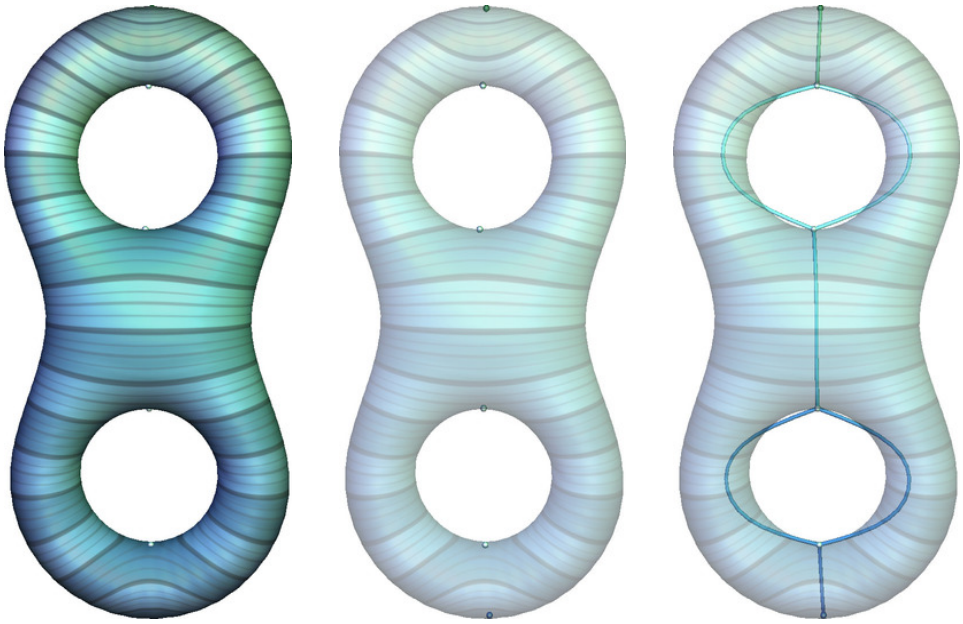

Figure 1.16 – *PL Morse scalar field defined on a PL 2-manifold (left and center) and its Reeb graph (right).*

continuously contracts each connected component of level sets to a point, yielding a one-dimensional simplicial complex that can be optionally embedded in $\mathbb{R}^3$, as illustrated in Figure 1.16.

Since the Betti numbers of the level sets change at critical points (in particular β_0 may change), the Reeb graph has a tight connection with the critical points of the function. In particular, branching occurs when $\beta_0(f^{-1}(i))$ changes as i evolves, as further detailed below:

Property 12 (Images through ϕ (Ree46)) *Let $\mathcal{R}(f)$ be the Reeb graph of a PL Morse scalar field $f = \psi \circ \phi$ defined on a PL d-manifold. Let the valence of a 0-simplex $v \in \mathcal{R}(f)$ be the number of 1-simplices in its star $St(v)$.*

- *All regular points of f map through ϕ to a point in the interior of a 1-simplex of $\mathcal{R}(f)$. The inverse is true.*
- *All critical points of index 0 or d (all extrema of f) map through ϕ to 0-simplices of $\mathcal{R}(f)$ of valence 1. The inverse is true.*
- *If $d = 2$, all critical points of index 1 (all saddles of f) map through ϕ to 0-simplices of $\mathcal{R}(f)$ of valence 2, 3 or 4. The inverse is true.*
- *If $d \geq 3$, all critical points of index 1 or $(d-1)$ (subsets of saddles of f) map through ϕ to 0-simplices of $\mathcal{R}(f)$ of valence 2 or 3. The inverse is not necessarily true.*
- *If $d > 3$, all critical points of index different from 0, 1, $(d-1)$ or d map*

through ϕ to 0-simplices of $\mathcal{R}(f)$ of valence 2. The inverse is not necessarily true.

The original description of the Reeb graph (Ree46) (including the above properties) implies that all critical points of f map to 0-simplices of $\mathcal{R}(f)$. In more contemporary descriptions, two 1-simplices sharing a valence-2 0-simplex as a face are considered to form only one 1-simplex. Therefore, in the contemporary vision of the Reeb graph, extrema map to valence-1 vertices while only the saddles where $\beta_0(f^{-1}(i))$ evolves map to vertices of higher valence. In particular, saddles where $\beta_0(f^{-1}(i))$ decreases (respectively increases) are called *join* (respectively *split*) saddles. For PL 3-manifolds, join (respectively split) saddles have index 1 (respectively 2). In this vision, not all 1 and 2-saddles map to vertices of the Reeb graph.

Since the Reeb graph has a tight connection with the critical points of its scalar field, some of the properties of PL Morse scalar fields translate in the Reeb graph setting.

Definition 59 (Loops in a Reeb graph) *Let $\mathcal{R}(f)$ be the Reeb graph of a PL Morse scalar field f defined on a PL d-manifold $\mathcal{M}$. Each independent cycle of $\mathcal{R}(f)$ is called a* loop. *The number of loops of a Reeb graph is noted $l(\mathcal{R}(f))$.*

The two saddles of each loop with the highest and lowest ψ values are usually called *loop saddles*.

Property 13 (Loops in a Reeb graph) *Let $\mathcal{R}(f)$ be the Reeb graph of a PL Morse scalar field f defined on a compact PL d-manifold $\mathcal{M}$. $l(\mathcal{R}(f))$ is bounded by $\beta_1(\mathcal{M})$:*

$l(\mathcal{R}(f)) \leq \beta_1(\mathcal{M})$

As discussed by Cole-McLaughlin et al. (CMEH*03), this property follows from the fact that the construction of the Reeb graph can lead to the removal of 1-cycles of $\mathcal{M}$, but not to the creation of new ones. In the case of PL 2-manifolds, tighter bounds have be shown (CMEH*03).

Property 14 (Loops in a Reeb graph on PL 2-manifolds) *Let $\mathcal{R}(f)$ be the Reeb graph of a PL Morse scalar field f defined on a PL 2-manifold $\mathcal{M}$. Let $b(\mathcal{M})$ be the number of connected components of $\mathcal{M}$ and $g(\mathcal{M})$ its genus. The number of loops of $\mathcal{R}(f)$ can be described as follows:*

- *If $\mathcal{M}$ is orientable (admits a non-null and continuous normal vector field):*
  - *if $b(\mathcal{M}) = 0$, then $l(\mathcal{R}(f)) = g(\mathcal{M})$;*
  - *otherwise $g(\mathcal{M}) \leq l(\mathcal{R}(f)) \leq 2g(\mathcal{M}) + b(\mathcal{M}) - 1$*

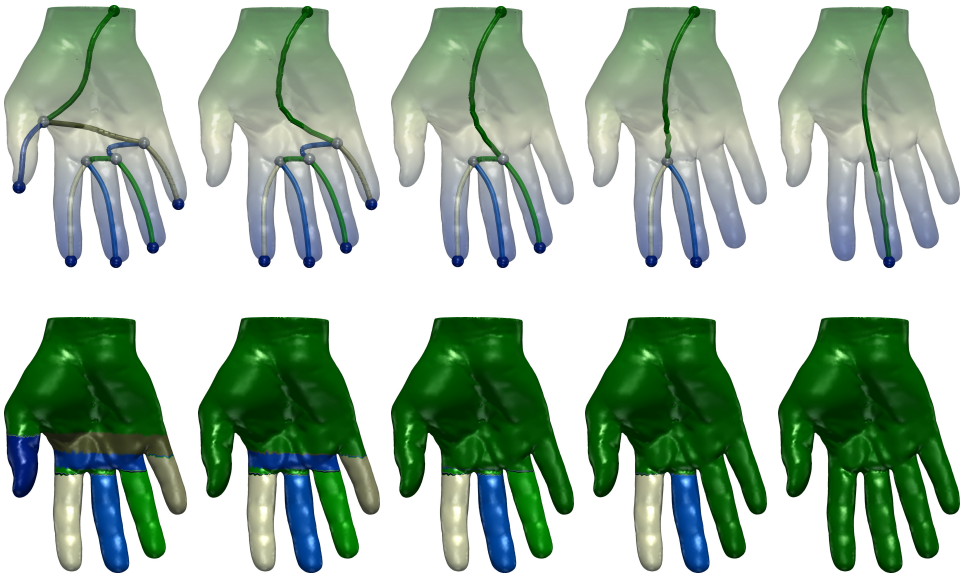

Figure 1.17 – *Hierarchy of Reeb graphs obtained by repeated persistence-driven removal of their* 1*-simplices (top) and hierarchy of data segmentations (bottom) obtained by considering the pre-image by* ϕ *of each* 1*-simplex of the Reeb graphs (matching colors).*

- *otherwise,*
  - *if* $b(\mathcal{M}) = 0$*, then* $0 \leq l(\mathcal{R}(f)) \leq g(\mathcal{M})/2$
  - *otherwise* $0 \leq l(\mathcal{R}(f)) \leq g(\mathcal{M}) + b(\mathcal{M}) - 1$

Figure 1.16 illustrates this property for a closed and orientable 2-manifold of genus 2 (here $\chi(\mathcal{M}) = 2 - 2g(\mathcal{M}) - b(\mathcal{M})$) and a Reeb graph having consequently two loops.

It follows from property 13 that the Reeb graph of a PL Morse scalar field defined on a simply-connected PL d-manifold $\mathcal{M}$ is loop free ($\beta_1(\mathcal{M}) = 0$). In this specific case, the Reeb graph is usually called a *Contour tree* and is noted $\mathcal{T}(f)$. Variants of the Reeb graph, called the *Join* (respectively *Split*) *trees* are defined similarly by contracting connected components of sub (respectively sur) level sets to points (instead of level sets) and are noted $\mathcal{J}(f)$ (respectively $\mathcal{S}(f)$). Note that the join (respectively split) tree of a PL Morse scalar field admitting only one maximum (respectively minimum) is equal to its contour tree.

Since the Reeb graph is a one-dimensional simplicial complex, a filtration of $\psi : \mathcal{R}(f) \to \mathbb{R}$ can be considered and therefore persistent homology concepts (previous subsection) readily apply to the Reeb graph without specialization. In particular, one can directly read the $(0, 1)$ critical point pairs of f (minima and join saddles) from the join tree by removing its 1-simplices attached to a minimum, one by one in order of their persistence, as illustrated in Figure 1.17 (top). A similar strategy applied to the

split tree enumerates all $(d-1, d)$ critical point pairs ($(d-1)$-saddles and maxima). The time-efficient enumeration of these types of critical point pairs is one of the primary applications of the Reeb graph in practice. Note that simplifying in such a way the Reeb graph, similarly to the critical points in the previous subsection, yields a hierarchy of Reeb graphs (for each of which a corresponding, simplified, PL Morse scalar field is guaranteed to exist), that enables to interactively explore it at multiple scales of importance, as showcased throughout Figure 1.17 (top).

Finally, note that the pre-image by ϕ of $\mathcal{R}(f)$ induces a complete partition of $\mathcal{M}$. In particular, the pre-image of a 1-simplex $\sigma \in \mathcal{R}(f)$ is guaranteed by construction to be connected (ϕ^{-1} is implemented in practice by marking during the construction of $\mathcal{R}(f)$ each vertex with the identifier of the 1-simplex where it maps to). This latter property is instrumental in various tasks in scientific visualization, including the efficient indexing of contours (for fast level set extraction) or the automatic and interactive data segmentation into regions of interest (especially when feature boundaries coincide with level sets). This latter capability of the Reeb graph can be nicely combined with persistent homology concepts, yielding hierarchies of data segmentations, as illustrated in Figure 1.17 (bottom).

1.2.4 Morse-Smale complex

As discussed in the previous subsection, in the modern interpretation of the Reeb graph, not all critical points are captured as 0-simplices. Therefore, the Reeb graph only describes the adjacency relations of a sub-set of critical point pairs. To capture such exhaustive adjacency relations, one needs to consider another topological abstractions, called the Morse-Smale complex, that is constructed by considering equivalence classes on integral lines instead of contours (see (Gyu08) for further details).

Property 15 (Integral lines) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL Morse scalar field defined on a closed PL d-manifold $\mathcal{M}$. Then, the following properties hold:*

- *Two integral lines are either disjoint or the same;*
- *Integral lines cover all of $\mathcal{M}$;*
- *The origin and the destination of an integral line are critical points of f.*

The latter property is particularly interesting. It means that an integral line can be characterized by its extremities, which are guaranteed to be critical points of f. Then, one can introduce an equivalence relation that

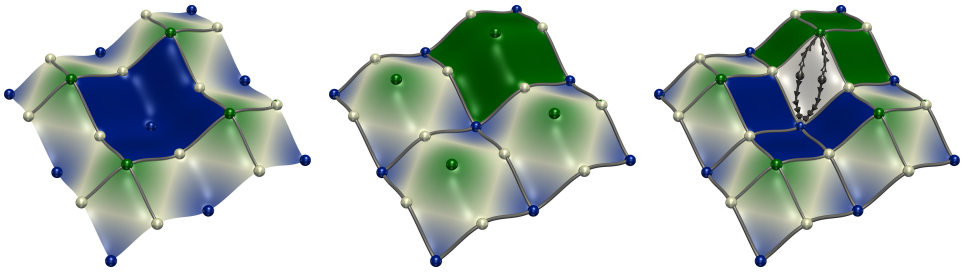

Figure 1.18 – *Ascending (left) and descending (center) manifolds and Morse-Smale complex (right) of a PL Morse scalar field f defined on a PL 2-manifold.*

holds if two integral lines share the same extremities. This is the key idea behind the Morse-Smale complex.

Definition 60 (Ascending manifold) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL Morse scalar field defined on a PL d-manifold $\mathcal{M}$. The* ascending *(respectively* descending*)* manifold *of a critical point p of f is the set of points belonging to integral lines whose origin (respectively destination) is p.*

Property 16 (Ascending manifolds) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL Morse scalar field defined on a PL d-manifold $\mathcal{M}$. Let p be an index-i critical point of f. The ascending (respectively descending) manifold of p is an open set of $\mathcal{M}$ of dimension $(d - i)$ (respectively i).*

Figure 1.18 illustrates these properties with a PL Morse scalar field defined on a PL 2-manifold $\mathcal{M}$. In particular, the ascending manifold of a minimum is a subset of $\mathcal{M}$ of dimension 2 (shown in dark blue, left). Similarly, the descending manifold of a maximum is also a subset of $\mathcal{M}$ of dimension 2 (show in green, center). For PL 2-manifolds, in both cases, ascending and descending manifolds of saddles have dimension 1 (grey integral lines in both images).

Definition 61 (Morse complex) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL Morse scalar field defined on a PL d-manifold $\mathcal{M}$. The complex formed by all descending manifolds of f is called the* Morse complex.

Given this definition, the complex of all ascending manifolds shown on the left of Figure 1.18 is the Morse complex of $-f$, while the complex of all descending manifolds shown in the center is that of f.

Definition 62 (Morse-Smale function) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL Morse scalar field defined on a PL d-manifold $\mathcal{M}$. f is a* Morse-Smale *function if the ascending and descending manifolds only intersect transversally.*

Intuitively, the transversal intersection condition implies that ascend-

ing and descending manifolds are not parallel at their intersection (this condition is enforced in practice through local remeshing). This implies that when these intersect exactly at one point, such a point is critical. This also implies that given an integral line, the index of its origin is smaller than that of its destination.

Definition 63 (Morse-Smale complex) *Let $f : \mathcal{M} \to \mathbb{R}$ be a PL Morse scalar field defined on a PL d-manifold $\mathcal{M}$. The complex formed by the intersection of the Morse complex of f and that of $-f$ is called the* Morse-Smale *complex and noted $\mathcal{MS}(f)$.*

Figure 1.18 (right) illustrates such an intersection. As shown with the white region, all integral lines (black curves) of a given cell of the complex (irrespectively of its dimension) share the same origin and destination. Note that one can derive a simplicial decomposition of the Morse-Smale complex by subdividing each d-dimensional cell into valid d simplices. By construction, all the critical points of f will therefore map to distinct 0-simplices of such a decomposition, since the Morse-Smale complex captures all ascending and descending manifolds (and therefore all critical points). Then, as for the Reeb graph (previous subsection), persistent homology concepts also readily apply to the simplicial decomposition of the Morse-Smale complex and simplifying it for increasing values of persistence also yields a hierarchy that enables to interactively explore the Morse-Smale complex at multiple scales of importance.

Finally, by construction, the Morse-Smale complex provides a partition of the domain that is instrumental in scientific visualization, especially when features or their boundaries coincide with the gradient. Alike the Reeb graph, this segmentation capability can be nicely combined with persistent homology concepts, yielding hierarchies of data segmentations, as detailed in the following subsection.

1.3 Algorithms and applications

In this section, I briefly discuss the state-of-the art algorithms for computing the topological abstractions described above, and I also briefly introduce some of their applications. Reference open source implementations of these algorithms can be found in the Topology ToolKit library (TFL*17).

Persistent homology

Critical points of PL scalar fields are usually extracted with a simple, robust, localized yet globally consistent, and easily parallelizable algorithm that directly implements the definitions presented in Section 1.2.1, and which derive from a seminal paper by Banchoff (Ban70).

Critical point pair extraction as well as their persistence evaluation (Section 1.2.2) are usually implemented through sparse matrix reduction (ELZ02) with an algorithm with $O(n^3)$ worst case time complexity (where n is the number of simplices). Note however, that for the purpose of feature selection, only extrema-saddle pairs seem to have a practical interest and these can be computed more efficiently with the Reeb graph as described below.

Persistence diagrams (which encode all critical point pairs along with their persistence) have been widely used for the purpose of function comparison, especially for high-dimensional domains where more advanced topological abstractions are more difficult to compute and simplify (see for instance (Ghr07), (CCSG*09) and (RL15)).

Cohen-Steiner et al. (CSEH05) showed that the bottleneck distance between the persistence diagrams of two PL scalar fields f and g computed on a common domain was bounded by the distance between the two functions with regard to the infinity norm ($||f-g||_\infty$). This result raises the reciprocal question: *given a persistence diagram $\mathcal{D}(f)$ where all pairs less persistent than a threshold ϵ have been removed (noted $\mathcal{D}(g)$), can we compute a function g sufficiently close from f that admits $\mathcal{D}(g)$ as persistence diagram?* This question has major practical implications since the time complexity of the algorithms for the construction or processing of topological abstractions is often dictated by the number of critical points in the input scalar field. Often in practice, it is possible to easily discriminate critical points that are not relevant application-wise. Therefore, there exists an applicative interest for an efficient pre-processing of an input scalar field, that would minimally perturb it to remove a given set of critical points. This question has first been addressed in the case of PL scalar fields defined on PL 2-manifolds by Edelsbrunner et al. (EMP06), who showed that such a function g existed and that its difference to the input was bounded by ϵ: $||f-g||_\infty \leq \epsilon$. These authors also provided an algorithm to compute it. However this algorithm is complicated and difficult to implement. Moreover, as persistence pairs are processed in order of their highest extremity, the same vertices are swept several times when canceable persistence pairs

are nested. Attali et al. (AGH*09) and Bauer et al. (BLW12) presented independently a similar approach to this problem for filtrations and Discrete Morse functions. However, converting the output of these algorithms to PL scalar fields (which is the standard scalar field representation for many applications) requires an important subdivision of the domain. Also, these approaches only deal with closed surfaces. We introduced in 2012 a general algorithm (TP12) for the topological simplification of PL scalar fields on closed or open PL 2-manifolds, capable of removing arbitrary critical point pairs (not necessarily the least persistent), which enables the usage of application-dependent metrics for feature selection. Thanks to its speed, ease of implementation, robustness and generality, we consider this algorithm as the reference for the problem of topological simplification of scalar data on surfaces.

A survey on the concepts of Persistent homology and their applications can be found in (EH08).

Reeb graph

Reeb graphs have been introduced in Computer Science independently by Boyell and Ruston (BR63) (for simply connected surfaces) and by Shinagawa et al. (SKK91).

Efficient algorithms for their computation have first been investigated in the simpler cases of simply connected domains (for which the Reeb graph is called the contour tree). Time-efficient algorithms have first been introduced for 2D domains (vKvOB*97), then 3D domains (TV98) and last for domains of arbitrary dimension, with an algorithm (CSA00) with optimal time complexity: $O(|\sigma_0|log(|\sigma_0|) + (\sum_{i=0}^{i=1}|\sigma_i|)\alpha(\sum_{i=0}^{i=1}|\sigma_i|))$, where $|\sigma_i|$ is the number of i-simplices in the domain and $\alpha()$ is an extremely slowly growing function (i.e. the inverse of the Ackermann function). In particular, the latter algorithm is considered by the community as the reference for the problem of contour tree computation thanks to its optimal time complexity, its ease of implementation, its robustness, and its practical performances. Sequential (Dil07) and multi-threaded (GFJT16, GFJT17) implementations are also available. This algorithm first computes the join and split trees by tracking the merge events of a union-find data-structure processing the vertices in ascending (respectively descending) order. Last, the two trees are combined to form the contour tree in a linear pass.

Regarding more general domains, an algorithm (CMEH*03) has been introduced for PL scalar fields defined on PL 2-manifolds, with optimal

time complexity: $O(|\sigma_1|log(|\sigma_1|)$. A more recent, non-optimal algorithm (PSF08) would be recommend instead however, due to its ease of implementation and acceptable performances in practice. This algorithm explicitly constructs the partition induced by the pre-image of ϕ to retrieve the 1-simplices of $\mathcal{R}(f)$, by computing the critical contours of all saddles of f as boundaries. This makes the algorithm output-sensitive but leads to a worst-case complexity of $O(|\sigma_0| \times |\sigma_2|)$.

Regarding higher dimensional (non simply-connected) domains, several attempts have been proposed, especially with the more practical target of PL 3-manifolds in mind. However, many properties of the 2-dimensional domains exploited by the above algorithms do not hold for 3-dimensional domains, making the problem more challenging. Pascucci et al. (PSBM07) introduced the first algorithm for the computation of the Reeb graph on domains of arbitrary dimension. Its streaming nature however, while appealing in specific applications, makes it implementation difficult. An open-source implementation, which I wrote in 2009 based on Giorgio Scorzelli's original implementation, is available in the official release of the open-source library the Visualization ToolKit (Tie09). Doraiswamy and Natarajan (DN08) extended the quadratic complexity algorithm by Patane et al (PSF08) from PL 2-manifolds to PL 3-manifolds. We introduced the first practical algorithm (TGSP09) for the efficient computation of the Reeb graph of PL scalar fields defined on PL 3-manifolds in 2009. Thanks to its speed, it enabled in practice to transfer all of the contour tree based interactive applications to more general non-simply connected domains. We considered this algorithm as the reference for the problem of Reeb graph computation on PL 3-manifolds, until a dimension-independent, optimal time complexity ($O((\sum_{i=0}^{i=2}|\sigma_i|)log(\sum_{i=0}^{i=2}|\sigma_i|))$) algorithm was introduced three years later (Par12). Algorithms for the practical computation of the bivariate analog of the Reeb graph, called the Reeb space (EHP08), have only been investigated recently (CGT*15, KTCG16, TC16).

The contour tree and the Reeb graph have been massively applied in scientific visualization, in particular because of the property that the pre-image by ϕ of a 1-simplex $\sigma \in \mathcal{R}(f)$ is guaranteed by construction to be connected. This enables for instance to extract an optimal number of vertex seeds for optimal time level set extraction: each vertex seed initiates the construction of a contour by breadth-first search traversal of the domain (limiting the traversal to the exact set of simplices projecting on the queried isovalue). This seed extraction process requires to store each

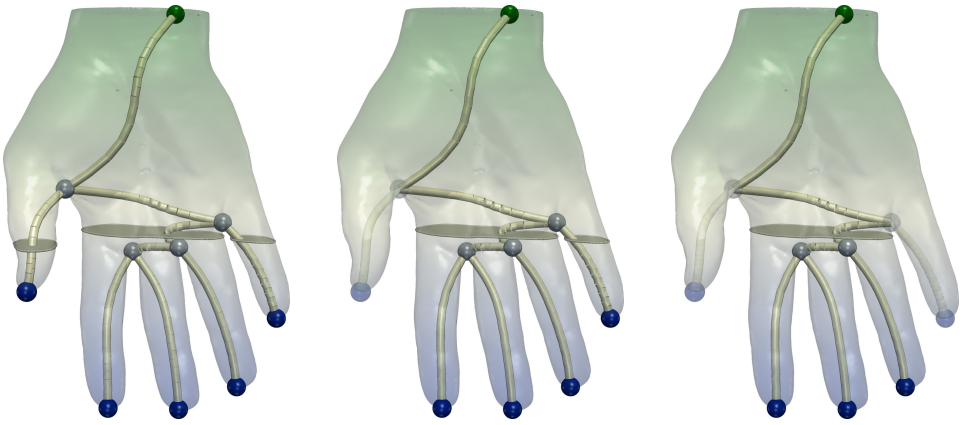

Figure 1.19 – *Topological simplification of isosurfaces (white surface, middle). The* 1*-simplices of* $\mathcal{R}(f)$ *that are less persistent than an increasing threshold (transparent, from left to right) are not considered for contour seed extraction, yielding a progressive removal of the least prominent contours from the isosurface.*

1-simplex of $\mathcal{R}(f)$ in a balanced interval tree (CLRS09). At query time, given an isovalue i, all 1-simplices of $\mathcal{R}(f)$ projecting on i can be efficiently retrieved in $O(|\sigma_1|log(|\sigma_1|))$ steps, where $|\sigma_1|$ is here the number of 1-simplices in $\mathcal{R}(f)$. Further, given a 1-simplex $\sigma \in \mathcal{R}(f)$ that projects on i, a seed vertex can be efficiently extracted if the vertices of the domain projecting to σ through ϕ (let $|\sigma_0'|$ be their number) are stored in a balanced search tree in $O(|\sigma_0'|log(|\sigma_0'|))$ steps. This mechanism can be nicely combined with persistent homology concepts to interactively simplify isocontours, as illustrated in Figure 1.19, by only considering the 1-simplices of $\mathcal{R}(f)$ that are more persistent than a threshold ϵ. Variants of this strategy have been presented in the case of the contour tree by van Kreveld et al. (vKvOB*97) and Carr et al. (CSvdP04). In particular, the latter approach introduced, as an alternative to persistence, several geometrical measures enabling to filter the 1-simplices of $\mathcal{T}(f)$ according to more application-relevant metrics, which reveals to be of major importance in practice.

The partitioning capabilities of the Reeb graph have also been instrumental in scientific visualization for data segmentation tasks, especially in cases where the boundaries of regions of interest coincide with level sets. In that context, the Reeb graph enables (with the fast isosurface extraction algorithm presented above) to rapidly extract and distinguish each of these boundaries, at multiple scales of importance when combined with persistent homology mechanisms. These segmentation capabilities also serve as the basis of more advanced techniques, for feature segmentation (for example in molecular chemistry (GABCG*14)), for the tracking of fea-

tures over time (SB06) for example in combustion data analysis (BWT*11), for the design of transfer functions in volume rendering (WDC*07) or the similarity estimation between data features (TN14).

Apart from scientific visualization, the Reeb graph has also been used as a core data-structure in computer graphics, as discussed in (BGSF08, Tie08, VDL*17).

Morse-Smale complex

The computation of Morse-Smale complexes was first investigated in the case of PL 2-manifolds. An initial algorithm was introduced by Edelsbrunner et al. (EHZ03). This algorithm constructs for each saddle of f the integral lines originating and terminating at the saddle, yielding the set of ascending and descending 1-manifolds. Ascending and descending 2-manifolds are then retrieved through breadth-first search, by growing 2-dimensional regions until ascending or descending 1-manifolds are attained. Further, the authors apply persistent homology concepts to remove the least persistent critical point pairs and describe a mechanism to update the Morse-Smale complex accordingly (by removing ascending and descending 1-manifolds attached to removed critical points, merging the adjacent 2-manifolds and re-routing their 1-manifold boundaries). Since the number of saddles of f can be proportional to the number of vertices in the domain and since the integral lines can intersect an number of triangles which is proportional to that of the domain, the construction algorithm has a worst case time complexity of $O(|\sigma_0| \times |\sigma_2|)$. This algorithm was latter improved and applied for the first time to scientific visualization by Bremer et al. (BEHP03).

The problem of computing the Morse-Smale complex in higher dimensions is far more challenging. First, as the dimension increases, new types of critical points appear (of increasing index), which translates into the apparition of ascending and descending manifolds of higher and higher dimensions. The construction of each of these types of manifolds implies a quadratic term in the runtime complexity. Second, degenerate cases become more challenging to resolve. This is in particular the case of the resolution of the degenerate critical points and the enforcement of the transversal intersection condition. A complicated algorithm has been proposed for PL scalar fields defined on PL 3-manifolds (EHNP03) but seems highly challenging to implement and no implementation has been reported.

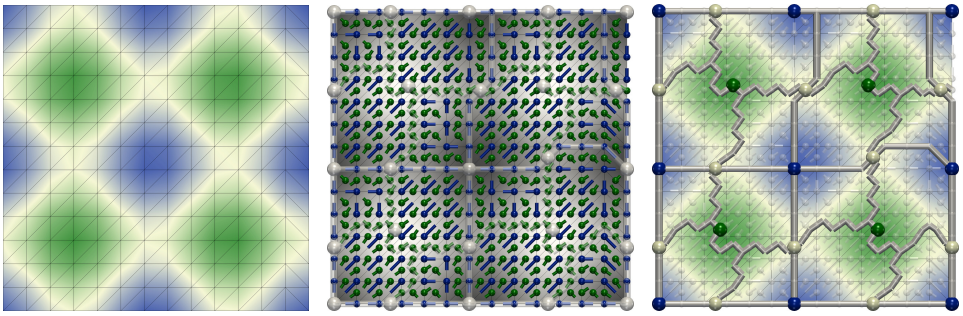

Figure 1.20 – *Simple synthetic function (left) and its discrete gradient field (center). Blue glyphs illustrate the pairing of 0-cells (vertices) with 1-cells (edges) while green glyphs show the pairing of 1-cells (edges) with 2-cells (faces). The i-cells that are left unpaired by the discrete gradient field correspond to the critical points of the discrete Morse function, shown with blue ($d = 0$), white ($d = 1$) and green ($d = 2$) spheres (right).* V-paths *(discrete analogs of integral lines) between these critical points represent the separatrices of the Morse-Smale complex (grey cylinders, right).*

Recently, efficient algorithms have been introduced by walking around the degeneracies of the PL setting and considering a competing formalism, Discrete Morse Theory (Foro1), which comes with many nice combinatorial properties. In this setting for example, critical points occur on simplices of arbitrary dimension and the index of a critical point coincides with the dimension of its simplex. Moreover, degenerate critical points cannot occur by construction. This formalism considers as *Discrete Morse Functions* scalar fields that map each simplex of the domain to a single scalar value, in such a way that for each simplex σ, there exists at most one simplex for which σ is a face with lower function value, and that there exists at most one simplex that is a face of σ with a higher function value. Then a simplex is critical if no such face and no such coface exist. This formalism additionally introduces the notion of *V-path*, used as an analog to PL integral lines, which is a sequence of pairs of simplices of alternating dimensions (d and $(d+1)$) with ascending function values. Then, the notion of *discrete gradient* can be introduced as a pairing of simplices that induces V-paths that are all monotonic and loop-free (Figure 1.20). Given a discrete gradient, critical points correspond to simplices that belong to no simplex pair. As mentioned before, degenerate critical points cannot occur by construction in this setting. Moreover, ascending and descending manifolds are guaranteed to intersect transversally (Gyu08). Therefore, by construction, this formalism avoids all the degeneracies found in the PL setting which make Morse-Smale complex computation challenging. Thus, several efficient algorithms have been proposed for the computation of a discrete gradient from a PL scalar field as well as the construction and

simplification of the Morse-Smale complex (GBHP08, RWS11), including a shared-memory parallel algorithm (SN12) whose implementation has been released in open-source (Shi12). Note that this discrete formalism has been described here in the context of PL manifolds for consistency, but it readily applies to arbitrary cellular complexes. This is another factor that motivates its popularity, as regular grids no longer need to undergo simplicial subdivisions in this framework.

The combinatorial consistency of the Discrete Morse Theory setting comes however with a price in terms of geometrical accuracy: in particular, locally, V-paths have to follow simplices of the domain and therefore do not match the integral lines induced by any interpolant. Gyulassy et al. (GBP12) addressed this issue by introducing a novel, probabilistic, discrete gradient construction algorithm whose V-paths are shown to converge to integral lines as the domain sampling increases. We recently improved this approach to avoid the need for domain re-sampling by introducing a discrete gradient construction algorithm that conforms to input constraints (GGL*14). When used with integral lines computed through numerical approximation, this gradient construction algorithm yields Morse-Smale complexes whose manifolds better align with the gradient induced by the domain interpolant.

Morse-Smale complexes have been a popular topological abstraction for data analysis and visualization, especially for data segmentation tasks in applications where features of interest (or their boundaries) coincide with the gradient. Then such features can be efficiently captured at multiple scales of importance (thanks to persistent homology mechanisms) by considering the cells of the Morse-Smale complex. This overall strategy has been applied with tailored analysis algorithms to various applications, including the analysis of the Rayleigh-Taylor instability (LBM*06), vortical structures (KRHH11), porous media (GND*07), cosmology (Sou11), combustion (GBG*14), computational fluid dynamics (COH*13, FGT16, LAS*17) chemistry (GABCG*14), etc. A recent survey (DFFIM15) provides further details regarding the construction, simplification and application of the Morse-Smale complex.

Index

Bibliography

[AGH*09] Attali D., Glisse M., Hornus S., Lazarus F., Morozov D.: Persistence-sensitive simplification of functions on surfaces in linear time. In *TopoInVis Workshop* (2009). (Cited page 32.)

[Ban70] Banchoff T. F.: Critical points and curvature for embedded polyhedral surfaces. *The American Mathematical Monthly* (1970). (Cited pages 20 and 31.)

[BEHP03] Bremer P.-T., Edelsbrunner H., Hamann B., Pascucci V.: A multi-resolution data structure for 2-dimensional Morse functions. In *IEEE Visualization* (2003), pp. 139–146. (Cited page 35.)

[BGSF08] Biasotti S., Giorgio D., Spagnuolo M., Falcidieno B.: Reeb graphs for shape analysis and applications. *Theoretical Computer Science* (2008). (Cited page 35.)

[BLW12] Bauer U., Lange C., Wardetzky M.: Optimal topological simplification of discrete functions on surfaces. *Discrete and Computational Geometry* (2012), 347–377. (Cited page 32.)

[BR63] Boyell R. L., Ruston H.: Hybrid techniques for real-time radar simulation. In *Proc. of the IEEE Fall Joint Computer Conference* (1963). (Cited page 32.)

[BWT*11] Bremer P., Weber G., Tierny J., Pascucci V., Day M., Bell J.: Interactive exploration and analysis of large scale simulations using topology-based data segmentation. *IEEE Transactions on Visualization and Computer Graphics* (2011). (Cited page 35.)

[CCSG*09] Chazal F., Cohen-Steiner D., Guibas L. J., Memoli F., Oudot S. Y.: Gromov–hausdorff stable signatures for shapes using persistence. *Computer Graphics Forum (Proc. of SGP)* (2009). (Cited page 31.)

[CGT*15] Carr H., Geng Z., Tierny J., Chattopadhyay A., Knoll A.: Fiber surfaces: Generalizing isosurfaces to bivariate data. *Computer Graphics Forum (Proc. of EuroVis)* (2015). (Cited page 33.)

[CLRS09] Cormen T., Leiserson C. E., Rivest R. L., Stein C.: *Introduction to Algorithms*. MIT Press, 2009. (Cited page 34.)

[CMEH*03] Cole-McLaughlin K., Edelsbrunner H., Harer J., Natarajan V., Pascucci V.: Loops in Reeb graphs of 2-manifolds. In *Proc. of ACM Symposium on Computational Geometry* (2003), pp. 344–350. (Cited pages 26 and 32.)

[COH*13] Chen F., Obermaier H., Hagen H., Hamann B., Tierny J., Pascucci V.: Topology analysis of time-dependent multi-fluid data using the reeb graph. *Computer Aided Geometric Design* (2013). (Cited page 37.)

[CSA00] Carr H., Snoeyink J., Axen U.: Computing contour trees in all dimensions. In *Proc. of Symposium on Discrete Algorithms* (2000), pp. 918–926. (Cited page 32.)

[CSEH05] Cohen-Steiner D., Edelsbrunner H., Harer J.: Stability of persistence diagrams. In *Proc. of ACM Symposium on Computational Geometry* (2005). (Cited pages 23 and 31.)

[CSvdP04] Carr H., Snoeyink J., van de Panne M.: Simplifying flexible isosurfaces using local geometric measures. In *Proc. of IEEE VIS* (2004), pp. 497–504. (Cited page 34.)

[DFFIM15] De Floriani L., Fugacci U., Iuricich F., Magillo P.: Morse complexes for shape segmentation and homological analysis: discrete models and algorithms. *Computer Graphics Forum* (2015). (Cited page 37.)

[Dil07] Dillard S.: A contour tree library. `http://graphics.cs.ucdavis.edu/~sdillard/libtourtre/doc/html/`, 2007. (Cited page 32.)

[DN08] Doraiswamy H., Natarajan V.: Efficient output sensitive construction of reeb graphs. In *International Symposium on Algorithms and Computation* (2008). (Cited page 33.)

[EH08] Edelsbrunner H., Harer J.: Persistent homology – a survey. *American Mathematical Society* (2008). (Cited page 32.)

[EH09] Edelsbrunner H., Harer J.: *Computational Topology: An Introduction*. American Mathematical Society, 2009. (Cited pages 3, 19, and 22.)

[EHNP03] Edelsbrunner H., Harer J., Natarajan V., Pascucci V.: Morse-smale complexes for piecewise linear 3-manifolds. In *Proc. of ACM Symposium on Computational Geometry* (2003). (Cited page 35.)

[EHP08] Edelsbrunner H., Harer J., Patel A. K.: Reeb spaces of piecewise linear mappings. In *Proc. of ACM Symposium on Computational Geometry* (2008). (Cited page 33.)

[EHZ03] Edelsbrunner H., Harer J., Zomorodian A.: Hierarchical morse-smale complexes for piecewise linear 2-manfiolds. *Discrete and Computational Geometry* (2003). (Cited page 35.)

[ELZ02] Edelsbrunner H., Letscher D., Zomorodian A.: Topological persistence and simplification. *Discrete & Computational Geometry 28* (2002), 511–533. (Cited page 31.)

[EM90] Edelsbrunner H., Mucke E. P.: Simulation of simplicity: a technique to cope with degenerate cases in geometric algorithms. *ACM Transactions on Graphics 9* (1990), 66–104. (Cited page 14.)

[EMP06] Edelsbrunner H., Morozov D., Pascucci V.: Persistence-sensitive simplification of functions on 2-manifolds. In *Proc. of ACM Symposium on Computational Geometry* (2006), pp. 127–134. (Cited page 31.)

[FGT16] Favelier G., Gueunet C., Tierny J.: Visualizing ensembles of viscous fingers. In *IEEE SciVis Contest* (2016). (Cited page 37.)

[For01] Forman R.: A user's guide to discrete morse theory. In *Proc. of the International Conference on Formal Power Series and Algebraic Combinatorics* (2001). (Cited page 36.)

[GABCG*14] Guenther D., Alvarez-Boto R., Contreras-Garcia J., Piquemal J.-P., Tierny J.: Characterizing molecular inter-

actions in chemical systems. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)* (2014). (Cited pages 34 and 37.)

[GBG*14] Gyulassy A., Bremer P., Grout R., Kolla H., Chen J., Pascucci V.: Stability of dissipation elements: A case study in combustion. *Computer Graphics Forum (Proc. of EuroVis)* (2014). (Cited page 37.)

[GBHP08] Gyulassy A., Bremer P. T., Hamann B., Pascucci V.: A practical approach to morse-smale complex computation: Scalability and generality. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)* (2008). (Cited page 37.)

[GBP12] Gyulassy A., Bremer P. T., Pascucci V.: Computing morse-smale complexes with accurate geometry. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)* (2012). (Cited page 37.)

[GFJT16] Gueunet C., Fortin P., Jomier J., Tierny J.: Contour forests: Fast multi-threaded augmented contour trees. In *IEEE LDAV* (2016). (Cited page 32.)

[GFJT17] Gueunet C., Fortin P., Jomier J., Tierny J.: Task-based augmented merge trees with Fibonacci heaps. In *Proc. of IEEE Symposium on Large Data Analysis and Visualization* (2017). (Cited page 32.)

[GGL*14] Gyulassy A., Guenther D., Levine J. A., Tierny J., Pascucci V.: Conforming morse-smale complexes. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)* (2014). (Cited page 37.)

[Ghr07] Ghrist R.: Barcodes: The persistent topology of data. *American Mathematical Society* (2007). (Cited page 31.)

[GND*07] Gyulassy A., Natarajan V., Duchaineau M., Pascucci V., Bringa E., Higginbotham A., Hamann B.: Topologically Clean Distance Fields. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS) 13* (2007), 1432–1439. (Cited page 37.)

[Gyu08] Gyulassy A.: *Combinatorial Construction of Morse-Smale Complexes for Data Analysis and Visualization*. PhD thesis, University of California at Davis, 2008. (Cited pages 28 and 36.)

[KRHH11] Kasten J., Reininghaus J., Hotz I., Hege H.: Two-dimensional time-dependent vortex regions based on the acceleration magnitude. *IEEE Transactions on Visualization and Computer Graphics* (2011). (Cited page 37.)

[KTCG16] Klacansky P., Tierny J., Carr H., Geng Z.: Fast and exact fiber surfaces for tetrahedral meshes. *IEEE Transactions on Visualization and Computer Graphics* (2016). (Cited page 33.)

[LAS*17] Lukasczyk J., Aldrich G., Steptoe M., Favelier G., Gueunet C., Tierny J., Maciejewski R., Hamann B., Leitte H.: Viscous fingering: A topological visual analytic approach. In *Physical Modeling for Virtual Manufacturing Systems and Processes* (2017). (Cited page 37.)

[LBM*06] Laney D. E., Bremer P., Mascarenhas A., Miller P., Pascucci V.: Understanding the structure of the turbulent mixing layer in hydrodynamic instabilities. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)* (2006). (Cited page 37.)

[Mil63] Milnor J.: *Morse Theory*. Princeton U. Press, 1963. (Cited page 17.)

[Par12] Parsa S.: A deterministic o(m log m) time algorithm for the reeb graph. In *Proc. of ACM Symposium on Computational Geometry* (2012). (Cited page 33.)

[PSBM07] Pascucci V., Scorzelli G., Bremer P. T., Mascarenhas A.: Robust on-line computation of Reeb graphs: simplicity and speed. *TOG 26* (2007), 58.1–58.9. (Cited page 33.)

[PSF08] Patane G., Spagnuolo M., Falcidieno B.: Reeb graph computation based on a minimal contouring. In *Proc. of IEEE Shape Modeling International* (2008). (Cited page 33.)

[Ree46] Reeb G.: Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. *Comptes-rendus de l'Académie des Sciences 222* (1946), 847–849. (Cited pages 24, 25, and 26.)

[RL15] Rieck B., Leitte H.: Persistent homology for the evaluation of dimensionality reduction schemes. *Computer Graphics Forum (Proc. of EuroVis)* (2015). (Cited page 31.)

[RWS11] Robins V., Wood P., Sheppard A.: Theory and algorithms for constructing discrete morse complexes from grayscale digital images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2011). (Cited page 37.)

[SB06] Sohn B. S., Bajaj C. L.: Time varying contour topology. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)* (2006). (Cited page 35.)

[Shi12] Shivashankar N.: Parallel implementation of 3d morse-smale complex computation. `http://vgl.serc.iisc.ernet.in/mscomplex/`, 2012. (Cited page 37.)

[SKK91] Shinagawa Y., Kunii T., Kergosien Y. L.: Surface coding based on morse theory. *IEEE Computer Graphics and Applications* (1991). (Cited page 32.)

[SN12] Shivashankar N., Natarajan V.: Parallel computation of 3d morse-smale complexes. *Computer Graphics Forum (Proc. of EuroVis)* (2012). (Cited page 37.)

[Sou11] Sousbie T.: The persistent cosmic web and its filamentary structure: Theory and implementations. *Royal Astronomical Society* (2011). (Cited page 37.)

[TC16] Tierny J., Carr H.: Jacobi fiber surfaces for bivariate Reeb space computation. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)* (2016). (Cited page 33.)

[TFL*17] Tierny J., Favelier G., Levine J. A., Gueunet C., Michaux M.: The Topology ToolKit. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)* (2017). `https://topology-tool-kit.github.io/`. (Cited page 30.)

[TGSP09] Tierny J., Gyulassy A., Simon E., Pascucci V.: Loop surgery for volumetric meshes: Reeb graphs reduced to contour trees. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS) 15* (2009), 1177–1184. (Cited page 33.)

[Tie08] Tierny J.: *Reeb graph based 3D shape modeling and applications*. PhD thesis, Lille1 University, 2008. (Cited page 35.)

[Tie09] Tierny J.: vtkReebGraph class collection. `http://www.vtk.org/doc/nightly/html/classvtkReebGraph.html`, 2009. (Cited page 33.)

[TN14] Thomas D. M., Natarajan V.: Multiscale symmetry detection in scalar fields by clustering contours. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)* (2014). (Cited page 35.)

[TP12] Tierny J., Pascucci V.: Generalized topological simplification of scalar fields on surfaces. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)* (2012). (Cited page 32.)

[TV98] Tarasov S., Vyali M.: Construction of contour trees in 3d in o(n log n) steps. In *Proc. of ACM Symposium on Computational Geometry* (1998). (Cited page 32.)

[VDL*17] Vintescu A., Dupont F., Lavoué G., Memari P., Tierny J.: Conformal factor persistence for fast hierarchical cone extraction. In *Eurographics (short papers)* (2017). (Cited page 35.)

[vKvOB*97] van Kreveld M., van Oostrum R., Bajaj C., Pasucci V., Schikore D.: Contour trees and small seed sets for isosurface traversal. In *Proc. of ACM Symposium on Computational Geometry* (1997). (Cited pages 32 and 34.)

[WDC*07] Weber G., Dillard S. E., Carr H., Pascucci V., Hamann B.: Topology-controlled volume rendering. *IEEE Transactions on Visualization and Computer Graphics* (2007). (Cited page 35.)