University of Kaiserslautern
Department of Mathematics
Research Group Algebra, Geometry and Computeralgebra

Dissertation

Topological Methods for the Representation and Analysis of Exploration Data in Oil Industry

by

Oleg Artamonov

Supervisor: Prof. Dr. Gerhard Pfister

Kaiserslautern, 2010

1. Reviewer: Prof. Dr. Gerhard Pfister

2. Reviewer: Prof. Dr. Bernd Martin

Day of the defense: *13 August 2010*

To my dear Grandfather.

Abstract

The purpose of Exploration in Oil Industry is to "discover" an oil-containing geological formation from exploration data. In the context of this PhD project this oil-containing geological formation plays the role of a geometrical object, which may have any shape. The exploration data may be viewed as a "cloud of points", that is a finite set of points, related to the geological formation surveyed in the exploration experiment. Extensions of topological methodologies, such as homology, to point clouds are helpful in studying them qualitatively and capable of resolving the underlying structure of a data set. Estimation of topological invariants of the data space is a good basis for asserting the global features of the simplicial model of the data. For instance the basic statistical idea, clustering, are correspond to dimension of the zero homology group of the data. A statistics of Betti numbers can provide us with another connectivity information. In this work represented a method for topological feature analysis of exploration data on the base of so called persistent homology. Loosely, this is the homology of a growing space that captures the lifetimes of topological attributes in a multiset of intervals called a barcode. Constructions from algebraic topology empowers to transform the data, to distillate it into some persistent features, and to understand then how it is organized on a large scale or at least to obtain a low-dimensional information which can point to areas of interest. The algorithm for computing of the persistent Betti numbers via barcode is realized in the computer algebra system "Singular" in the scope of the work.

Contents

Acknowledgements

This work was dome with the financial support of Deutscher Akademischer Austausch Dienst. The financial support of the University of Kaiserslautern is also gratefully acknowledged.

First of all, I would like to thank my supervisor Prof. Dr. Gerhard Pfister for his encouragement, support, and valuable help with a Sngular programming. He is one who gave me a great opportunity to do my Ph.D. research in a supporting and comforting atmosphere.

Next, I would like to say "Thank you so much!" to Dr. Hennie Poulisse, Principal Research Mathematician in Shell. It may be said with absolute certainty that the results of this thesis would not appear without my cooperation with Hennie. He become the thesis advisor during my one-year internship at the Department of Exploratory Research of Shell Research, Rijswijk, The Netherlands.

My special thanks to Prof. Dr. Dirk Siersma from Utrecht University. It was nice of him to help with a solution of unexpectedly arisen problems in The Netherlands.

I want to express my sincere appreciation of the fruitful concomitant discussions to Matthew Heller.

Finally, I want to acknowledge large majority of people from collective of the Research Group Algebra, Geometry und Computer-algebra of the Department of Mathematics of the University of Kaiserslautern, between whom elapsed my life all these years.

Chapter 1

Introduction

1.1 Motivations of the research.

In oil industry, in order to attain a more effective extraction, it is necessary to obtain information about oil fields location and also about shapes of underground capacities of petroleum gathering. The reason is clear: since to drill one oil well is extremely expensive, it is crucial to understand how the underground geological formation roughly look like. This kind of knowledge demands to process huge amounts of experimental exploration data which always contains a lot of noise and also has missing information. The obtained after a row of explosions data is corresponds to times of arrival of post-explosion reflected waves to a network of special sensor detectors.

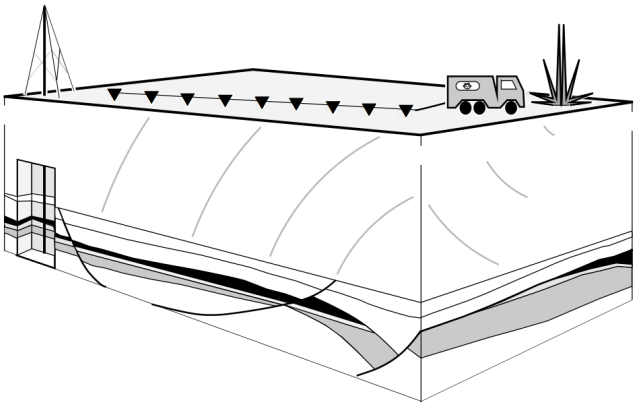


Fig. 1.1: Seismic acquisition on land using a dynamite source and a cable of geophones.

The data coming from real applications is massive and it is not possible to discern and visualize structure even in low resolution. The purpose of this work is to modify and to apply of recently developed techniques of topological data analysis for ad hoc applied objectives. The main message is based on the idea of partial clustering of the data guided by a construction of simplicial complexes in order of topological approximation.

Algebraic topology can be loosely described as the study of spaces through their algebraic images [18]. Since most of the information about topological spaces can be obtained through diagrams of discrete sets, the gist of the method is to reduce high dimensional data sets and to find such a simplicial representation for the reduced data with much fewer points which still encodes some essential topological and geometric information at a specified resolution from the original data.

So the method is based on ideas of algebraic topology. We map our cloud of noisy data to a combinatorial object – some simplicial complex, whose interconnections reflect important aspects of the topological features of the geological object under investigation.

1.2 Input exploration data.

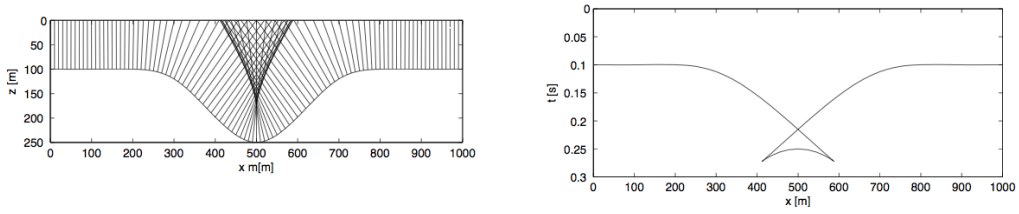
The initial input information about the geological formation is multilevel nature data which may be viewed as a "cloud of points", that is a finite set of points, related to the geological formation surveyed in the exploration experiment. Therefore it can be obtained only with the real geometrical object noisy sampling. By sampling with noise we mean points sampling from a probability distribution concentrated near the underground geological formation surface. As was mentioned, the data is obtained after surface or underground explosions, and corresponds to times of arrival of post-explosion reflected waves to a network of geophones or hydrophones.

In the simplest case, the detectors are located on a straight line after the explosion point with equal distances between each other. In this case the data can be coarsely represented as a plot of points: on the horizontal axis of the plot we have distances of the detectors from the explosion point, and on the vertical one – time of arrival of reflected sound waves to the detectors. Points of the diagram are distributed around series of hyperbolas which, after straightening and rectification of the curve, gives a approximate image of underground rock strata.

The curvature carries messages about constantly unknown density of the inhomogeneous rock medium. Since the underground geological formation can have an extremely complicated shape, the signals arriving after the reflections

Fig. 1.2: 3D marine seismic acquisition, with multiple streamers towed behind a vessel.

from such a surface can arrive to the detectors in a complicated succession. For example, a syncline reflector yields "bow-tie" shape in zero offset section.



depth section with zero offset ray paths *zero offset section*

Fig. 1.3: A syncline reflector (left) yields "bow-tie" shape in zero offset section (right).

Actually, it happens very often on practice when on one detector comes several signals from different directions. A minimal bit of information for us is every such a signal, and therefore we will be treat post-explosion reflected waves arrived on a sensor detector as our initial input points. In the considered case, the signal can be parametrized by $(\Delta_i, t_i, \mathfrak{A}_i)$*, where* Δ_i *is a distance of the* i*-th sensor detector to the explosion point,* t_i *is a interval between the explosion time and the time when the signal came to the* i*-th detector, and* $\mathfrak{A}_i$ *is an amplitude of the signal.*

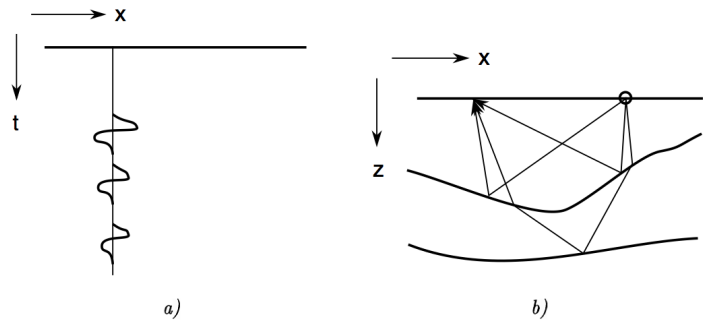


Fig. 1.4: Reflections in time (a) and in depth (b).

In the case of a network of the sensor detectors with the explosion point in the "middle", we just take a polar system of coordinates and increase a dimension of the parametrization. Let us first choose a system of coordinates with the horizontal axis from West to East and with the vertical one from South to North. Also let Δ_i *be a module of the radius-vector from the origin of coordinates to the* i*-th detector, and* θ_i *be a angle between the positive horizontal semi-axis and the radius-vector. Then we have here* $(\Delta_i, \theta_i, t_i, \mathfrak{A}_i)$ *as the parametrization of the* t_i*-time signal at the* i*-th sensor detector.*

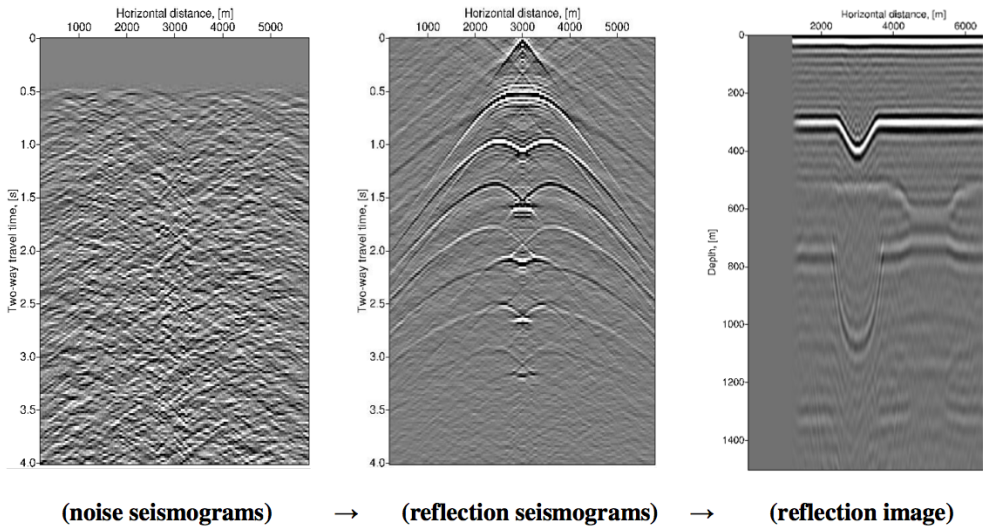


Fig. 1.5: (a) Transmission response of the noise sources in the subsurface observed at the surface. (b) Synthesized reflection response, obtained by seismic interferometry. (c) Synthesized reflection depth image from reflection responses as in (b).

So, in the simplest case, we have on input the experimentally obtained point cloud $X \stackrel{\text{def}}{=} \{x_i \,|\, x_i = (\Delta_i, t_i, \mathfrak{A}_i)\}$, *which can be treated as a finite set of* N *points equipped with the Euclidean distance function* $d(x_i, x_j)$ *between each* $x_i, x_j \in X$. *Of course, any another metric which is a reasonable proxy for an intuitive notion of similarity can also be used.*

It is an entirely separate science about how to derive a proper sampling of underground objects from the exploration experiments. This very challenging task requires a sufficient number of explosions and sensors, as well as proper locations for these. So the problem of extracting of topological and geometrical information about an underground geological formation has two separate aspects: geophysical and mathematical. The first one is beyond our control, and we can only rely on a professionalism of geophysicists here. As this work is devoted to the mathematical side of the problem, we further assume that we obtained a suitable "nice" sampling of the object under investigation, which is uniformly distributed with a sufficient density.

For everybody who is interested in the geophysical aspect of the problem, there is a lot of easily available literature. For example, confer the website listed below:

- *five Jon Claerbout's books:* `http://sepwww.stanford.edu/sep/prof`;
- *Biondo Biondi's publications:*

  `http://sepwww.stanford.edu/data/media/public/sep//biondo/biblio_frame.html`;
- *books from Samizdat Press*: `http://samizdat.mines.edu`;
- *Guy G. Drijkoningen's lecture notes and pictures:*

  `http://geodus1.ta.tudelft.nl/PrivatePages/G.G.Drijkoningen.`

See also [2] and [33].

Chapter 2

Structures for Point Sets

2.1 Homological approximation of real objects.

A principle problem within computational topology is recovering the topology of a finite point set. The assumption is that the acquired point set is sampled from some underlying topological space, whose connectivity is lost during the sampling process. Strictly speaking, we are not able to make any computation from the input point cloud directly. Therefore, we need techniques for computing structures that topologically approximate the underlying space of a given point set. In other words, we should come up ourselves with additional "intermediate" input information by usage of techniques based on a special kind of mathematical formalism. For the sake of clarity, we begin with a topological construction, and proceed then to develop the analogous construction for the experimentally obtained sampling.

In order to be able to do any calculations, we have to encode first our space to a special approximation complex which may be considered as a combinatorial version of the topological space whose properties may now be studied from combinatorial, topological or algebraic aspects. In most general terms, algebraic topology offers two methods for gauging the global properties of a particular topological space, $\mathbb{X}$, by associating with it a collection of algebraic objects. The first set of invariants are the homotopy groups,* $\pi_i(\mathbb{X})$. A much less computationally expensive approach and, therefore, more practical is the second set of invariants,

*Homotopy groups contain information on the number and kind of ways one can map a k-dimensional sphere S^k into $\mathbb{X}$, with two spheres in $\mathbb{X}$ considered equivalent if they are homotopic – belonging to a same path equivalence class – relative to some fixed **basepoint**. The main object here is a so called **fundamental group** – the group of homotopy classes of loops in space. Here a **path** in $\mathbb{X}$ is a continuous map $\varphi\colon [0,\,1] \to \mathbb{X}$, and a **loop** is a path with $\varphi(0)=\varphi(1)$, i.e. φ starts and ends at the same basepoint.

which will be the main object of study in this work. The k-dimensional homology groups, $H_k(\mathbb{X})$, provide information about properties of chains which was formed from simple oriented units known as simplexes. As opposite to homotopy groups, homology groups can be computed using the methods of linear algebra.

We want to recover accessible information about a solid shape of a geometric object in $\mathbb{R}^3$ from the finite point cloud of approximately noisy exploration data empirically sampled from the object. What attributes of the original space could be recovered from this data? Briefly, the idea behind this is what we discuss next. Imagine a volume of oil and gas in some reservoir. This volume can be considered as a manifold Ξ, i.e. as an algebraic surface in 3-space. The surface is embedded in the reservoir rock and also captures faults, as well as impermeable layers in the reservoir rock, in which, as a result, there can be no oil or gas. In this context, these anomalies can be interpreted as holes of the algebraic surface. Information about the number and type of these holes which are contained in a topological space go beyond standard homological approaches. Features of a geometric model which can be obtained by mathematical techniques at our disposal are the three types of holes characterizing its connectivity:

- The gaps that separate components.
- The tunnels that pass through the shape.
- The voids that are components of the complement space inaccessible from the outside.

The decomposition into pass connected components is level zero connectivity information, loops the level one connectivity information, and so forth. Homology groups offers a formal algebraic framework for studying and counting holes in a topological space. Due to an Alexander duality property [18], the ranks of the first three homology groups or Betti numbers, $\beta_0, \beta_1, \beta_2$, count the number of above-mentioned gaps, tunnels, and voids (see [42]).[†] In this manner, homology gives a finite compact description of the connectivity of the object's shape. For instance in the below picture and table are represented simple subspaces of the three-dimensional sphere, $\mathbb{S}^3$, and their Betti numbers, correspondently. E.g. we can see that the torus is one connected component, has two tunnels, and encloses one void, correspondently, $\beta_0 = 1$, $\beta_1 = 2$, and $\beta_2 = 1$. Skeleton of tetrahedron has no voids and forms three tunnels[‡], so $\beta_1 = 3$, and $\beta_2 = 0$ in this case.

[†] In an informal sence, the k-th Betti number $\beta_k(X)$ measures the number of k-dimensional holes in the space X. Tunnels and voids belongs to the complement of a considered complex.

[‡] The edges of the skeleton of tetrahedron form four triangle-shaped cycles but, since one triangle may be represented as the linear some of the other ones, the four together are a vector space of *1*-cycles with rank three.

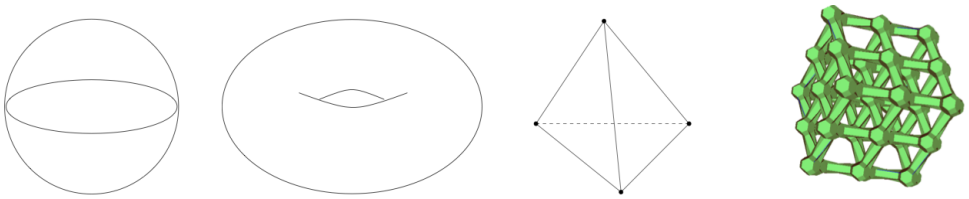


Fig. 2.1: Four simple subspaces of $\mathbb{S}^3$.

	β_0	β_1	β_2
Sphere $\mathbb{S}^2$	1	0	1
Torus $\mathbb{T}^2$	1	2	1
Skeleton of a Tetrahedron	1	3	0
Crystal's Grid	1	3	0

Tab. 2.1: Betti numbers β_0, β_1 and β_2 of the geometrical objects from Figure *2.1*.

Moreover, since homology is an invariant, we may represent the shape combinatorially with a simplicial complex which has the same connectivity, and therefore lead to the same result. It is interesting for us when subcomplexes of a triangulation of $\mathbb{S}^3$ encloses a void, and the void is the empty space enclosed by the complex. One can be interested to find out which holes are long-lasting, that is, persist over a certain parameter range with the course of time, and which is can be easily ignored as topological noise. It is for establishing and counting of these holes, which are represented by so called persistent Betti numbers, the algorithms from computational homology are adopted and implemented in the scope of this PhD research.

Observe that our data which we treat as initial input information are not spatial coordinates, but parametrized post-explosion reflected-back signals which are arrived on sensor devices. Therefore, Betti numbers which can be caught from the data has complex physical meaning and can not be directly interpreted as the above-mentioned multidimensional connectivity information about the underground geological formation under investigation. Nevertheless, since each signal implicates information about the spatial point of the object surface from which it was reflected back, the available Betti numbers still implicates desirable information and thus are valuable for further refinement from noise and persistent features distillation. How to infer Betti numbers of the desired underground shape from the ones of the input noisy cloud X is partially a geophysical question and, as we mentioned above, a geophysical interpretation of the topological information regarding the data is beyond the scope of this paper. So after the assumptions about the sampling, we concentrate our opinion on preliminary ap-

proximation constructions that also are finite combinatorial representations which fit for machine computations.

2.2 Preliminary constructions.

A homology is a topological invariant that is frequently used in practice, since it is computable by linear algebraic methods in all dimensions. The homological method of data investigation characterizes the connectivity of a space X through the structure of its holes (see e.g. [18]) by studying them via equivalence classes of cycles called homology classes. In order to be able to calculate homology by a computer, we need to deal with structures amenable to finite computation. It is not possible to carry out direct computations of homology groups from the definition, but, since the homologies of simplicial complexes are algorithmically computable, it is necessary to use special techniques for spaces which are equipped with a homeomorphism to a some structure which captures the topology of the data. In our case, the capturing of the topology means an approximation of X in terms of homology. So any of our calculations stipulates first for a solving this theoretically challenging problem, i.e. construction of such a simplicial complex.

Huge amount of literature is devoted to the problem of constructing simplicial complexes that represent or approximate a geometric object in some finite-dimensional Euclidean space. As mentioned in [14], this is a special case of the grid generation problem and can be divided into two parts:

1) choose the points or vertices of the grid;

2) connect the vertices using edges, triangles, and higher-dimensional simplices.

Unfortunately, most existing simplicial approximation algorithms are prohibitively expensive since they give too many cells in the approximating complex or are valid only in low dimensional cases. So, in order to estimate topological invariants of X, it is necessary to find a simplicial complex construction which uses relatively few cells and can be efficiently computed in an arbitrary metric space.

One natural operation for the approximation of the original surface is the construction of a triangulated surface using the data points as vertices. Since the topological invariants, which we want to measure, are pretty coarse features of the data, therefore it is much more efficient and absolutely sufficient to construct less detailed approximations, e.g. a triangle is topologically equivalent to a circle.

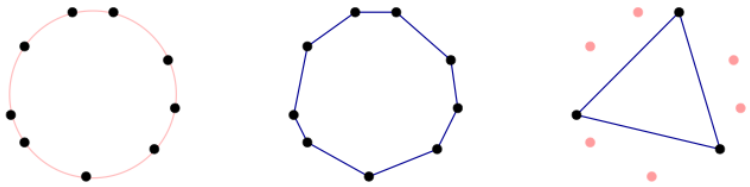


Fig. 2.2: Data sampled from a circle.

In this stream, we can just approximate the intrinsic metric structure of the data by computing shortest paths in a local connectivity graph and define the so called Delaunay complex, a complex which is defined in terms of this intrinsic path-length geometry instead of extrinsic Euclidean geometry. On the other hand, this construction may be interpreted as their dual – graph Voronoi diagram[§] (see [16]), whereby, by definition, Voronoi cells are required to overlap in order to use of their intersection structure. We are going to use restricted versions of these complexes for estimation of Betti numbers from the point-cloud data. Furthermore, we refer for the good survey of different construction algorithms to [30], and, in order to give some context and some comparison, we will start, in due course, with several famous simplicial complex constructions.

2.3 Nerves of coverings, a geometric realization of the point cloud data and the similarity theorem.

All definitions relating to algebraic topology are represented in the Appendix A of the work. As it is infeasible to include an entire course of algebraic topology here (for this see e.g. [7], [18], [20], [28], [29], [32]), we give the most important meanings in order to make an exposition smooth.

An affinely independent point set $T \subseteq X \subseteq \mathbb{R}^d$ defines the k-simplex $\sigma_T = \operatorname{conv} T$ with dimension $k = \dim \sigma_T = \operatorname{card} T - 1$ and which vertexes are points of T. The standard k-simplex can be taken to be the convex hull of the basis vectors in $\mathbb{R}^d$. An abstract simplicial complex, $\mathcal{K}$, is a finite collection of simplices such that satisfies the following two properties:

1) if $\sigma_T \in \mathcal{K}$ and $S \subseteq T$ then $\sigma_S \in \mathcal{K}$;

[§]Let us call a subset $\mathcal{L} \subset X$ as the set of **landmark points**. For some $\ell \in \mathcal{L}$ the **Voronoi cell** $\mathcal{V}_\ell \stackrel{\text{def}}{=} \{x \in X \,|\, d(x,\ell) \leq (x,\ell') \text{ for all } \ell' \in \mathcal{L}\}$ form a covering of X with the **Delaunay complex** attached to $\mathcal{L}$ as the nerve. The **Voronoi diagram** of X is the decomposition of X into Voronoi cells. The **Delaunay triangulation** is the abstract simplicial complex whose vertex set is X, and where a family $\{x_0, x_1, \ldots, x_k\}$ spans a k-simplex if and only if $\mathcal{V}_{x_0} \cap \mathcal{V}_{x_1} \cap \ldots \cap \mathcal{V}_{x_k} \neq \varnothing$ for all $k \geq 0$; it is geometrically realised as a triangulation of the convex hull of X.

2) if $\sigma_T, \sigma_L \in \mathcal{K}$ then $\sigma_T \cap \sigma_L = \sigma_{T\cap V}$.

Each k-simplex has k+1 faces which are $(k$–1)-simplices, each face is obtained by deleting one of the vertices, $\sigma_T \in \mathcal{K}$ if all its faces belongs to the complex. $\mathcal{K}$ has the dimension $\dim \mathcal{K} \stackrel{\text{def}}{=} \max_{\sigma\in\mathcal{K}} \dim \sigma$, the vertex set $\mathsf{vert}\,\mathcal{K} \stackrel{\text{def}}{=} \bigcup_{\sigma_T\in\mathcal{K}} T$, and the underlying space $|\mathcal{K}| \stackrel{\text{def}}{=} \bigcup \mathcal{K} = \bigcup_{\sigma\in\mathcal{K}} \sigma$. A subcomplex of $\mathcal{K}$ is a simplicial complex $\mathcal{L} \subseteq \mathcal{K}$. A triangulation of X is a simplicial complex $\mathcal{K}$ together with a homomorphism between X and $|\mathcal{K}|$; X is triangulable if there exists a simplicial complex $\mathcal{K}$ such that $|\mathcal{K}|$ is homeomorphic to X (see [29], [31]).

Several points in d-dimensional space are in non-degenerate position if there are no $d+2$ cospherical points; in other words, if there are no four points on one circle, five points on one sphere and so forth. The non-degenerate position can be easily simulated computationally by a slight symbolic perturbation. So, since a joggle input is at our service, further we assume that points in X are in the non-degenerate position. Without this assumption we get cells that are not simplices.

We assume we are given a point set X embedded in $\mathbb{R}^d$. As was mentioned, the point set does not have any interesting topology by itself, and so we begin by approximating the underlying space by pieces of the embedding space. An approach than lets decouple geometry from topology is a covering of X – a collection

$$\mathcal{U} \stackrel{\text{def}}{=} \{\, \{U_i\}_{i\in I} \mid U_i \subseteq \mathbb{R}^d,\ X \subseteq \bigcup_i U_i,\ \text{where } I \text{ is an indexing set} \,\}.$$

It is an open or closed covering if each $U_i \in \mathcal{U}$ is open or closed, and it is a finite covering if each U_i is finite. Topological attributes can be localized by an affixment to elements of the covering.

The nerve of a finite covering of $\mathcal{U} = \{U_i\}_{i\in I}$ is the set of cover elements with non-empty common intersections:

$$\mathcal{N} = \mathcal{N}(\mathcal{U}) \stackrel{\text{def}}{=} \mathtt{nerve}\ \mathcal{U} = \{\, \{U_j\}_{j\in J} \mid \cap U_j \neq \emptyset,\ J \subseteq I \}.$$

A geometric realization of $\mathcal{N}(\mathcal{U})$ is a simplicial complex, $\mathcal{G}$, together with a bijection r between $\mathcal{U}$ and $\mathsf{vert}\,\mathcal{G}$, so that $\{U_j\}_{j\in J, J\subseteq I} \in \mathcal{N}$ if and only if the simplex spanned by $r(\{U_j\}_{j\in J})$ is in $\mathcal{G}$.

Since $\mathcal{V} \subseteq \mathcal{U} \in \mathcal{N}$ implies $\mathcal{V} \in \mathcal{N}$, we need to construct an embedding in order to obtain an abstract simplicial complex from the nerve. Let us assume that each element $\mathcal{N}_i \stackrel{\text{def}}{=} \{U_i \mid U_i \in \mathcal{N}(\mathcal{U})\}$ in $\mathcal{N}(\mathcal{U})$ is represented by a point $i \in \mathbb{R}^d$ and that any subset of $\mathcal{N}$ is then represented by the convex hull, conv, of the corresponding points. Now find an injection

$$r : \mathcal{N} \to \mathbb{R}^d \quad \text{so that}$$

$$\mathtt{conv}\, r(\mathcal{U}) \cap \mathtt{conv}\, r(\mathcal{V}) = \mathtt{conv}\, r(\mathcal{U} \cap \mathcal{V}) \quad \text{for all } \mathcal{U}, \mathcal{V} \in \mathcal{N}.$$

The simplicial complex $\mathcal{G} = \mathsf{conv}\, r(\mathcal{N}(\mathcal{U}))$ is the geometric realization of $\mathcal{N}$, and the underlying space of $\mathcal{G}$ is the part of $\mathbb{R}^d$ covered by its simplicies, $|\mathcal{G}| = \bigcup_{\sigma \in \mathcal{G}} \sigma$.

Now we are ready to formulate the crucial point in the complex approximation. This is the famous result of algebraic topology, the so-called nerve theorem of combinatorial topology, also known as the similarity theorem or Leray's theorem (see e.g. [25]).

Theorem 2.1. *Let $\mathcal{U} = \{U_i\}_{i \in I}$ be a finite closed covering of a triangulable space $X \subseteq \mathbb{R}^d$ such that $\{ \bigcap \{U_j\}_{j \in J} \mid J \subseteq I \}$ is either empty or contractible. Let $\mathcal{G}$ be a geometric realization of $\mathcal{N}(\mathcal{U})$. Then X is homotopy equivalent to the underlying space $|\mathcal{G}|$, and therefore has homology isomorphic to that of $|\mathcal{G}|$.*

Loosely, the nerve of a "good" cover is homotopy equivalent to the cover.

This theorem is the basis of most methods for point set representations. In each case we search for a good covering whose nerve will be our representation.

So it is natural now to define the approximation/similarity simplicial complex of X, associated with the covering $\mathcal{U}$, as the underlying space of the geometric realization of the nerve of the covering: $|\mathtt{conv}\, r(\mathcal{N}(\mathcal{U}))|$. Let us give more extensive definition.

An approximation simplicial complex of $X \subseteq \mathbb{R}^d$ associated with the covering $\mathcal{U}$ is the abstract simplicial complex $\mathcal{K}(X) = |\mathcal{G}(\mathcal{N}(\mathcal{U}))|$, whose vertex set is the indexing set $R = r(\mathcal{N}(\mathcal{U}))$, where r is the injection $\mathcal{N}_i \xrightarrow{r} i$ which define the geometric realization, and where a family $\{r_0, r_1, \ldots, r_k\}$ spans a k-simplex if and only if correspondent elements of $\mathsf{conv}\, r(\mathcal{N}(\mathcal{U}))$ have non-empty common intersections:

$$\mathcal{G}(\mathcal{N}_{i_0}) \cap \mathcal{G}(\mathcal{N}_{i_1}) \cap \ldots \cap \mathcal{G}(\mathcal{N}_{i_k}) \neq \emptyset.$$

This work is devoted to the strictly applicable problem of estimating the topological structure of the underground geological formation via homology groups or Betti numbers of the similarity complex. These coarse topological structures are invariants under homotopy equivalence, and therefore it is appropriate and sufficient notion of equivalence for our purposes. The simplicial complex approximation assumes the following aspects.

1. A construction of a simplicial complex, $\mathcal{K} \simeq X$, which depending on X and possibly on additional parameters, but not depending on Ξ. The similarity theorem asserts that such an approximation complex captures topological features of the set.

2. A similarity simplicial complex reflects the homology of Ξ if there exists a homotopy equivalence, $\Xi \simeq \mathcal{K}(X)$, or homeomorphism, $\Xi \approx \mathcal{K}(X)$, between Ξ and $\mathcal{K}$. These relations stipulate for reasonable conditions on X as a sample of Ξ, and for some choice of values for the additional parameters.

The important point here is that when we construct some covering of X, we thereby construct a covering of the "initial" space Ξ, and now, in order to be able to switch from one to another, we need the sampling to be "good enough". In our case, obviously, we could provide such a goodness by a sufficient amount of sensor devises which receives signals, a lucky location of these geophones or hydrophones and also a sufficient amount of explosion located in proper places. Access of the extent of such a sufficiency is however a geophysical problem. Since our research is devoted to the mathematical side of the problem, it is not within the scope of this work.

So from now on, we assume that we have received a sufficiently fine sampling from professional geophysicists and concentrate our attention instead on a construction of the similarity complex. This complex $\mathcal{K}(X)$ ensures that the relations finally imply an approximation of the topological structure of Ξ. It is possible to construct several such simplicial complexes with their own advantages and disadvantages. We must analyze these complexes to compute topological invariants attached to the geometric object around which our data is concentrated.

2.4 Čech and Rips complexes.

Simplicial complex approximations are well understood if they can be interpreted as the nerve of a covering of a space (see [36]). Below we will use non-empty intersections of elements of the covering $\mathcal{U} = \{U_i\}_{i \in I}$ of $X \subseteq \mathbb{R}^d$ in building an another complex from X.

The Čech complex of the covering $\mathcal{U}$ is the abstract simplicial complex, $\check{\mathcal{C}}$, whose vertex set is the indexing set I, and where a family $\{i_0, i_1, \ldots, i_k\}$ spans a k-simplex if and only if $U_{i_0} \cap U_{i_1} \cap \ldots \cap U_{i_k} \neq \emptyset$.

That is, $\check{\mathcal{C}}$ is precisely the nerve $\mathcal{N}(\mathcal{U})$. We treat a corresponding elements of $\{U_i\}_{i \in I}$ as a vertex in our complex whenever $U_{i_\ell} \bigcap U_{i_m} \neq \emptyset$. And then, whenever $\{U_{i_0}, U_{i_1}, \ldots, U_{i_k}\}$ are overlapping, we add a k-simplex $\sigma = [i_0, i_1, \ldots, i_k]$ to the Čech complex.

We need some additional definitions to estate a correspondence between X and $\mathcal{N}(\mathcal{U})$ by kind of partial coordinatization of X with values in $\mathcal{N}(\mathcal{U})$.

A partition of unity subordinate to the finite open covering $\mathcal{U}$ is a family of real valued functions $\{\theta_i\}_{i \in I}$ with the following properties:

1) $0 \le \theta_i(x) \le 1$ for all $i \in I$ and $x \in X$;
2) $\sum_{i \in I} \theta_i(x) = 1$ for all $x \in X$;
3) the closure of the set $\{x \in X \mid \theta_i(x) > 0\}$ is contained in the open set U_i.

The barycentric coordinatization is a bijection between points p_i of a k-simplex σ_T, $T = [p_0, p_1, \ldots, p_k]$, and the set of ordered k-tuples of real numbers $(\mu_0, \mu_1, \ldots, \mu_k)$ so that:

1) $0 \le \mu_j \le 1$, $j \in [0, 1, \ldots, k]$;
2) $\sum_{j=0}^{k} \mu_j = 1$;
3) $\sum_{j=0}^{k} \mu_j p_j = p$.

The numbers $(\mu_0, \mu_1, \ldots, \mu_k)$ are barycentric coordinates of the point p with respect to the simplex σ_T, which are unique and non-negative for all $p \in \sigma_T$.[¶] The barycenter of σ_T is $\mathfrak{b}_T = \sum_{i=0}^{k} \frac{p_i}{k+1}$.

Now for any point $x \in X$ let $\Lambda(x) \stackrel{\text{def}}{=} \{i \in I \mid x \in X_i\} = (i_0, i_1, \ldots, i_l) \subseteq I$. Then we define the desired correspondence map $\rho \colon X \to \mathcal{N}(X)$ as the map $x \mapsto \rho(x)$, where $\rho(x) \in \mathcal{N}(X)$ is the point in the simplex with the vertex set $\Lambda(x)$, whose barycentric coordinates are

$$(\mu_0, \mu_1, \ldots, \mu_l) = \{\theta_{\lambda_{i_0}}, \theta_{\lambda_{i_1}}, \ldots, \theta_{\lambda_{i_l}}\} = \{\, \theta_\lambda(x) \mid \lambda \in \Lambda(x) \,\}.$$

The map $\rho(x)$ is continuous, provides a partial coordinatization of X and is with values in the simplicial complex $\mathcal{N}(\mathcal{U})$.

For this case, the nerve theorem can be formulated as following: if, for all non-empty $J \subseteq I$ we have $\bigcap_{j \in J} \rho^{-1}(U_j)$ is either contractible or empty, then $\check{\mathcal{C}}(X)$ is homotopically (and therefore homologically) equivalent to X.

It is convenient to represent each element U_i of $\mathcal{U}$ as a closed Euclidean ball

$$B_\varepsilon = B(x, \varepsilon) = \{y \in \mathbb{R}^d \mid d(x, y) \le \varepsilon,\ x \in X,\ \varepsilon \in \mathbb{R}\}.$$

Since balls are convex, the nerve theorem implies that $\check{\mathcal{C}}(X, \varepsilon)$ is homotopy equivalent to the union of these balls, and therefore has a straightforward geometrical

[¶] Let $\mathcal{K}$ and $\mathcal{L}$ be two simplicial complexes with a map $\varphi \colon \operatorname{vert}\mathcal{K} \to \operatorname{vert}\mathcal{L}$ which take vertices of any simplexes in $\mathcal{K}$ to the vertexes of a simplex in $\mathcal{L}$. The **simplicial map** implied by φ is $\phi \colon \cup\mathcal{K} \to \cup\mathcal{L}$, which maps a point $p \in \sigma_T$, $T = [p_0, p_1, \ldots, p_k]$, to $\phi(p) = \sum_{i=0}^{k} \mu_i \varphi(p_i)$. If φ is a bijection then ϕ is a homeomorphism.

There is a **standard realization** for an k-simplex as follows. The **standard k-simplex** Δ^k is the convex hull of $\{e^i\}_{i \in \{0,1,\ldots,k\}}$, where $e^i = \{(0, \ldots, 1, \ldots, 0) \mid 1 \text{ in the } i\text{th position},\ i \in I = \{0, 1, \ldots, k\}\}$ is the **i-th standard basis vector** for $\mathbb{R}^k$. For any indexing set $J \subseteq I$, Δ^J is the face of $\Delta^k = \Delta^I$ spanned by $\{e^j\}_{j \in J}$. The standard simplex may be subdivided using the barycenters of its faces to produce the simplicial complex $\mathcal{K}^k$ with $|\mathcal{K}^k| = \Delta^k$. Each non-empty face Δ^J of Δ^k has an associated vertexes in $\mathcal{K}^k$. Δ^J is triangulated by subcomplex $\mathcal{K}^J \subseteq \mathcal{K}^k$ with $|\mathcal{K}^J| = \Delta^J$.

interpretation. Here the radius ε is a typical additional parameter which can serve as a feature scale that directly defines geometrical features of which the scale should be captured by $\check{\mathcal{C}}(X,\varepsilon)$. This is a nested parameter in the sense that $\check{\mathcal{C}}(X,\varepsilon_1)\subseteq\check{\mathcal{C}}(X,\varepsilon_2)$ whenewer $\varepsilon_1\leq\varepsilon_2$. Later, this property will be propagated to inclusions of homology groups of corresponding complexes.

However, we can not be tempted to use $\check{\mathcal{C}}(X,\varepsilon)$ as the approximation complex because it requires storage of inappropriately large amount of simplices of various dimensions even when the underlying topological information is simple. The complex may have dimension much higher than the original space, and such a cumbersome construction is prohibitively expensive computationally. Much less computationally awkward is a construction of a simplicial complex which can be recovered solely from the edge information. So we will relax the *Čhech* condition for simplex inclusion by allowing a simplex of which the vertexes are pairwise within some distance ε.

The Rips complex for the set $X\subseteq\mathbb{R}^d$, attached to the fixed parameter ε, is the abstract simplicial complex $\mathcal{R}_\varepsilon(X)$ whose vertex set is X and whose k-simplexes are spanned by $(k+1)$-tuples $\{x_0,x_1,\ldots,x_k\}$ if and only if $d(x_i,x_j)\leq\varepsilon$ for all $0\leq i,j\leq k$.

$\mathcal{R}_\varepsilon(X)$ is the variant of $\check{\mathcal{C}}$, which is easier to calculate. It is also the largest simplicial complex having the same *1*-skeleton as the correspondent Čech complex $\mathcal{N}_{\frac{\varepsilon}{2}}(X)$. The definition makes sense for an arbitrary metric structure on X, and it avoids the calculations needed to determine whether a set of Euclidean balls has nonempty common intersection. There are obvious inclusions $\check{\mathcal{C}}_\varepsilon(X)\subseteq\mathcal{R}_\varepsilon(X)\subseteq\check{\mathcal{C}}_{2\varepsilon}(X)$.

The disadvantage of the Rips complex is that it is not the nerve of any covering, and it is therefore not amenable to the nerve theorem. This means that $\mathcal{R}_\varepsilon(X)$ is actually not always homotopy equivalent to X. Despite this, the Rips complex is widely uses for approximations in cases where it is homology isomorphic to X, and after some optimal factorisation where this is not the case (at greater length see [9]).

On the below picture are represented Čech and Rips complexes constructed from point cloud data for a particular ε.

Another drawback of the Vietoris-Rips complex is wastefulness from a computational point of view. One way to circumvent this problem is to use the Voronoi diagram to define nested families of some special complexes which makes frugal use of simplices, but is however easily computed.

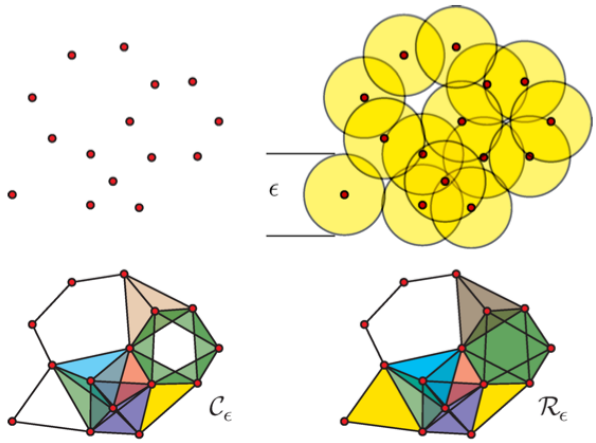


Fig. 2.3: A fixed set of points can be completed to $\check{\mathcal{C}}_\varepsilon(X)$ or to $\mathcal{R}_\varepsilon(X)$. The $\check{C}$ech complex has the homotopy type of the $\varepsilon/2$ cover, $S^1 \vee S^1 \vee S^1$, while the Rips complex has homotopy type $S^1 \vee S^2$.

2.5 Witness complexes.

The strong witness complex $\mathcal{W}(X, \mathcal{L}, \varepsilon)$ for X is the abstract simplicial complex whose vertex set is a finite $\mathcal{L} \subset X$, and where a family $\Lambda = \{\ell_0, \ell_1, \ldots, \ell_k\} \subset X$ spans a k-simplex if and only if there is a point $x \in X$ such that $d(x, \ell_i) \leq \min(x, \mathcal{L}) + \varepsilon$ for all i.

A point $x \in X$ is a ε-weak witness for Λ if $d(x, l) + \varepsilon \geq d(x, \ell_i)$ for all i and all $l \notin \Lambda$.

The weak witness complex $\bar{\mathcal{W}}(X, \mathcal{L}, \varepsilon)$ for X with vertex set is $\mathcal{L}$, is the abstract simplicial complex where Λ spans a k-simplex if and only if $\mathcal{L}$ and all its faces are amenable to ε-weak witnesses.

Here the landmark points $\mathcal{L} \subseteq X$ are chosen to be treated as the vertex set and assumed to be well-distributed over the data.

Withess complexes are based on the idea that the non-landmark data points $X \setminus \mathcal{L}$ can be used to determine the edges and higher-dimensional cells of the complex: the edge $[\ell_i, \ell_i]$ between two landmark points is included in the complex if there exists a data point whose two nearest neighbors in the landmark set are ℓ_i and ℓ_j.

It is very convenient to consider the versions of $\mathcal{W}(X, \mathcal{L}, \varepsilon)$ and $\bar{\mathcal{W}}(X, \mathcal{L}, \varepsilon)$, in which Λ spans a k-simplices if and only if all the pairs (ℓ_i, ℓ_j) are *1*-simplices (see [4]).

There is the important result which implies that, instead of looking for a single strong witness, it is possible to consider the entire aggregate of weak witnesses (at greater length see [8]).

Theorem 2.2. *Lets consider points* $\ell_0, \ell_1, \ldots, \ell_k$ *from a finite* $\mathcal{L} \subset X$*. Then* $\sigma = [\ell_0, \ell_1, \ldots, \ell_k]$ *has a strong witness with respect to* $\mathcal{L}$ *if ond only if* σ *and all its cells have weak witnesses with respect to* $\mathcal{L}$*.*

Below we consider a construction in the framework of witness complexes, which have the defining characteristics:

1) landmarking: the complex $\mathcal{K}$ uses a vertex set $\mathcal{L} \subseteq X$ considerably smaller than the sample X itself;

2) homotopy approximation: under favourable circumstances, there is a homotopy equivalence $\mathcal{K} \simeq X$, and may be a homeomorphism $\mathcal{K} \approx X$, however it need not be a close geometric approximation;

3) intrinsic geometry: the construction estimates and works with the intrinsic geometry of X, what let us to avoid problems relate to the embedding dimension d;

4) non-redundancy: as opposed to the *Čech* complex, $\mathcal{K}$ is not predisposed to an accumulation of redundant high-dimensional cells.

2.6 Voronoi diagrams and Delaunay triangulations.

The Voronoi cell of $p \in X \subseteq \mathbb{R}^d$ is the set of points in the ambient space whose Euclidean distance from p is less than or equal to the distance from any other point in X. That is

$$\mathcal{V}_p \stackrel{\text{def}}{=} \{x \in \mathbb{R}^d \mid d(x,p) \leq d(x,q),\ q \in X\}.$$

Each Voronoi cell is a closed and, in case that p lies on the boundary of $\texttt{conv}\, X$, unbounded convex polyhedron and distinct cells have disjoint interiors. The Voronoi cells meet at most along common boundary faces, the collection of Voronoi cells form the Voronoi diagram

$$\mathcal{V}_X \stackrel{\text{def}}{=} \{\mathcal{V}_p \mid p \in X\},$$

which decomposes $\mathbb{R}^d$ into Voronoi cells and forms a good covering of entire $\mathbb{R}^d$, since all the cells are convex.

The Voronoi cell restricted to Ξ is $\mathcal{V}_{p,\Xi} \stackrel{\text{def}}{=} \{\, \mathcal{V}_p \cap \Xi \mid p \in X \}$, and the collection of restricted Voronoi cells is the restricted Voronoi diagram which is the decomposition of Ξ as a union of cells

$$\mathcal{V}_\Xi = \mathcal{V}_{X,\Xi} \stackrel{\text{def}}{=} \{\, \mathcal{V}_{p,\Xi} \mid p \in X \,\}.$$

If Ξ is a union of ε-balls then the nerve, $\mathcal{N}(\mathcal{V}_{\cup B_\varepsilon})$, of such a restricted Voronoi diagram is the so called alpha complex.

The collection of Voronoi cells restricted to Ξ is a finite closed covering of Ξ. For a subset $T \subseteq X$, the corresponding subsets are

$$\mathcal{V}_T = \{\, \mathcal{V}_p \mid p \in T \,\} \quad \text{and} \quad \mathcal{V}_{T,\Xi} = \{\, \mathcal{V}_{p,\Xi} \mid p \in T \,\} \subseteq \mathcal{V}_\Xi.$$

Since we assume non-degenerate position, the common intersection of any k

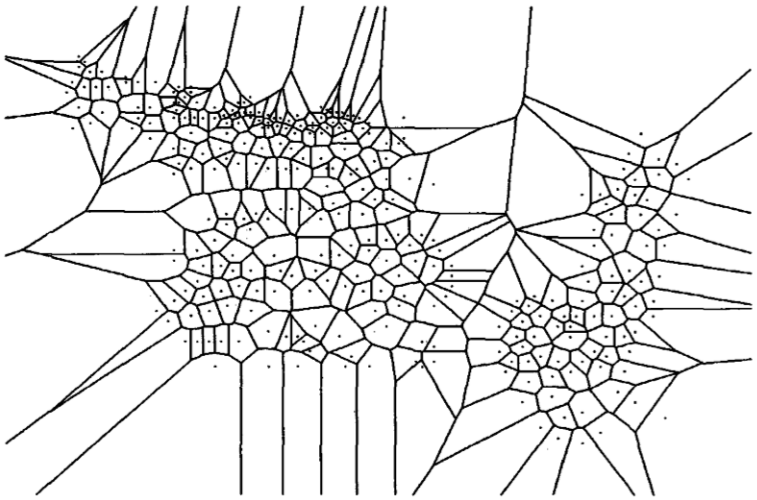


Fig. 2.4: Decomposition of the plane by Voronoi cells of a finite set.

Voronoi cells is either empty or a convex polyhedron of dimension $d+1-k$ and, for any $\mathcal{V} \in \mathcal{N}(\mathcal{V}_X)$, $\texttt{card}\, \mathcal{V} \leq d+1$.

The Delaunay complex of X, $\mathcal{D} = \mathcal{D}_X$, is the geometric realization of $\mathcal{N}(\mathcal{V}_X)$ defined by the injection

$$\{\, r : \mathcal{V}_X \to \mathbb{R}^d \mid r(\mathcal{V}_p) = p \,\} \text{ mapping every } \mathcal{V}_X \text{ to its generator.}$$

That is,

$$\mathcal{D}_X = \{\, \texttt{conv}\, r(\mathcal{V}) \mid \mathcal{V} \in \mathcal{N}(\mathcal{V}_X) \,\}.$$

In other words, if two Voronoi cells share a common $(d-1)$-face then their generating points are connected by an edge, if three cells share a common $(d-2)$-face then their generators are connected by a triangle, etc.

Therefore, we have

$$\mathcal{D}_\Xi = \{\, \sigma_T \mid T \subseteq X,\, \cap \mathcal{V}_{T,\Xi} \neq \emptyset \,\},$$

i.e the convex hull of k points is a cell in the Delaunay complex iff the corresponding k Voronoi cells have a non-empty common intersection not contained in any other Voronoi cell.[‖]

For given X, we have that $\mathcal{D}$ is unique,

$$\dim \mathcal{D} = \min\{\, d,\, \mathtt{card}\, X - 1 \,\} \quad \text{and} \quad \cap \mathcal{D} = \mathtt{conv}\, X.$$

The Delaunay complex decomposes the convex hull of X by connecting the points with simplices of all possible dimensions. In computational geometry, $\mathcal{D}$ is referred to as Delaunay triangulation (see [11],[31]), which is dual to the Voronoi diagram of the points. Advantages of these complexes are that they are small, geometrically realizable, and their highest-dimensional simplexes have the same dimension as the ambient space. In other words, the Delaunay triangulation is a simplicial complex whose vertex set is X and it contains the cell $\sigma = [x_0, x_1, \ldots, x_k]$ whenever $\mathcal{V}_{x_0} \cap \mathcal{V}_{x_1} \cap \ldots \cap \mathcal{V}_{x_k} \neq \emptyset$. Equivalently, $\sigma \in \mathcal{D}$ if there is a point p which is equidistant from $x_0, x_1, \ldots, x_k$ and which has no nearer neighbour in X. Then p is a witness to the cell σ and the assumed non-degenerate points position means that each witness is equidistant from no more then $d+1$ nearest neighbors in X.

By itself, the Delaunay triangulation $\mathcal{D}$ is a contractible simplicial complex, and so its topological invariants carry no information. However, we can define restricted complexes whose structure does reflect the topology of X.

The restricted Delaunay triangulation (or complex) is the geometric realization in $\mathbb{R}^d$ of the nerve of the restricted Voronoi diagram, that is

$$\mathcal{D}_\Xi = \mathcal{D}_{X,\Xi} \stackrel{\text{def}}{=} \{\, \mathtt{conv}\; r(\mathcal{V}) \mid \mathcal{V} \in \mathcal{N}(\mathcal{V}_\Xi),\, r\,(\mathcal{V}_{p,\Xi}) = p,\, p \in X \,\}.$$

So the restricted Delaunay samplicial complex $\mathcal{D}_\Xi$ is the dual of $\mathcal{V}_{X,\Xi}$, and it contains the cell $\sigma = [x_0, x_1, \ldots, x_k]$ whenever $\mathcal{V}_{x_0} \cap \mathcal{V}_{x_1} \cap \ldots \cap \mathcal{V}_{x_k} \cap \Xi \neq \emptyset$. In other words, we demand a witness for σ which lies on Ξ itself.

Let us repeat once again, we can talk about an approximation of the surface $\Xi \subset \mathbb{R}^3$ by the simplicial complex $\mathcal{D}_{X,\Xi}$ only if the sampling $X \subset \mathbb{R}^d$ is a "sufficiently fine". In our applied case, the quality of the approximation depends on sufficiency of density and by evenness of distribution of the sensor devices by which we obtain X.

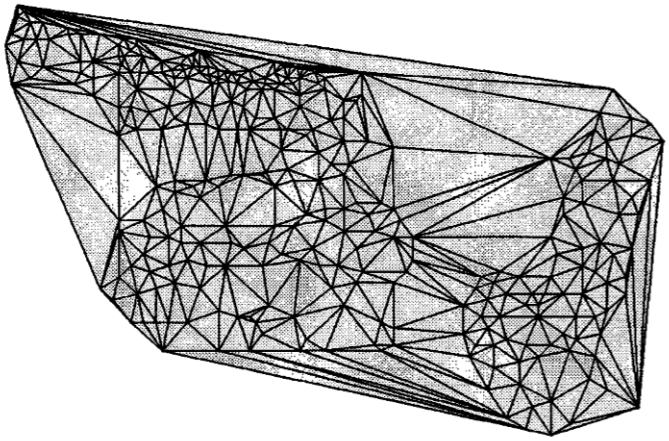


Fig. 2.5: Delaunay complex corresponding to the shown in Figure 2.3 decomposition by Voronoi cells.

After the partitioning of X to subsets that corresponds to elements of the covering, we can use the interaction of the subsets formed in this way between each other for an approximate representation of the exploration data.

Note that $\mathcal{D}_\Xi$ is a subcomplex of the Delaunay simplicial complex, $\mathcal{D}=\mathcal{D}_{\mathbb{R}^d}$, of X and the relationship between Ξ, $\mathcal{V}_\Xi$ and $\mathcal{D}_\Xi$ is elucidated by the nerve theorem. The nerve theorem of Leray [25] implies that, if all restricted Voronoi cells are contractible, then the underlying space of the restricted Delaunay complex, $|\mathcal{D}_\Xi|=\bigcup\mathcal{D}_\Xi$, is homotopy equivalent to Ξ. This means that the two topological spaces can be geometrically different but, in the meanwhile, have the same kind of arrangement of holes, i.e. to be topologically equivalent.

Here is one of crucial points of the work, so let us make a brief resume. As mentioned before, we refer to our underground geometric object under investigation Ξ as a topological space and subspace of $\mathbb{R}^3$. We have only a representation of Ξ by a finite set $X\subseteq\mathbb{R}^d$, where d is a number of parameters in the reflected signals representation. Nevertheless, since each signal implicates information about the spatial point of the object surface from which it was reflected back, the available Betti numbers still implicate desirable information, and therefore are valuable for further refinement from noise and persistent features distillation. In order to represent the geological formation and to be able to do any calculations, we need to construct on the base of X a simplicial complex which captures topology of Ξ. The Voronoi cells of X decompose Ξ into closed convex regions, and the Delaunay simplicial complex restricted by Ξ is defined as the geometric realization of the nerve of these regions by the map $\mathcal{V}\xrightarrow{r}p$, for all $\mathcal{V}\in\mathcal{N}(\mathcal{V}_\Xi)$. The complex $\mathcal{D}_\Xi$ is dual to the restricted Voronoi diagram, its simplices are spanned by

[‖] This construction even does not stipulates non-degenerate position for points of X.

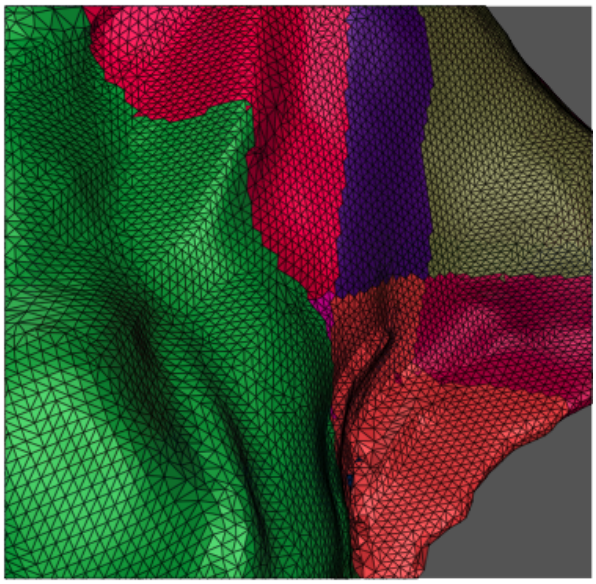


Fig. 2.6: Delaunay based triangulation of a complicated shape.

subsets $T \subseteq X$, and, if the common intersection of any subset of these restricted Voronoi regions, $\bigcap \mathcal{V}_{T,\Xi}$, is empty or is convex and, therefore, contractible for every $T \subseteq X$, then, by the nerve theorem, Ξ and $|\mathcal{D}_\Xi|$ are homotopy equivalent and so have the same topological type.**

So topological properties of a simplicial complex representing the underground space, such as whether its domain is homotopy equivalent or homomorphic to Ξ, is under consideration are based on local interactions between Ξ and the Voronoi neighborhoods of the sampled points X. This leads us to an idea that additional points can be chosen so that improve the local interaction patterns, can be done by using the landmark points, and are related to our special case of the grid generation problem.

2.6.1 The dual complex.

As a good example here can serve us the so called dual complex of a union of balls in $\mathbb{R}^d$, as is demonstrated in [12]. Let $X \subseteq \mathbb{R}^d$ and define

$$\mathcal{X} \stackrel{\text{def}}{=} \{\, x \in \mathbb{R}^d \mid \min_{p \in X} d(x,p) \leq \rho,\ \rho \in \mathbb{R},\ \rho \geq 0 \,\}.$$

**Moreover, $|\mathcal{D}_\Xi|$ and Ξ are homomorphic if the sets can be further subdivided in a certain way so that they form a so called regular CW complex. A closed ball is called a cell, or a k-cell if its dimension is k. A finite collection of non-empty cells, $\mathcal{R}$, is a regular CW complex if the cells have pairwise disjoint interiors, and the boundary of each cell is the union of other cells in $\mathcal{R}$ (see e.g. [26]).

The Voronoi cells $\mathcal{V}_X$ decompose $\mathcal{X}$ into closed convex regions, and the dual complex is defined as the nerve of these regions, $\mathcal{N}(\mathcal{V}_X)$, geometrically realized by the map $\mathcal{V}_{p,X} \xrightarrow{r} p$ for all $p \in X$. It is the same as the restricted Delaunay simplicial complex $\mathcal{D}_{\mathcal{X}}$. The common intersection of any subset of these regions is convex and therefore contractible, and so the nerve theorem implies that the underlying space of the dual complex, $|\mathcal{D}_{\mathcal{X}}|$, is homotopy equivalent to $\mathcal{X}$.

Note that, in terms of open balls, the Delaunay complex, $\mathcal{D} = \mathcal{D}_X$, can be defined as a simplicial complex defined by $X \in \mathbb{R}^d$ and consists of all simplices $\{\sigma_T \,|\, T \subseteq X\}$, for which there exist an open ball

$$B = B(x, \rho) = \{\, y \in \mathbb{R}^d \mid d(x, y) < \rho,\ x \in \mathbb{R}^d,\ \rho \in \mathbb{R} \,\},$$

with $X \cap \mathtt{cl}\, B = T$ and $X \cap B = \emptyset$.

There is an interesting result from [14], stating that, if Ξ is a k-manifold with boundary, then Ξ and $|\mathcal{D}_{X,\Xi}|$ are homeomorphic if $\mathcal{V}_{X,\Xi}$ satisfy the following closed ball property:

1) the common intersection of Ξ and any $k+1-l$ Voronoi cells is either empty or a closed l-ball;

2) the common intersection of the boundary of Ξ and any $k+1-l$ Voronoi cells is either empty or a closed $(l-1)$-ball.

The closed ball property generalizes to a sufficient condition that implies homeomorphic reconstruction for general triangulable space Ξ.

2.6.2 The α-shape complex.

The α-shape complex for the set $X \subseteq \mathbb{R}^d$, attached to the fixed parameter ε, is the abstract simplicial complex $\aleph_\varepsilon(X)$ whose vertex set is X , and where a family $\{x_0, x_1, \ldots, x_k\}$ spans a k-simplex if and only if the convex sets

$$\alpha(x_i, \varepsilon) \stackrel{\text{def}}{=} \{\, B(x_i, \varepsilon) \cap \mathcal{V}_{x_i} \mid x_i \in X,\ i = 0, 1, \ldots, k \,\}$$

have non-empty common intersections.

This complex is homotopy equivalent to the Čech complex. The α-shapes are also nerves of different coverings of a union of the balls. However, α-complexes include much fewer elements and has the same dimension as the ambient space.

So, by the nerve theorem, there is a homotopy equivalence $\aleph(X, \varepsilon) \simeq B(X, \varepsilon)$ and advantage of the α-complex is that it uses considerably fewer cells. Observe that, by construction, the alpha complex is always a subcomplex of the Delaunay

complex, $\aleph_\varepsilon(X) \subseteq \mathcal{D}_X$, and therefore use of the α-shape complexes can significantly reduce calculations, i.e. we may compute the former by computing the latter.

The paradigm of α-shape complexes is defined for $\varepsilon \in [0, \infty]$. The smallest complex $\aleph_0(X)$ is a discrete collection of points, and the largest complex $\aleph_\infty(X)$ is a complete simplex on the vertex set X. The appearance, survival and disappearance of homology classes, as ε varies through intermediate values, provide detailed topological information that is statistically more robust than the Betti numbers of the complex for any single value of ε. Thus we seek similar nested families based on restricted Delaunay complexes.

Chapter 3

Multiresolution and Persistence

3.1 Levels of resolution.

We want to explore sets of point cloud at a various level of resolution, and so to get a possibility to consider outcomes at different levels for comparison. Whenever intervals at two different resolutions have a non empty intersection, there exist a natural map from one set of intervals to the other. This construction produces a multiresolution or multiscale image of the data set.

One can actually construct a family of simplicial complexes which are viewed as images at varying levels of coarseness, and maps between them moving from a complex at one resolution to one of less or more coarser resolution.

Let us assume that we got two coverings $\mathcal{U}=\{U_i\}_{i\in I}$ and $\bar{\mathcal{U}}=\{\bar{U}_j\}_{j\in J}$ of X. A map of coverings from $\mathcal{U}$ to $\bar{\mathcal{U}}$ is a set map $\psi: I\to J$ such that, for all $i\in I$, we have $U_i\subseteq\bar{U}_{\psi(i)}$. This map allow to discern a multiresolution structure of the clustered point cloud, and implies a so called functorial* clustering algorithm.

Then $\mathcal{U}\subseteq\bar{\mathcal{U}}$ implies an inclusion $\mathcal{K}(U_i)\subseteq\mathcal{K}(\bar{U}_i)$ for all $i\in I$. So, if we apply a functorial clustering scheme to the both coverings, it is clear that each connected component of $\mathcal{K}(\mathcal{U})$ is included in exactly one connected component of $\mathcal{K}(\bar{\mathcal{U}})$. So we got a map between sets of clusters of $\mathcal{U}$ and $\bar{\mathcal{U}}$, and therefore a map from the vertex set of $\mathcal{K}(\mathcal{U})$ to the vertex set of $\mathcal{K}(\bar{\mathcal{U}})$. Finally, we obtained an associated induced simplicial map $\lambda\colon \mathcal{K}(\mathcal{U})\to\mathcal{K}(\bar{\mathcal{U}})$ of complexes given on correspondent clusters.

Similar, if we have a family of different coverings

$$X^i=\{\,\mathcal{U}^i \mid X\subseteq\mathcal{U}^i,\ i=0,1,\ldots,n\,\}\ \text{ with inclusions }\ \mathcal{U}^i\subseteq\mathcal{U}^{i+1}.$$

*A clustering algorithm is **functorial** if any inclusion $X\to Y$ of point clouds maps each single cluster in X in one of the unique clusters in Y.

If we represent X with a simplicial complex, then we may also represent its grows with a filtered complex, i.e. we can obtain the correspondent diagrams of complexes and simplicial maps

$$\mathcal{K}(X^0) \xrightarrow{\lambda_0} \mathcal{K}(X^1) \xrightarrow{\lambda_1} \dots \xrightarrow{\lambda_{n-1}} \mathcal{K}(X^n).$$

With the data encoded into simplicial complexes, we are interested in topological features which persist over a sequence of simplicial complexes of different sizes. This sequence reflects changes when new attributes introduced or removed. So the above structure clearly demonstrates that the exploration of the behavior of intrinsic geometric features of X under such maps allows us to distinguish actual features which holds out at multiple scales from artifacts which appears not so often or even at a single level.

The filtration presupposes inclusions by increasing of a certain parameter ε for all involved complexes, i.e. there is an inclusion of the upper complex into the lower one. If ε is too small, the complex is a discrete set X themselves, and, for ε is too large, $\mathcal{K}_\varepsilon(X)$ is a single simplex. In this context, the golden mean may not exist. Algebraic topology suggest a functional approach based on idea, that is the topology of a given space is framed in the mappings to or from that space.

3.2 Persistence homology.

Suppose that we have two topological spaces X and Y and two maps $g, h\colon X \to Y$ between them. Observe first, that the homology groups $H_k(X)$ are a family of Abelian groups for positive integers k with the following properties (see [18]).

Functoriality: each $H_k(X)$ is a functor, that is, for any continuous g, there is the induced homomorphism

$$H_k(g)\colon H_k(X) \to H_k(Y), \text{ such that } \begin{cases} H_k(gh) = H_k(g)H_k(h) \\ \text{and } \ H_k(1_X) = 1_{H_k(X)} \end{cases}.$$

Homotopy invariance: if g and h are homotopic, then $H_k(g) = H_k(h)$. If g is a homotopy equivalence, then $H_k(g)$ is an isomorphism.

The elements of homology groups are cycles, i.e. chains with vanishing boundary, and two k-cycles are considered homologous if their difference is the boundary of the $(k+1)$-chain. In more general terms, $H_k(X)$ determines the number of k-dimensional subspaces of X, which have no boundary in X and themselves are not a boundary of any $(k+1)$-dimensional subspace. Homology groups are computable and provides an insight into topological spaces with maps between

them. Our interest is in a discerning which topological features are essential and which can be safely ignored. This is pretty similar to the signal processing procedure, when a signal is removed from noise.

For any field $\mathfrak{F}$, there is a version of homology with coefficients in $\mathfrak{F}$, that takes values in the $\mathfrak{F}$-vector spaces. Throughout the work, we always compute over field coefficients. The dimension of the vector space is the kth Betti number $\beta_k(X)$ of the space.

Ideally, the complex $\mathcal{K}(X)$ has the same homotopy tipe, and therefore it has the same homology as X. Unfortunately, this is rather an exception and, despite the above theory, in practice it is rather unusual for the complex to capture the homology of the underlying space. The root of the problem is in the noise which implies the not adequate sampling from X, and also there are other faults in the data recovery. We can not distinguish between features of the original space and the noise spanned by the representation.

In order to make an exposition more visual, let us consider the next graphic example. We have the point cloud sampled from an annulus that has the homotopy class of a circle. We assume that the similarity complex, $\mathcal{K}_\varepsilon = \mathcal{K}(X, \varepsilon)$, constructed on such X is the $\check{C}$ech complex, and, as it shown in the below picture, $\check{C}_\varepsilon$ has two smaller additional holes which creates new generators in homology. Therefore, for this value of ε, a computation of the homology of the complex yield a first Betti number is equal to three instead of the desired $\beta_1 = 1$.

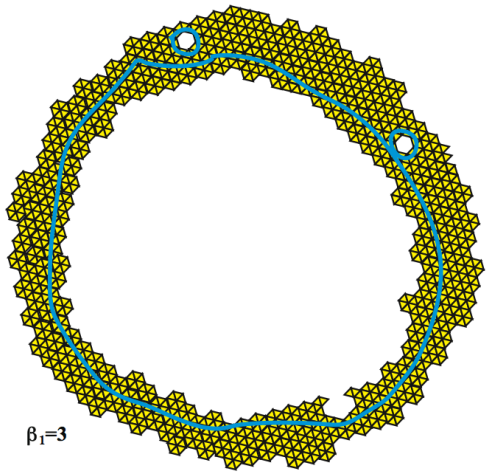


Fig. 3.1: A Čech complex $\check{C}_\varepsilon$ constructed on a finite collection on points in the Euclidean plane.

It seems, that an increasing of the parameter value will give rise to a complex with the correct topological characteristics. Indeed, some voids will enclosed and

filled in, but at the same time new components will appear and connect to the old ones. Therefore, in reality the thickened complex which corresponds to some $\varepsilon' > \varepsilon$ will lead to an another error. As illustrated in the next picture, whereas the two holes have been closed, a new hole has been arisen. Finally, if we compute the homology of this complex, we will get the another incorrect result: $\beta_1 = 2$.

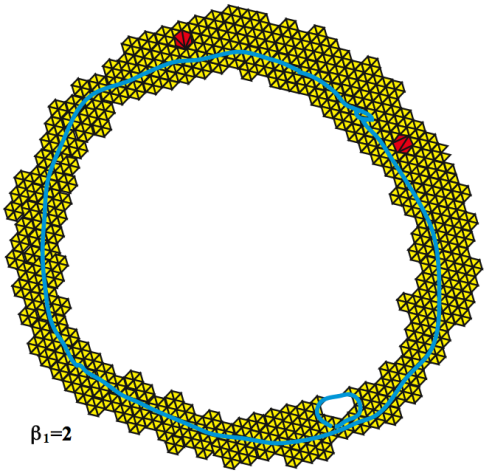


Fig. 3.2: The Čech complex after increasing of the parameter value to $\varepsilon' > \varepsilon$.

Therefore in the similarity complex filtration clusters with the related k-dimensional topological attributes appear and vanish "with the lapse of time", and we need some measure of significance which would enable us to differentiate meaningful information. In other words, we need some segregation technique for the captured attributes which have relatively long lifetime within the filtration grows history.

The filtration $\emptyset = X^0 \subseteq X^1 \subseteq \ldots \subseteq X$ presupposes inclusions by increasing of the filtration parameter for the all involved complexes, i.e. there is an inclusion of the upper complex into the lower one, since the upper complex is corresponds to a bigger filtration index then the lower one. This yields a directed space

$$\emptyset = \mathcal{K}(X^0) \overset{\iota}{\hookrightarrow} \ldots \overset{\iota}{\hookrightarrow} \mathcal{K}(X^j) \overset{\iota}{\hookrightarrow} \ldots \overset{\iota}{\hookrightarrow} \mathcal{K}(X),$$

where the maps ι are the respective inclusions. Applying the kth dimensional homology functor H_k to both the spaces and the maps, we get an another directed space

$$\emptyset = H_k(X^0) \xrightarrow{\iota_k} \ldots \xrightarrow{\iota_k} H_k(X^j) \xrightarrow{\iota_k} \ldots \xrightarrow{\iota_k} H_k(X),$$

where ι_k are the respective induced homology maps. Since in this context the golden mean may not exist, algebraic topology suggest a functional approach

based on the idea, that is the topology of a given space is framed in the mappings to or from that spaces. This leads to extremely powerful tools for studying real data, where a single homology group by itself is likely to be highly unstable with respect to parameter settings and noise.

The trick that will let us to obtain the desired topological information is that, instead of computing homology for the concrete approximation complex, jth element of the filtration of $\mathcal{K}$, $\mathcal{K}^j = \mathcal{K}(X^j)$, we will compute homology for a sufficient amount of the filtration parameters, since the "sufficiency" defines by a wishful level of a resolution coarseness. Now the multiscale approach lets us to consider the so called

persistence complex – a filtered simplicial complex, along with its associated chain and boundary maps, so one considers the ordered sequence of spaces $\{\mathcal{K}^j\}_{j\in J}$, stitched together in a nested family of injections

$$\mathcal{K}^j \overset{\lambda^{j,p}}{\hookrightarrow} \mathcal{K}^{j+p}.$$

By turns the persistence complex leads to the homomorphism

$$H_k(\mathcal{K}^j) \xrightarrow{\lambda_k^{j,p}} H_k(\mathcal{K}^{j+p}),$$

that maps a homology class into the one that contains it. Now we are able to define the crucial concept of this research.

The persistent homology of X is an image of the above homomorphism, $\mathtt{Im}\,\lambda_k^{j,p}$.

This this the quite famous[†] algebraic invariant which derive their popularity from their computability. Persistent homology enable to capture the connectivity of the space, and that peaks out those already existing in $\mathcal{K}^j$ homology classes which persist when we map the complex to the lower one. Here the homological history is modeled by the complex filtration, where simplexes are always added but never removed, implying a partial order on the simplexes. Persistent homology is an algebraic invariant that identifies the birth and death of each topological attribute in this evolution. Another names for persistence are space-time analysis and historical analysis with the filtration as the history of topological and geometrical changes.

Algebraically, the p-persistent kth homology group of the jth complex $\mathcal{K}^j$ in a filtration can be defined as the factor of its kth cycle group, Z_k^j, by the kth boundary group, B_k^{j+p}, of $\mathcal{K}^{j+p}$, p complexes later in the filtration:

$$H_k^{j,p} \overset{\text{def}}{=} Z_k^j / (B_k^{j+p} \cap Z_k^j),$$

[†]It was introduced first by Edelsbrinner, Letscher, Zomorodian ([13]), and then studied in detail by Carlsson and Zomorodian ([44]).

which is well-defined since the denominator is the intersection of two subgroups of C_k^{j+p} and thus is a group itself, a subgroup of the numerator.[‡] Here we derive the cycles which are not turned in the boundaries for p steps in time since the moment j. Here we have that, if two cycles are homologous in $\mathcal{K}^j$, then they are also exist and homologous in $\mathcal{K}^{j+p}$, and therefore we have the isomorphism $\mathtt{Im}\,\lambda_k^{j,p} \cong H_k^{j,p}$.

In each dimension the homology of the complex becomes a vector space over a field, and it fully described by its Betti numbers. Each topological attribute in the similarity complex filtration has a lifetime during which it contributes to some Betti number. We mostly interest in those attributes with longer lifetimes, as they persist in being features of the object's shape. We may represent these life-spans as intervals, and therefore persistent homology can describe the connectivity of the object under investigation via a multiset of intervals in each dimension. For an interval

$$\{\,[a_i, b_i) \mid a_i \in \mathbb{Z}^+,\, b_i \in \mathbb{Z}^+ \cup \{\infty\}\,\}.$$

Let $\mathfrak{F}\,[a_i, b_i) = \{\mathfrak{F}_i\}_{i\geq 0}$ be a directed vector space over the field, $\mathfrak{F}$, which is equivalent to $\mathfrak{F}$ within the interval and is empty elsewhere. Under suitable finiteness hypotheses that are satisfied for all our spaces, the homology directed space may be written as a direct sum

$$\bigoplus_{i=0}^{l} \mathfrak{F}_i,$$

where the description is unique up to a reordering of the summands. Therefore we can track topological attributes and measure their lifetimes as the finite multiset of these so called $\mathcal{P}$-intervals.

Consider some non-bounding k-cycle z, which arise at time i as a consequence of the appearance of the simplex σ^+ in the complex $\mathcal{K}$, such that the correspondent homology class $[z] \in H_k^i$. We mark the simplex σ^+ as a creator. At some moment of time after, p steps later in the filtration, the another just arrived simplex σ^- turn in $[z]$ the homologous to z k-cycle, z', into a boundary. The simplex σ^- labeled as a destroyer, just because it is eliminated both z' and the created before element of H_k^i, and thereby decreased the rank of the kth homology group.

The persistence of the k-cycle z and the correspondent homology class $[z]$ in $\mathcal{K}$ is the difference between endpoints of its life-span interval $[i, i+p)$ in the complex. A non-bounded interval $[i, \infty)$ implies an infinite persistence.

By a varying of the parameter p it is possible to regulate what amount of topological noise we wish to get rid of.

[‡]Here the superscripts indicates the filtration index, so the associated with $\mathcal{K}^j$ groups are $C_k^j, Z_k^j, B_k^j, H_k^j$, and the boundary operators are ∂_k^j for all $j, k \geq 0$.

So the persistent (or persistence) homology is a correct and an effective tool to capture topological invariants from the real data. The main idea is to assess the extent to which features are "genuine" and represented by large holes, as opposed to "artifacts" that may be regarded as an inadequate sampling or a noise, that relate to little holes which collapse almost as soon as they are formed. Lifetime intervals serve us with such an extent, since the features which persist over a range of values of the coarseness, and retains in the complex more then certain threshold time, would be viewed as being less likely to be artifacts.

Chapter 4

Persistence Structures

4.1 Persistence Betti numbers of different dimensions and barcode.

Individual Betti numbers by themselves are highly unstable, and it is natural to define the so called

p-persistent kth Betti numbers as dimensions of the correspondent p-persistent kth homology groups:

$$\beta_k^{i,p} \stackrel{\text{def}}{=} \mathsf{rank}\, H_k^{i,p}.$$

The kth Betti numbers have the meaning that somewhere in the complex k-dimensional subcomplex is missing through all stages of the complex growth or reduction. In other words, an object formed by simplices of dimension at most k is absent from the complex, and this k-dimensional holes remains open when we thicken the complex until the certain filtration index. A sufficient increase of the filtration index p is the mode to clean the "topological signal" of the complex from the "topological noise" which related to the features whose lifespan lasts beyond the chosen threshold.

The persistence compute the compatible homology bases across this growth history, i.e. it represent an algebraic invariant which detect the birth and death of each topological feature as the complex evolve in time. Therefore, it is advantageous to encode the persistent homology in the form of a parameterized version of the rank of homology groups, its Betti numbers, by a representing of the persistence of kth homology group's generators in a multiset of intervals called a **persistence barcode** [6]. Time is a parameter for this purposes, since it encompasses both a birth and, perhaps, a death of each single simplex in the complex. During its temporal existence, each topological attribute plays a part in

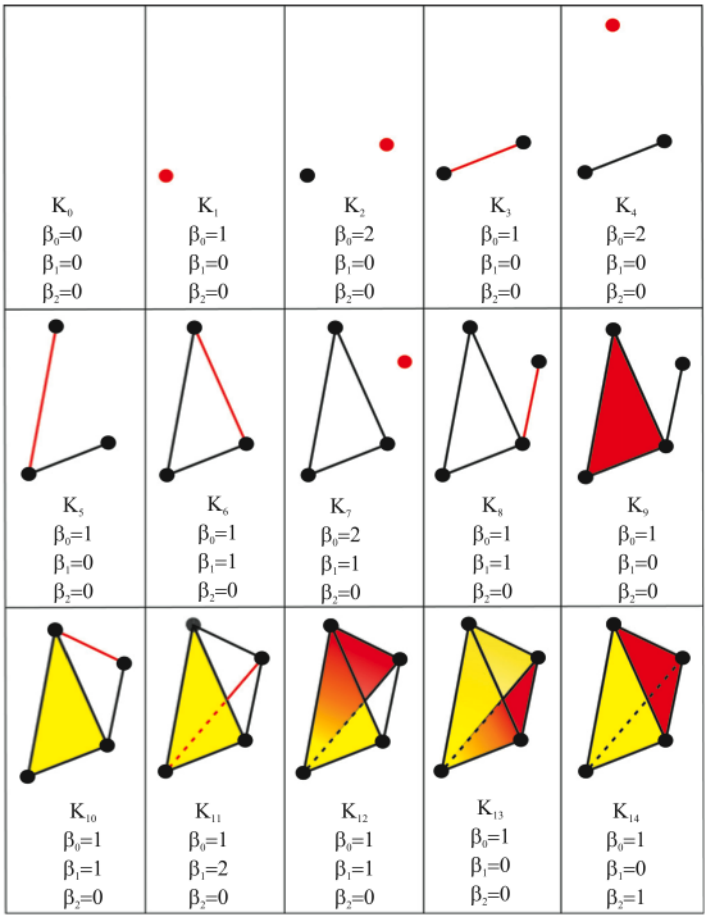


Fig. 4.1: Filtration of a simple simplicial complex with topological characteristics. Just added at a current step vertex or face are represented in red.

the formation of some Betti number, and our interest lies in those properties with long life-spans. The parameter intervals represent lifetimes of various stages of the filtration, and they may be represented on the horizontal axis while arbitrary ordered homology generators H_k may be represented on the vertical axis. E.g. a barcode for the filtration of a tetrahedron of the above picture is presented in the below picture.

The rank of the persistent homology group $H_k^{i,p}$ is equal to the number of $\mathcal{P}$-intervals in the barcode of the homology group H_k, i.e., in detail, is equal to the number of the lifetimes $[i, i+p)$ which corresponds to H_k, and where p is within the limits of the chosen threshold of a "resolution". Barcode reflects the persistent properties of Betti numbers and serve us as a filter that enable a clear distinction between a topological noise and a topological "signal". So the *persistent homology* can be defined as the homology of the growing space that

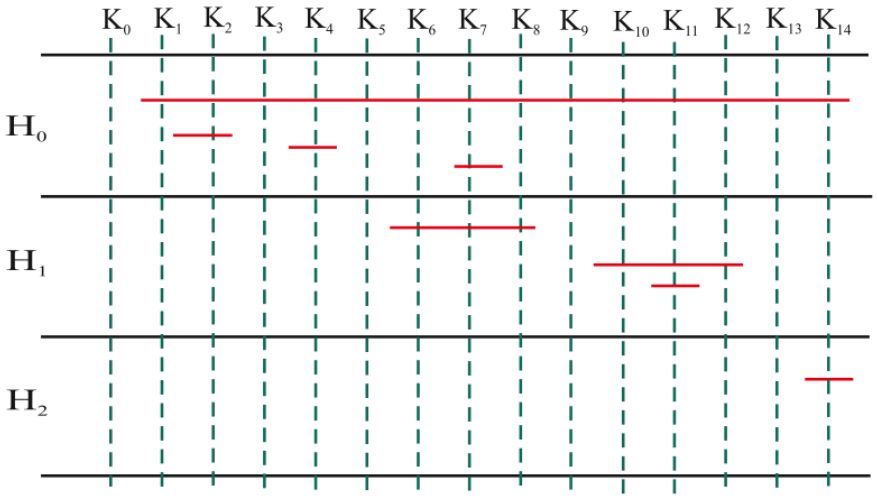


Fig. 4.2: Barcode for the filtration of the simplicial complex presented in Figure 4.1.

captures lifetimes of topological attributes in a barcode.

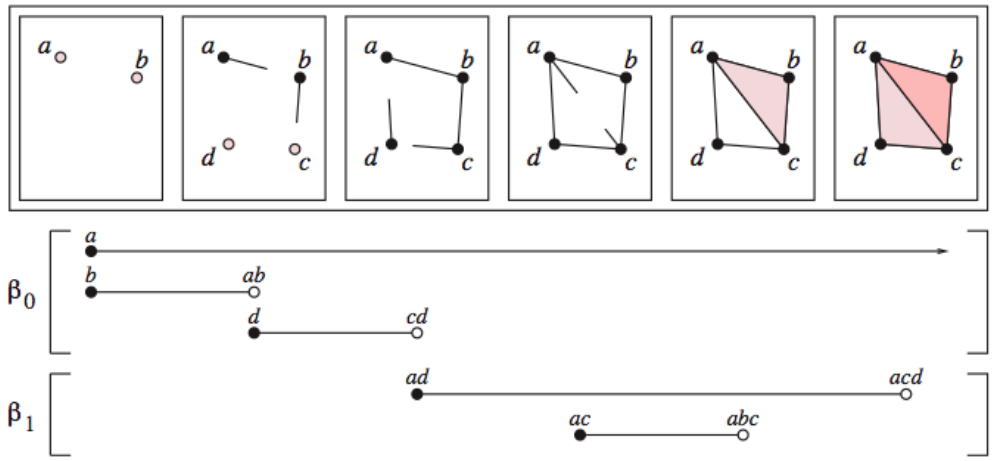


Fig. 4.3: A filtered simplicial complex and its barcode – persistence interval multiset in each dimension. Each persistent interval shown is the lifetime of a topological attribute, created and destroyed by the simplices at the low and high endpoints, respectively.

So here our interest lies in a detection of long-lived homology groups of a simplicial complex during the course of its history which includes both addition and removal of simplices. The method relies on the visual approach of a recognizing of persistent features in the form of a barcode which may be regarded as the persistence analog of Betti numbers. E.g. for the filtration in Figure 4.3, the β_0-barcode is $\{[0, \infty), [0, 1), [1, 2)\}$ with the intervals describing the lifetimes of

the components created by simplices a, b and d, respectively. The β_1-barcode is $\{[2, 5)\,[3, 4)\}$ for the two *1*-cycles created by edges ad and ac, respectively, and provided ad enters the complex after cd at time *2*.

For a simplicial complex, there is the standard algorithm for computing Betti numbers in each dimension [18]. In order to represent this reduction algorithm and according to the *Appendix A* terminology, let us consider the k-dimensional boundary operator

$$\partial_k \colon C_k \to C_{k-1}$$

for the similarity simplicial complex $\mathcal{K}$. Therefore, here we have

$$Z_k \stackrel{\text{def}}{=} \ker \partial_k, \quad B_k \stackrel{\text{def}}{=} \operatorname{Im} \partial_{k+1} \quad \text{and} \quad H_k \stackrel{\text{def}}{=} Z_k / B_k$$

are kth chain (cycle) group, kth boundary group and kth homology group, respectively.

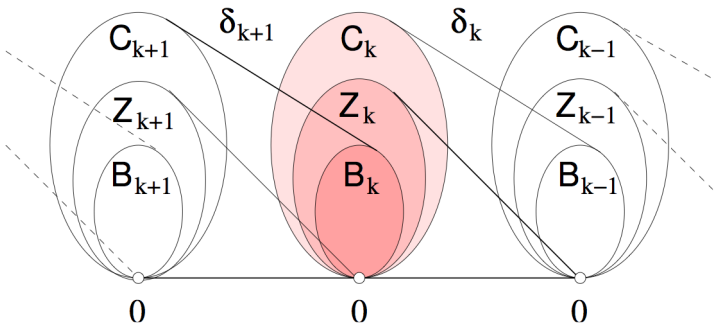


Fig. 4.4: A chain complex with chain, cycle, boundary groups and their images under the boundary operators.

Observe that the chain complex

$$\ldots \xrightarrow{\partial_{k+2}} C_{k+1} \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \longrightarrow \ldots$$

splits into a direct sum of subcomplexes, each is with at most two nonzero terms. The chain complex arises from a finite simplicial complex, and these C_k are finitely generated. Since each C_k splits as $C_k = Z_k \oplus B_k$, we have a diagram:

$$\begin{array}{ccccccccccccc}
\ldots & \xrightarrow{\partial_{k+2}} & C_{k+1} & \xrightarrow{\partial_{k+1}} & C_k & \xrightarrow{\partial_k} & C_{k-1} & \xrightarrow{\partial_{k-1}} & C_{k-2} & \xrightarrow{\partial_{k-2}} & \ldots \\
 & & & & \vdots & & \vdots & & & & \\
 & & 0 & \longrightarrow & B_k & \longrightarrow & Z_{k-1} & \longrightarrow & 0 & & \\
 & & \oplus & & \oplus & & \oplus & & & & \\
 & 0 & \longrightarrow & B_{k+1} & \longrightarrow & Z_k & \longrightarrow & 0 & & & \\
 & \oplus & & \oplus & & \oplus & & & & & \\
0 & \longrightarrow & B_{k+2} & \longrightarrow & Z_{k+1} & \longrightarrow & 0 & & & &
\end{array}$$

As chain groups C_k is free, the oriented k-simplices form the standard basis for it. We represent the boundary operator ∂_k relative to the standard bases of the chain groups, as an integer matrix, M_k, with entries $\{-1, 0, 1\}$. The matrix M_k is called the standard matrix representation of ∂_k, it has m_k columns and m_{k-1} rows which are numbers of k- and $(k-1)$-simplices, respectively. As shown in Figure 4.4, the null-space of M_k is corresponds to Z_k, and its range-space is corresponds to B_{k-1}. In order to reduce M_k to a more manageable form, we use the following elementary row and similary defined elementary column operations:

1) exchange row i and row j;

2) multiply row i by -1;

3) replace row i by (row i)$+ q\cdot$(row j), where q is an integer and $i \neq j$.

Utilizing the reduction, we can derive alternate bases for the chain groups, relative to which the matrix for ∂_k is diagonal. Each column/row operation is corresponds to a change in the basis for $C_k/\ C_{k-1}$, and, finally, we can reduce M_k to its (Smith) normal form, $\widetilde{M}_k$, where all entries are zero except, possibly, a block at the top left corner which may contain nonzero entries down the diagonal:

$$\widetilde{M}_k = \begin{bmatrix} b_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & b_{l_k} & \\ & 0 & & 0 \end{bmatrix}, \text{ where } \begin{cases} l_k = \operatorname{rank} \widetilde{M}_k,\ b_i \geq 1 \text{ and} \\ b_i | b_{i+1} \text{ for all } 1 \leq i < l_k \end{cases}.$$

In this matrix, the columns with non-zero entries corresponds to a basis for the image, and each gives a summand of the form $0 \to \mathbb{Z} \xrightarrow{b_{l_k}} \mathbb{Z} \to 0$. The zero columns of the matrix corresponds to a basis for $\mathbb{Z}_k$, and each one gives a summand of the form $0 \to \mathbb{Z} \to 0$.

Computing the normal form for boundary operators in all dimensions, we get a full characterization of H_k as the following:

a) the torsion coefficients of H_k corresponds to d_i in the considered below structure theorem 4.1, and the diagonal entries are greater then one;

b) since $\{e_i \mid l_k+1 \leq i \leq m_k\}$ is a basis for Z_k, then $\operatorname{rank} Z_k = m_k - l_k$;

c) as $\{b_i \bar{e}_i | 1 \leq i \leq l_k\}$ is a basis for B_{k-1}, therefore $\operatorname{rank} B_k = \operatorname{rank} M_{k+1} = l_{k+1}$.

Combining *b)* and *c)*, we obtain an elegant expression for Betti numbers:

$$\beta_k \ = \ \operatorname{rank} Z_k - \operatorname{rank} B_k \ = \ m_k - l_k - l_{k+1}.$$

For instance, for the filtered simplicial complex in Figure 4.3, the standard matrix representation of ∂_1 is

$$M_1 = \left[\begin{array}{c|ccccc} & ab & bc & cd & ad & ac \\ \hline a & -1 & 0 & 0 & -1 & -1 \\ b & 1 & -1 & 0 & 0 & 0 \\ c & 0 & 1 & -1 & 0 & 1 \\ d & 0 & 0 & 1 & 1 & 0 \end{array}\right],$$

where the bases are shown within the matrix. Reducing the matrix, we receive the normal form

$$\widetilde{M_1} = \left[\begin{array}{c|ccccc} & cd & bc & ab & z_1 & z_2 \\ \hline d-c & 1 & 0 & 0 & 0 & 0 \\ c-b & 0 & 1 & 0 & 0 & 0 \\ b-a & 0 & 0 & 1 & 0 & 0 \\ a & 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

$$\text{where } \begin{cases} z_1 = ad-bc-cd-ab, \\ z_2 = ac-bc-ab \text{ form a basis for } Z_1 \text{ and} \\ \{d-c,\, c-b,\, b-a\} \text{ is a basis for } B_0 \end{cases}.$$

4.2 The persistence module.

It is time now to place the persistence homology within the classical framework of algebraic topology, what will allow us to utilize the standard structure theorem in order to be able to establish the existence of a simple description of persistent homology groups over arbitrary fields as a set of intervals. We are going to use the classification of modules* over a polynomial ring with rational numbers field coefficients for a computation of the persistent homology by a correlation it with the birth and death of topological features in the data. Such a set of intervals for a filtered complex, where positive cycle-creating simplices are paired with negative cycle-destroying simplices, allowed the correct computation of the rank of persistent homology groups.

Let us first to combine the homology of all the complexes in the filtration in a single algebraic structure. As was defined, a persistence complex is a filtered simplicial complex, along with its associated chain and boundary inclusion maps. I.e. we have a family of chain complexes $\{\mathcal{K}_*^i\}_{i\geq 0}$, assume over a commutative ring with unity R, together with chain maps $f^i : \mathcal{K}_*^i \to \mathcal{K}_*^{i+1}$. In general, it is a family of chain complexes $\{\mathcal{K}_*^i\}_{i\geq 0}$ and inclusion chain maps f_i:

$$\mathcal{K}_*^0 \xrightarrow{f^0} \mathcal{K}_*^1 \xrightarrow{f^1} \mathcal{K}_*^2 \xrightarrow{f^2} \dots$$

*All necessary algebraical background it is possible to find e.g. in well written [7].

More widely, along with the boundary maps, we have the diagram where the filtration index increases horizontally under the chain maps, and the dimension decreases vertically under the boundary operators:

$$\begin{array}{ccccccc}
\vdots & & \vdots & & \vdots & & \\
\downarrow{\scriptstyle \partial_3} & & \downarrow{\scriptstyle \partial_3} & & \downarrow{\scriptstyle \partial_3} & & \\
\mathcal{K}_2^0 & \xrightarrow{f^0} & \mathcal{K}_2^1 & \xrightarrow{f^1} & \mathcal{K}_2^2 & \xrightarrow{f^2} & \dots \\
\downarrow{\scriptstyle \partial_2} & & \downarrow{\scriptstyle \partial_2} & & \downarrow{\scriptstyle \partial_2} & & \\
\mathcal{K}_1^0 & \xrightarrow{f^0} & \mathcal{K}_1^1 & \xrightarrow{f^1} & \mathcal{K}_1^2 & \xrightarrow{f^2} & \dots \\
\downarrow{\scriptstyle \partial_1} & & \downarrow{\scriptstyle \partial_1} & & \downarrow{\scriptstyle \partial_1} & & \\
\mathcal{K}_0^0 & \xrightarrow{f^0} & \mathcal{K}_0^1 & \xrightarrow{f^1} & \mathcal{K}_0^2 & \xrightarrow{f^2} & \dots
\end{array}$$

For instance, in the filtered simplicial complex cinsidered in the below picture, at the step 0, there are two contractible connected components which become two circles at the step 1. At the step 2, the two components join to form just one component. Finally, one of the circles is filled, killing off a *1*-cycle class in homology.

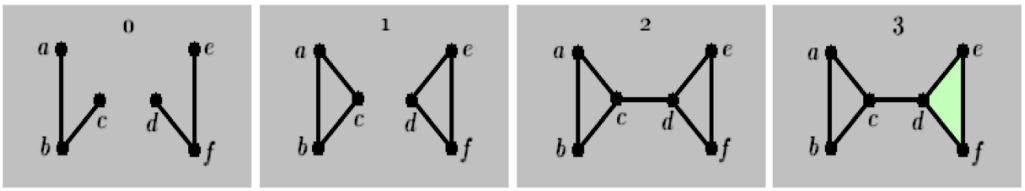


Fig. 4.5: A filtered simplicial complex.

It determines an inductive system of homology groups, i.e. a family of Abelian groups

$$\{H_*^i\}_{i\geq 0} \text{ together with homomorphisms } H_*^i \to H_*^{i+1}.$$

Since the homology is computed with field coefficients, we obtain an inductive system of vector spaces over the field. Each vector space is determined up to the isomorphism by its dimension. In order to obtain a simple classification of an inductive system of vector spaces in terms of a set of intervals, we need an additional structure that represents the homology of a persistent complex.

A persistence module, $\mathcal{M}$, over ring R is a family of R-modules $\mathcal{M}^i$, together with homomorphisms $\varphi^i : \mathcal{M}^i \to \mathcal{M}^{i+1}$. Written as $\mathcal{M} = \{\mathcal{M}^i, \varphi^i\}_{i\geq 0}$.

The persistence module is a structure that represents the homology of a filtered complex, where φ^i merely maps a homology class to the one that contains it.

A persistent complex $\{\mathcal{K}^i_*, f^i\}$ is of finite type if each component complex is a finitely generated R-module, and if, for sufficiently large i, the corresponding maps f^i become R-module isomorphisms. Similarly, a persistent module $\{\mathcal{M}^i_*, \varphi^i\}$ is of finite type if each component of the module is a finitely generated R-module, and if the maps φ^i are isomorphisms for i more or equal of some integer value. Since our complex $\mathcal{K}$ is finite, it generates of a persistence complex, $\{\mathcal{K}^i_*, f^i\}$, of finite type, whose homology is a persistence module, $\mathcal{M}$, of finite type.

The setup is now as follows. Suppose we are given a persistence module $\mathcal{M} = \{M^i, \varphi^i\}_{i\geq 0}$ over a ring R. We need to have a simple classification for these modules and to develop a way to identify elements with the corresponding elements at other timesteps of the filtration. In other words, for a computation of the persistent homology, we have to choose bases which are compatible across the filtration. The main classification comes in two main steps (see [44]).

4.2.1 The Artin-Rees correspondence.

As we assumed, a ring R to be commutative with unity. A polynomial $f(t)$ with coefficients in R is the formal sum

$$\left\{ \sum_{i=0}^{\infty} a_i t^i \mid a_i \in R, \forall t \right\}.$$

The set of all polynomials $f(t)$ over R forms a commutative ring, $R[t]$, with unity. If R has no divisors of zero, and all its ideals are principal, it is a principal ideal domain, PID[†].

A graded ring is a ring $\langle R, +, \cdot \rangle$ equipped with a direct sum decomposition of Abelian groups

$$R \cong \bigoplus_i R_i, \quad i \in \mathbb{Z},$$

so that the multiplication is defined by the bilinear pairings $R_n \otimes R_m \to R_{n+m}$. Elements in a single R_i are called homogeneous, and degree of these elements is i.

Let us grade $R[t]$ non-negatively with the standard grading

$$(t^n) \stackrel{\text{def}}{=} \{t^n \cdot R[t] \mid n \geq 0\}.$$

[†]For our purposes, PID is simply a ring where we may compute the greatest common divisor, gcd, of a pair of elements. This is the key operation needed by the below structure theorem. PIDs include the familiar rings $\mathbb{Z}, \mathbb{Q}$ and $\mathbb{R}$. Finite fields $\mathbb{Z}_p$ for p a prime, as well as polynomials with coefficients from a field F, $F[t]$, are also PIDs and have effective algorithms for computing the gcd [7].

Now we are going to combine all of the complexes in the filtration in order to get a single structure and to encode the time step at which an element is born by a polynomial coefficient. So define

a graded module, $\Gamma(\mathcal{M})$, over a graded ring $R[t]$ as the module which is equipped with a direct sum decomposition, $\mathcal{M} \cong \{\bigoplus_i \mathcal{M}^i \mid i \in \mathbb{Z}\}$, so that the action of R on $\mathcal{M}$ is defined by bilinear pairings $R_n \otimes \mathcal{M}_m \to \mathcal{M}_{n+m}$. A graded ring or module is **non-negatively graded**, if, respectively, $R_i = 0$ or $\mathcal{M}_i = 0$ for all $i < 0$.

Literally, we have

$$\Gamma(\mathcal{M}) \stackrel{\text{def}}{=} \bigoplus_{i=0}^{\infty} \mathcal{M}^i,$$

so the $R[t]$-module structure is simply the sum of the structures on the individual components, where the action of t is given by

$$t \cdot (\, m^0,\, m^1,\, m^2,\, \dots) = (\, 0,\, \varphi^0(m^0),\, \varphi^1(m^1),\, \varphi^2(m^2),\, \dots)$$

that is, t simply shifts elements of the module up in the gradation.

We start by computing a direct sum of the complexes, arriving at a much larger space that is graded according to the filtration ordering. Then, we remember the time when each simplex enters using a polynomial coefficient. For example, while a simplex σ exists at time 0, if it retains in the complex until some time p, we write $t^p \cdot \sigma$ at this time. *The key idea is that the filtration ordering is encoded in the coefficient polynomial ring.*

The main correspondence given by the Artin-Rees theory in commutative algebra (see [15]). In the next section we will demonstrate how Γ defines an equivalence of categories between the category of persistence modules of finite type over R and the category of finitely generated non-negatively graded modules over $R[t]$.

4.2.2 A structure theorem for graded modules over a graded PID.

A structure of a persistence module is described by the below structure theorem.

Theorem 4.1. *If D is PID, then every finitely generated D-module is isomorphic to a direct sum of cyclic D-modules. That is, it decomposes D uniquely into the form*

$$D^{\alpha} \oplus \Big(\bigoplus_{i=1}^{n} D/d_i D \Big), \text{ where } \begin{cases} \alpha \in \mathbb{Z}, \\ d_i \in D, \text{ such that } d_i | d_{i+1} \end{cases}.^{\ddagger}$$

[‡] The theorem decomposes the structure into two parts: the **free** portion on the left and the

Similarly, every graded module, $\mathcal{M}$, over a graded PID, D, decomposes uniquely into the form

$$\Big(\bigoplus_{i=1}^{n} \Sigma^{\beta_i} D\Big) \oplus \Big(\bigoplus_{j=1}^{m} \Sigma^{\gamma_j} D/d_j D\Big),$$

where Σ^{α} denotes a shift upward in grading by α, β_i,$\gamma_i \in \mathbb{Z}$, and $d_j \in D$ are homogeneous elements such that $d_i | d_{i+1}$.

Since the ground ring, R, is assumed to be a field, $\mathfrak{F}$, then the graded ring $\mathfrak{F}[t]$ becomes a PID with only graded ideals are homogeneous in the form (t^n). So, by the above theorem, the structure of graded $\mathfrak{F}[t]$-modules can be represented as the following direct sum:

$$\Big(\bigoplus_{i=1}^{n} \Sigma^{\beta_i} \mathfrak{F}[t]\Big) \oplus \Big(\bigoplus_{j=1}^{m} \Sigma^{\gamma_j} \mathfrak{F}[t]/t^{n_j}\Big).$$

Via the correspondence with finite-type persistence modules given above, the coefficients for this module decomposition can be made meaningful: γ_j and β_i describes when a basis element is created along the filtration. The element then either persists along the filtration until it death at the time γ_j+n_j-1, or it retains in the complex filtration "forever" if it lives in the free left summand. As before, we express the lifespan of a basis element in the filtration by the pairing of its creation and elimination times:

1) $[\beta_i, \infty)$ from the left summand corresponds to a topological attribute that is created at time β_i and exists in the final structure;

2) $[\gamma_j, \gamma_j + n_j)$ from the right summand corresponds to an attribute that is created at time γ_j, lives for time n_j, and is destroyed.

Finally, we can complete the theoretical part of this work by a parametrization of isomorphism classes of finitely generated $\mathfrak{F}[t]$-modules by a finite set of combinatorial invariants. As was mentioned above, a $\mathcal{P}$-interval in an ordered pair

$$\{\, [i,j) \mid 0 \le i < j \in \mathbb{Z}^{\infty} = \mathbb{Z} \cup \{+\infty\} \,\}.$$

For a $\mathcal{P}$-interval $[i,j)$ we define a map

$$\varrho(i,j) \stackrel{\text{def}}{=} \begin{cases} \sum_i \mathfrak{F}[t]/(t^{j-i}) & \text{if } j < \infty \\ \sum_i \mathfrak{F}[t] & \text{if } j = \infty \end{cases},$$

torsional portion on the right. If the ring is a PID, D, the kth homology group H_k is a D-module and the theorem applies that α – the `rank` of the free submodule – is the Betti number of the module, and d_i are its torsion coefficients. When the ground ring is $\mathbb{Z}$, the theorem describes the structure of finitely generated Abelian groups. Over a field, such as $\mathbb{R}$, $\mathbb{Q}$ or $\mathbb{Z}_p$ for p a prime, the torsion submodule disappears. The module is a vector space that is fully described by a single integer, its `rank`, α, which depends on the chosen field.

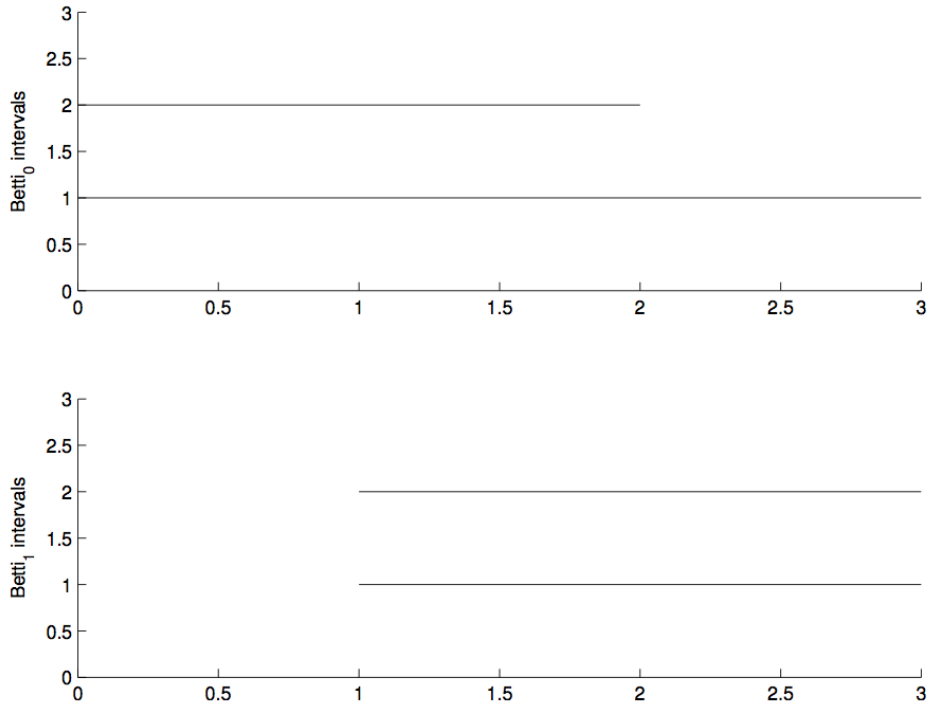


Fig. 4.6: Diagram of $\mathcal{P}$-intervals corresponding to the filtered simplicial complex from Figure 4.5. $[0, \infty)$ and $[0, 2)$ are 0-intervals; $[1, \infty)$ and $[1, 3)$ are 1-intervals.

and consider a finite set of $\mathcal{P}$-intervals

$$\mathcal{S} \stackrel{\text{def}}{=} \{\, [i_1,\, j_1),\, [i_2, j_2),\, \ldots,\, [i_m, j_m)\,\}.$$

We associate $\mathcal{S}$ with the finitely generated graded modules over the graded ring, $\mathfrak{F}[t]$, via the correspondence $\mathcal{S} \to \varrho(\mathcal{S})$ that defines a bijection

$$\varrho(\mathcal{S}) = \bigoplus_{\ell=1}^{m} \varrho(i_\ell, j_\ell).$$

So it was constructed the correspondence with the classification which demonstrates that the isomorphism classes of persistence modules of finite type over a field $\mathfrak{F}$ are bijective to the finite sets of $\mathcal{P}$-intervals. We refer to this multiset of intervals as the considered above barcode. Like a homology group, a barcode is a homotopy invariant.

Since we are working over a field, in each dimension the homology groups $H_k(\mathcal{K}^i)$ are in fact vector spaces, completely described by its ranks, $\beta_k(\mathcal{K}^i)$, which counts the number of topological attributes in the correspondent dimensions. By the structure theorem, there exist a basis for a persistence module that is a compatible basis for all these filtered vector spaces glued by direct sums, which

gives an ability to track attribute's life-spans through the filtration history, i.e. to compute a persistent homology for the whole filtration.

Each $\mathcal{P}$-interval $[i, j)$ describes a basis element for the homology vector spaces from time i until time $j-1$. I.e. this element is a k-cycle, $\mathtt{e}$, that is completed at the moment of time i, forms a new homology class, and also remains non-bounding until the moment of time j, when it joins the boundary group B_k^j. Our point of interest is when the k-cycle $\mathtt{e}+B_k^\ell$ is a basis element for the persistent homology groups $H_k^{\ell,p}$. Here we have three obvious inequalities which define the represented in the below picture triangle region in the index-persistence plain:

$$\begin{cases} p \geq 0 & \text{according to the filtration} \\ \ell \geq i & \text{since } i \text{ is a time of birth of the considered homology class} \\ \ell + p < j & \text{since } \mathtt{e} \in B_K^{\ell+p} \text{ otherwise} \end{cases} .$$

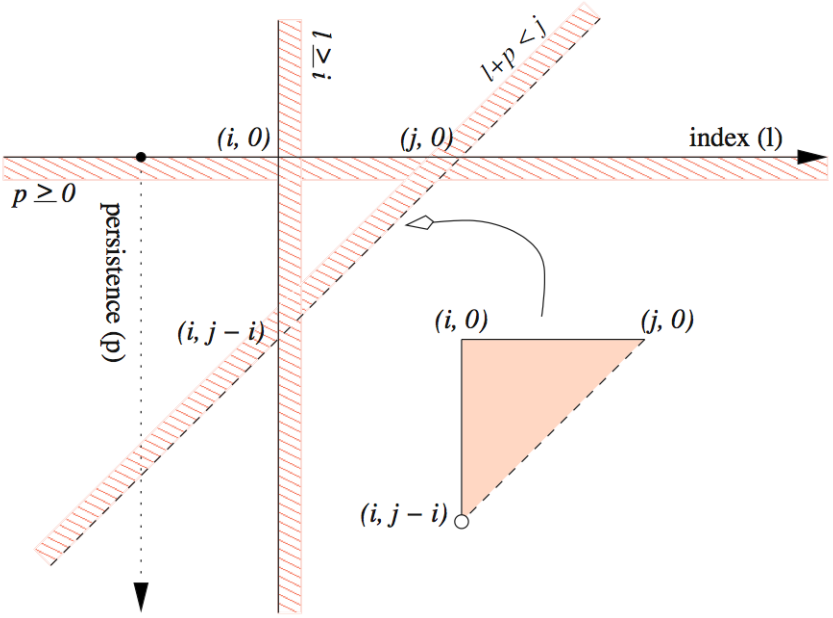


Fig. 4.7: The triangular region in the index-persistence plain, that defines when the cycle is a basis element for the homology vector space.

Lemma 4.2. *Let $\mathcal{T}$ be the set of such triangles defined by $\mathcal{P}$-intervals for the k-dimensional persistent module. Then $\beta_k^{\ell,p}$ is the number of triangles in $\mathcal{T}$ containing the point (ℓ, p) (see [13]).*

The above lemma asserts that a computing of a persistent homology over a field is equivalent to a finding of the corresponding barcode. Observe also, that while component homology groups are torsionless, a persistence appears as torsional and free elements of the persistence module.

It was shown that the persistent homology of a filtered k-dimensional simplicial complex is merely the standard homology of a particular graded module over a polynomial ring.

Chapter 5

The Realized "Singular" Software

5.1 The program structure.

I have implemented the persistence algorithm, the code is represented in the Appendix B. The implementation is in the computer algebra system "Singular" which is perfectly adopted for our purposes, and utilizes a library from "Qhull", a free open source software [39]. The goal was to take a finite cloud of points, X, as input, and compute persistent Betti numbers as a function of a resolution parameter. So the program requires for initial input a point cloud data which is a material for a construction of the approximation complex. The similarity complex triangulate the points cloud sampled from the underlying geological formation, and is equipped with a filtration that explains how the complex might be built in steps. As output the program produce a barcode of a persistent module over a field $\mathfrak{F}$, what is sufficient information for an obtaining of the all desirable persistence information. Observe that we can simulate the algorithm over the considered field itself, without the need for computing the $\mathfrak{F}[t]$-module. The software consists of three independent blocks which is finally melted in the one integral program. These three parts performs the following functions.

1. An approximation of the input point cloud, X, by the restricted Voronoi diagram complex (or by its dual, the restricted Delaunay complex). The similarity complex is represented in the so called Object File Format – the very common data format to represent the geometry of a model by a specifying the polygons of the model's surface. So the result of this block execution is the transformation of the initial input data file to the file `ComplexInput` which has includes a lot of information and statistics but is still not suitable for the represented by the third block main "Singular" program.

That is why we will transform the initial input file once again on the second step. This initial block is *C++* procedure from the free open source software "Qhull" which is compiled and inserted inside of "Singular".

2. A transformation of the OFF file `ComplexInput` to the main procedure. The result of this second transformation of the initial input data is essential information about the approximation complex. Literally, we got the representation of all facets of the complex in the suitable for the main "Singular" procedure form: each of the facets is described by the collection of the vertexes which it is formed by. A convenient way to represent the simplicial similarity complex is via the so called incidence matrix, whose columns are labeled by its vertices and whose rows are labeled by its simplices, as shown in the below picture example. This block is an "Singular" program.

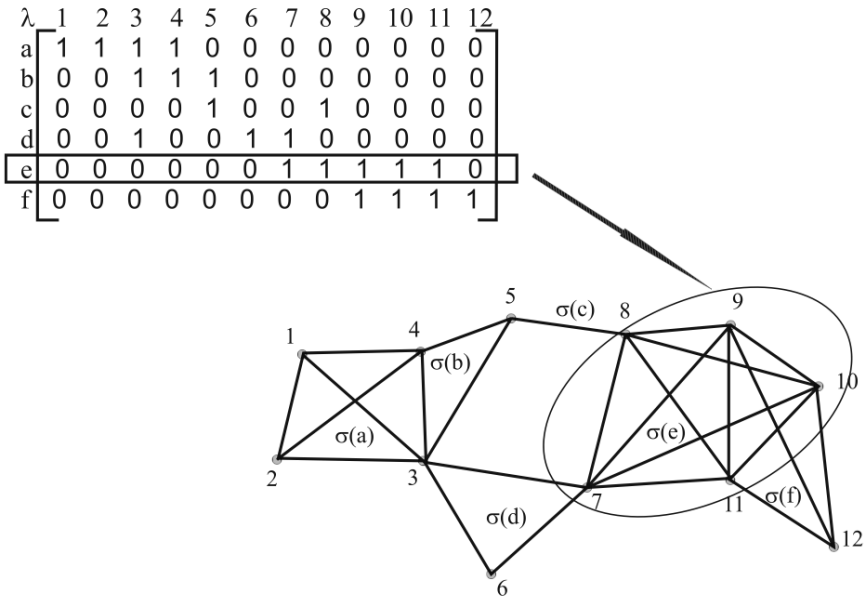


Fig. 5.1: An example of a simplicial complex and its incidence matrix representation. Columns are labeled by its vertices and rows are labeled by its simplices.

3. Computation of the barcodes and persistent Betti numbers from the transformed on the previous steps data, which reflects topological structure of X and, therethrough, of the underground geological object under investigation. The block is the main "Singular" program.

All the three blocks are incapsulated in the program which demand for input a name of an input file with a full directory path.

An initial input is a text file which should include the following information:

1) the first line is the dimension, d, of the data;

2) the second line is the number, N, of points in X;

3) each of the next N lines are coordinates of the points divided just by space characters.

5.2 The persistence algorithm.

The realized in the third block of the program persistence algorithm stipulate for the preprocessing task which is to generate a list of simplices up to the dimension $k+1$ for the k-dimensional homology. A filtration implies a partial order on the cells of the finite simplicial complex $\mathcal{K}$. We start by sorting cells within each time snapshot by dimension with a breaking other ties arbitrarily, obtaining a full order. The algorithm takes for input a full order of the filtered approximation complex's cells. For each simplex $\sigma \in \mathcal{K}$, one needs to identify its faces and to determine its times of appearance and disappearance. The algorithm generate persistence barcodes or barcode – a set of $\mathcal{P}$-intervals that pairs creators and destroyers for each homology class – for the filtered complex, where the positive cycle-creating simplices are paired with the negative cycle-destroying ones. These intervals allow the correct computation of ranks of persistent homology groups.

The algorithm is represented in different works of authors of the persistence idea (see e.g. [13]). The persistence algorithm from the Smith normal form reduction scheme for a computation over arbitrary fields and non-fields for complexes in arbitrary dimensions is represented in [44], a revised version of the algorithm is represented in [43], [45].

5.2.1 Matrix representations.

Let us observe once again that, in each dimension, the homology of complex $\mathcal{K}^i$ becomes a vector space over a field, fully described by its rank β_i. We need to choose compatible bases across the filtration in order to compute persistent homology for the entire filtration. So we form the corresponding to $\mathcal{K}$ persistence module, a direct sum of these vector spaces. The structure theorem states that there is a basis for this module, which provides compatible bases for all the vector spaces. The main purpose of the algorithm is to find a description of such a structure. We will trace a result of the algorithm's work on the example of the simple filtered simplicial complex which was already considered before and is again represented in the below picture.

A homogeneous basis is a basis of homogeneous elements. The first step in the derivation of the algorithm for a computing of persistence homology over a field

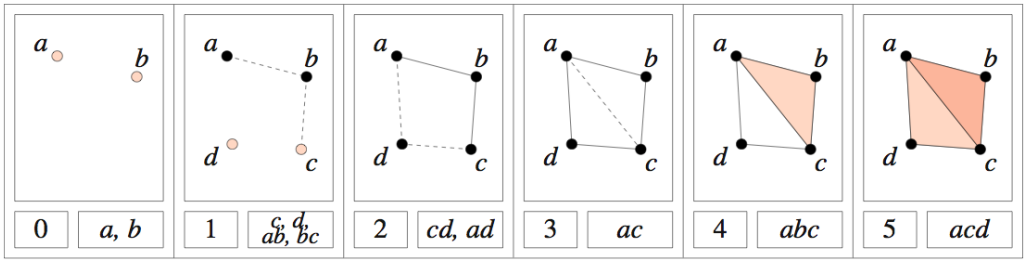


Fig. 5.2: A simple filtration with newly added simplices highlighted and listed.

is to represent the boundary operator $\partial_k : C_k \to C_{k-1}$ relative to the standard basis of C_k and a homogeneous one for Z_{k-1}. Reducing to the normal form, we read off the description provided by the direct sum from the structure Theorem 4.1 using the new basis $\{\hat{e}_j\}$ for Z_{k-1} by the following rules:

1) zero row i contributes a free term with shift $\beta_i = \mathsf{deg}\,\hat{e}_i$;

2) row with diagonal term b_i contributes a torsional term with homogeneous $d_j = b_j$ and shift $\gamma_j = \hat{e}_j$.

We are going "just" to simplify the considered above standard reduction algorithm using the persistence module. As output, we get a finite set of $\mathcal{P}$-intervals for a filtered complex directly over the field $\mathfrak{F}$, which, up to isomorphism, characterize the persistence module – i.e. the homology of the filtered complex – without any necessity to construct one.

Note that, relative to homogeneous bases, a matrix representation M_k of ∂_k has the following property:

$$\mathsf{deg}\,\hat{e}_i + \mathsf{deg}\,M_k(i,j) = \mathsf{deg}\,\{e_j\},$$

where $\{e_j\}$ and $\{\hat{e}_i\}$ are homogeneous bases of C_k and C_{k-1}, respectively, and $M_k(i,j)$ denotes the elements at location (i,j).

Let us return to the simple filtered complex given in Figure 5.2. For ∂_1 with coefficients in $\mathbb{Z}_2$ we get the following matrix expression:

$$M_1 = \left[\begin{array}{c|ccccc} & ab & bc & cd & ad & ac \\ \hline d & 0 & 0 & t & t & 0 \\ c & 0 & 1 & t & 0 & t^2 \\ b & t & t & 0 & 0 & 0 \\ a & t & 0 & 0 & t^2 & t^3 \end{array}\right].$$

The below table reviews the degrees of the simplices of this filtration as homogeneous elements of the persistence $\mathbb{Z}_2$-module.

a	b	c	d	ab	bc	cd	ad	ac	abc	acd
0	0	1	1	1	1	2	2	3	4	5

Table 5.1: Degree of simplices of filtration in Figure 5.2.

For example, it is possible to verify the basic property of the matrix representation of the boundary operator: e.g.

$$\mathsf{deg}\, M_1(5,4) = \mathsf{deg}\, t^3 = \mathsf{deg}\, ac - \mathsf{deg}\, a = 3 - 0 = 3.$$

To arrive at the desired representation of the boundary operator, we proceed inductively in dimension. The base case is simple as $\partial_0 = 0$, $Z_0 = 0$, and the standard basis represents ∂_1. In the inductive step, we assume that there are given: 1) a matrix M_k for ∂_k relative to the standard basis $\{e_j\}$ for chain groups C_k (which is, clearly, homogeneous); 2) a homogeneous basis $\{\hat{e}_i\}$ for Z_{k-1}. For induction, we need to compute a homogeneous basis for Z_k and a matrix representation M_{k+1} of ∂_{k+1} relative to the standard basis of C_{k+1} and the computed basis.

The motivation for this step is the reduction algorithm for homology which was represented in section 4.1, where one computes M_k with respect to a nicer basis. One key distinction though, is that we will be able to accomplish our goal via only column operations. That is, instead of a reducing the matrix completely to its (Smith) normal form using both row and column operations, we can get a halfway there by using Gaussian elimination on the columns, utilizing the elementary column operations of types 1 and 3 only. So we can reduce M_k to its so called column-echelon form, $\widetilde{M}_k$, that is a lower staircase, where the steps have variable height, all landings have a width equal to one, and all non-zero elements must lie beneath the staircase:

$$\widetilde{M}_k = \begin{bmatrix} \star & 0 & 0 & 0 & 0 & 0 & & 0 \\ * & \star & 0 & 0 & 0 & 0 & & \\ * & * & 0 & 0 & 0 & 0 & & \\ * & * & \star & 0 & 0 & 0 & \cdots & \vdots \\ * & * & * & \star & 0 & 0 & & \\ * & * & * & * & 0 & 0 & & \\ * & * & * & * & \star & 0 & & 0 \end{bmatrix}.$$

The $\star$ elements in the example is a pivot, and a row/column with a pivot is called a pivot raw/column. Starting with the leftmost column, we eliminate non-zero entries occurring in pivot rows in order of increasing row. To eliminate an entry, we use an elementary column operation of type 3 that maintains the homogeneity of the basis and matrix elements. We continue until we either arrive

at a zero column, or we find a new pivot. If needed, we then perform a column exchange by an operation of type 1 to reorder the columns appropriately.

The pivot elements in the column-echelon form are exactly the same as the diagonal elements of the (Smith) normal form, which indicated the presence of torsion and free summands in homology. Moreover, the degree of the basis elements on pivot rows is the same in both forms. The only difference here is that now these elements do not correspond to the component homology groups, but they, rather, give torsion and free summands in the persistence module, which are all that we need for the desired pairing.

Observe that the number of pivots in an echelon form is $\mathsf{rank}\, M_k = \mathsf{rank}\, B_{k-1}$, and that the basis elements corresponding to the non-pivot columns of the column-echelon form comprise the desired basis for Z_k. For the complex in Figure 5.2, we continue:

$$\widetilde{M}_1 = \left[\begin{array}{c|ccccc} & cd & bc & ab & z_1 & z_2 \\ \hline d & \mathbf{t} & 0 & 0 & 0 & 0 \\ c & t & \mathbf{1} & 0 & 0 & 0 \\ b & 0 & t & \mathbf{t} & 0 & 0 \\ a & 0 & 0 & t & 0 & 0 \end{array}\right], \text{ where } \begin{cases} z_1 = ad - cd - t\cdot bc - t\cdot ab \text{ and} \\ z_2 = ac - t^2\cdot bc - t^2\cdot ab \text{ form} \\ \text{a homogeneous basis for } Z_1 \end{cases}.$$

So, if we are only interest in the degree of the basis elements, we may read them off from the echelon form directly, and we may use the following corollary of the standard structure Theorem 4.1 to obtain the description.

Corollary 5.2.1.1. *Let $\widetilde{M}_k$ be the column-echelon form for ∂_k relative to the basis $\{e_j\}$ and $\{\hat{e}_i\}$ for C_k and Z_{k-1}, respectively. If row i has pivot $\widetilde{M}_k(i,j) = t^k$, it contributes the summand $\Sigma^{\mathsf{deg}\,\hat{e}_i}\mathfrak{F}[t]/t^k$ to the description of H_{k-1}. Otherwise, it contributes the summand $\Sigma^{\mathsf{deg}\,\hat{e}_i}\mathfrak{F}[t]$. In the language of $\mathcal{P}$-intervals for H_{k-1}, we get the pairs $(\mathsf{deg}\,\hat{e}_i, \mathsf{deg}\,\hat{e}_i + k)$ and $(\mathsf{deg}\,\hat{e}_i, \infty)$, respectively.*

In our example, $\widetilde{M}_1(1,1) = t$. As $\mathsf{deg}\, d = 1$, the element contributes $\Sigma^1\mathbb{Z}_2[t]/(t)$ or $\mathcal{P}$-interval $(1,2)$ to the description of H_0.

From here, it is easy to obtain a matrix representation for ∂_{k+1} with respect to the computed basis for Z_k (see [44]).

Lemma 5.2.1.2. *To represent ∂_{k+1} relative to the standard basis for C_{k+1} and the basis computed for Z_k, by a matrix with respect to the computed basis for ∂_k as above, merely delete the rows of M_{k+1} corresponding to the pivot columns of $\widetilde{M}_k$.*

Therefore, we have no need for row operations and can simply eliminate the rows corresponding to pivot columns one dimension lower. By this way, we are able to get the desired representation for ∂_{k+1} in terms of the basis for Z_k,

what completes the induction. In the considered example, the standard matrix representation for ∂_2 is

$$M_2 = \left[\begin{array}{c|cc} & abc & acd \\ \hline ac & t & t^2 \\ ad & 0 & t^3 \\ cd & 0 & t^3 \\ bc & t^3 & 0 \\ ab & t^3 & 0 \end{array}\right].$$

To get a representation in terms of C_2 and the basis (z_1, z_2) for the computed earlier Z_1, we simply eliminate the bottom three rows. These rows are associated with the pivots in the represented abowe $\widetilde{M}_1$, and we obtain

$$\widetilde{M}_2 = \left[\begin{array}{c|cc} & abc & acd \\ \hline z_2 & t & t^2 \\ z_1 & 0 & t^3 \end{array}\right],$$

$$\text{where} \left\{ \begin{array}{l} \text{we have also replaced } ab \text{ and } ac \text{ with the correspondent} \\ \text{bases elements } z_1 = ad - bc - cd - ab \text{ and } z_2 = ac - bc - ab \end{array} \right. .$$

The persistence algorithm is based on these two lemmas which show that a full reduction to the normal form is unnecessary and that only column operations are needed. The algorithm has the same running time as Gaussian elimination over fields, so it takes $O(m^3)$ in the worst case, where m is the number of simplices in the filtration.

For the considered example filtration in Figure 5.2, the marked 0-simplices $\{a, b, c, d\}$ and 1-simplices $\{ad, ac\}$ generate $\mathcal{P}$-intervals

$$L_0 = \{[0,\infty), [0,1), [1,1), [1,2)\} \text{ and } L_1 = \{[2,5), [3,4)\}, \text{ respectively.}^*$$

5.2.2 A pseudo-code of the revised version of the persistence algorithm.

We need just measure lifetimes of certain topological properties of a filtered simplicial complex, which appears and disappears when simplices are added to the complex. The incremental, one cell at a time, algorithm computes the generator for each homology class and pair the correspondent cell to its partner – the cell which eliminate the class. Once we got this pairing of such creators and destroyers, we can read off the barcode and, herewith, the Betty numbers themselves

*This example is also visually illustrated by the below Table 5.2.

from the filtration. Information about the representatives of homology classes for each cell, σ, stores in a k-chain – so called cascade – which is initially σ itself. Observe that if a destroyer does not exist then the homology class persist until the final simplex in the complex filtration, and the correspondent Betti number is $\beta_i = \infty$.

For instance, the below Table 5.2 reflects all required topological attributes for the filtration in the above Figure 5.2. It is easy to see that e.g. the vertex a has no partner, the vertex b is paired with edge ab, and the vertex d is paired with edge cd. Therefore, as was illustrated in the Figure 4.3, we got intervals $[0, \infty)$, $[0, 1)$ and $[1, 2)$ for β_0 barcode, respectively.†

The algorithm fixes the impact of σ's entry on the topology by the determination whether a boundary, $\partial\sigma$, of the regular cell is already a boundary in the complex $\mathcal{K}$. For this it sweep $\partial(cascade[\sigma])$ through the **while** loop of the represented below algorithm's pseudo-code. After this loop, there are two possibilities:

1. $\partial(cascade\,[\sigma]) = 0$ and we can write $\partial\sigma$ as a sum of the boundary basis elements, so $\partial\sigma$ is already a $(k-1)$-boundary. Therefore $cascade\,[\sigma]$ is a new k-cycle that σ completed. In this case σ is a creator of a new homology cycle and its cascade is a representative of the homology class it created.

2. $\partial(cascade\,[\sigma]) \neq 0$ and $\partial\sigma$ becomes a boundary after we add σ. In this case σ is a destroyer of the homology class of its boundary and its cascade is a chain whose boundary is a representative of the homology class it destroyed. So we pair σ with the youngest – the most recently entered the filtration – cell τ in $\partial(cascade\,[\sigma])$.

σ	a	b	c	d	ab	bc	cd	ad	ac	abc	acd
$partner\,[\sigma]$		ab	bc	cd	b	c	d	acd	abc	ac	ad
$cascade\,[\sigma]$	a	b	c	d	ab	bc	cd	ad	ac	abc	acd
								cd	bc		abc
								bc	ab		
								ab			

Tab. 5.2: Data structure after running the persistence algorithm on the filtration in Figure 5.2. The simplices without partners, or with partners that come after them in the full order, are creators. The others are destroyers.

Repeat briefly: on each step the algorithm identifies the i-th cell as a creator or as a destroyer and then computes its cascade; in the first case, the cascade is

†The vertex c here is died immediately after its birth, what corresponds to the interval $[1, 1)$.

a generator for the homology class it creates; in the second case, the boundary of the cascade is a generator for the boundary class.

Below is the pseudo-code of the persistence algorithm:

```
for σ ← σ_1 to σ_n ∈ K
    {
        partner[σ] ← ∅;
        cascade[σ] ← σ;
    }

    while ∂(cascade[σ]) ≠ 0
        {
            τ ← Yngst(∂(cascade[σ]));

            if partner[τ] ≠ ∅
            {
                cascade[σ] ← cascade[σ] + cascade[partner[τ]];
            }
            else { ↩ while; }
        }

    if ∂(cascade[σ]) ≠ 0
        {
            τ ← Yngst(∂(cascade[σ]));
            partner[σ] ← τ;
            partner[τ] ← σ;
        }
```

The **while** loop corresponds to the processing of one row/column in Gaussian elimination. Here we repeatedly check whether the youngest cell τ in $\partial(cascade[\sigma])$ has partner. If not, we leave the loop. Otherwise, the cycle that τ created was destroyed by its partner, and we add the cascade of the τ's partner to the cascade of σ. Of course, addition of boundaries does not change homology classes.

For example Table 5.3 enable us to trace iterations while we sweep the simplexes cd and then ad throughof the **while** loop.

The algorithm return exhaustive persistence information and works for a large class of cell complexes. In order to formalize this assertion, we need the following definition. Lets assume that we have a filtered cell complex with a partial order on the cells.

A based persistence complex *is a persistence complex equipped with a choice*

$cascade[cd]$	τ	$partner[\tau]$	$cascade[ad]$	τ	$partner[\tau]$
cd	d	$\emptyset$	ad	d	cd
			$ad+cd$	c	bc
			$ad+cd+bc$	b	ab
			$ad+cd+bc$	-	-

Tab. 5.3: Tracing the successive simplexes cd and ad of the considered complex through the iterations of the **while** loop.

of basis in every dimension and persistence level, such that the basis in one fixed dimension and level maps to a subset of the bases in the same dimension and higher levels under inclusion.

Since we only use basis elements and the boundary operator for each given complex, and nothing special to the geometry of the underlying complex, the algorithm computes persistent information for any based persistence complex (see [44]).

5.3 Examples of data processing.

In order to demonstrate that the implementation of the persistence algorithm for fields perform pretty well for data-sets sampled of geometrical objects, I provide here a simple graphic examples of the realized software work.

Example 5.1. *Let us consider a collection of points which is uniformly distributed around boundaries of two distanced from each other areas, e.g. around circles,* $x^2+y^2=1$ *and* $(x-10)^2+y^2=1$. *The input data file is represented in the left column of the below table. The Voronoi diagram of the correspondent point set is shown in Figure* 5.3.

As a result of the program work we obtained the following information. We got ten barcodes which corresponds to zero homology group, and the length of the barcodes or the persistence of zero Betti numbers *are:*

$$97,\ 112,\ 104,\ 88,\ 82,\ 86,\ 93,\ 67,\ 75,\ 46.$$

Now we have to read the data with a filter parameter which fit for the situation. In the considered case, the correct interpretation is possible if we take e.g. 100 as the lowest minimal threshold for the Betti numbers persistence, and receive $\beta_0 = 2$, *two simple connected areas, what is the true result.*

Example 5.2. *Now we add the third circle,* $(x-5)^2+(y+5)^2=1$, *and proceed similarly with the three areas. The augmented input data file and the correspondent Voronoi diagram is represented in the right column of Table* 5.4 *and in Figure* 5.4, *respectively.*

2	%	$2-d$ input sample
56	%	number of points
0	1	
0.26	0.97	
0.51	0.86	
0.7	0.72	
0.81	0.58	
0.87	0.49	
0.97	0.24	
1	0	
0.97	-0.25	
0.86	-0.52	
0.73	-0.68	
0.65	-0.76	
0.49	-0.87	
0.22	-0.97	
0	-1	
-0.24	-0.97	
-0.5	-0.87	
-0.67	-0.74	
-0.78	-0.62	
-0.87	-0.48	
-0.96	-0.28	
-1	0	
-0.98	0.21	
-0.87	0.5	
-0.76	0.64	
-0.67	0.74	
-0.51	0.86	
-0.27	0.96	
10	1	
10.25	0.97	
10.5	0.87	
10.67	0.74	
10.75	0.66	
10.87	0.49	
10.97	0.22	
11	0	
10.97	-0.25	
10.86	-0.5	
10.75	-0.66	
10.63	-0.77	
10.48	-0.88	
10.24	-0.97	
10	-1	
9.75	-0.97	
9.52	-0.88	
9.36	-0.77	
9.29	-0.71	
9.16	-0.54	
9.04	-0.28	
9	0	
9.03	0.24	
9.12	0.47	
9.23	0.64	
9.32	0.73	
9.53	0.88	
9.74	0.97	

2	
84	
0	1
0.26	0.97
0.51	0.86
0.7	0.72
0.81	0.58
0.87	0.49
0.97	0.24
1	0
0.97	-0.25
0.86	-0.52
0.73	-0.68
0.65	-0.76
0.49	-0.87
0.22	-0.97
0	-1
-0.24	-0.97
-0.5	-0.87
-0.67	-0.74
-0.78	-0.62
-0.87	-0.48
-0.96	-0.28
-1	0
-0.98	0.21
-0.87	0.5
-0.76	0.64
-0.67	0.74
-0.51	0.86
-0.27	0.96
10	1
10.25	0.97
10.5	0.87
10.67	0.74
10.75	0.66
10.87	0.49
10.97	0.22
11	0
10.97	-0.25
10.86	-0.5
10.75	-0.66
10.63	-0.77
10.48	-0.88
10.24	-0.97
10	-1
9.75	-0.97
9.52	-0.88
9.36	-0.77
9.29	-0.71
9.16	-0.54
9.04	-0.28
9	0
9.03	0.24
9.12	0.47
9.23	0.64
9.32	0.73
9.53	0.88
9.74	0.97
5	-4
5.23	-4.03
5.49	-4.13
5.69	-4.28
5.78	-4.38
5.88	-4.52
5.97	-4.76
6	-5
5.96	-5.28
5.86	-5.51
5.77	-5.64
5.64	-5.77
5.49	-5.87
5.25	-5.97
5	-6
4.75	-5.97
4.52	-5.88
4.37	-5.77
4.25	-5.66
4.15	-5.53
4.04	-5.26
4	-5
4.02	-4.79
4.12	-4.53
4.23	-4.36
4.34	-4.25
4.49	-4.14
4.72	-4.04

Tab. 5.4: The "Singular" input data file which represents the collection of points sampled of two (left) and three (right) distanced from each other circles.

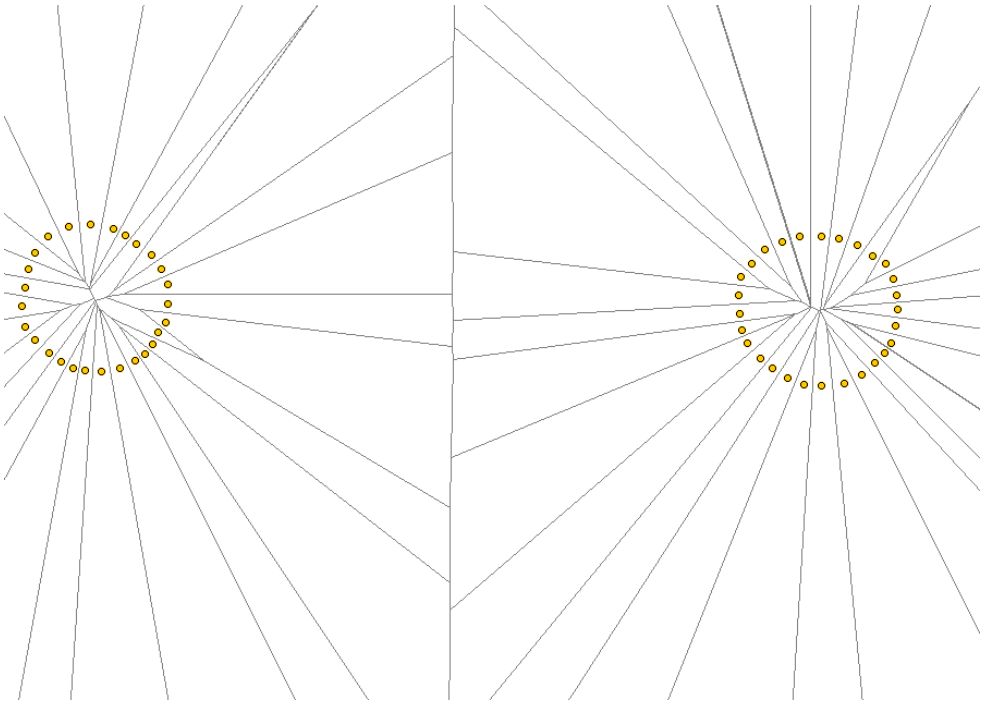


Fig. 5.3: The Voronoi diagram of the points from the left column of Table *5.4*.

In this case, we receive the following persistence related to zero homology group:

$$162,\ 117,\ 112,\ 150,\ 116,\ 145,\ 98,\ 81,\ 81,\ 108.$$

As we can see, the picture is become even more clear then in the previous example, and an isolation of true features from topological "noise" was increased. If we determine the lowest minimal threshold for zero Betti numbers persistence is equal to e.g. 140, we get $\beta_0 = 3$, three simple connected areas, i.e. a correct interpretation of the obtained result.

The establishing of a parameter threshold value is a key moment for data interpretation, and is exactly the "boundary" which separates the mathematical and geophysical parts of the considered project.

It is natural that, in a case when we have huge amount of points, data processing is a time consuming process. So, when we switch to another dimensions or/and to much bigger point clouds, the program demand more then just several seconds like in the considered examples, but, nevertheless, the time remains within reasonable limits.

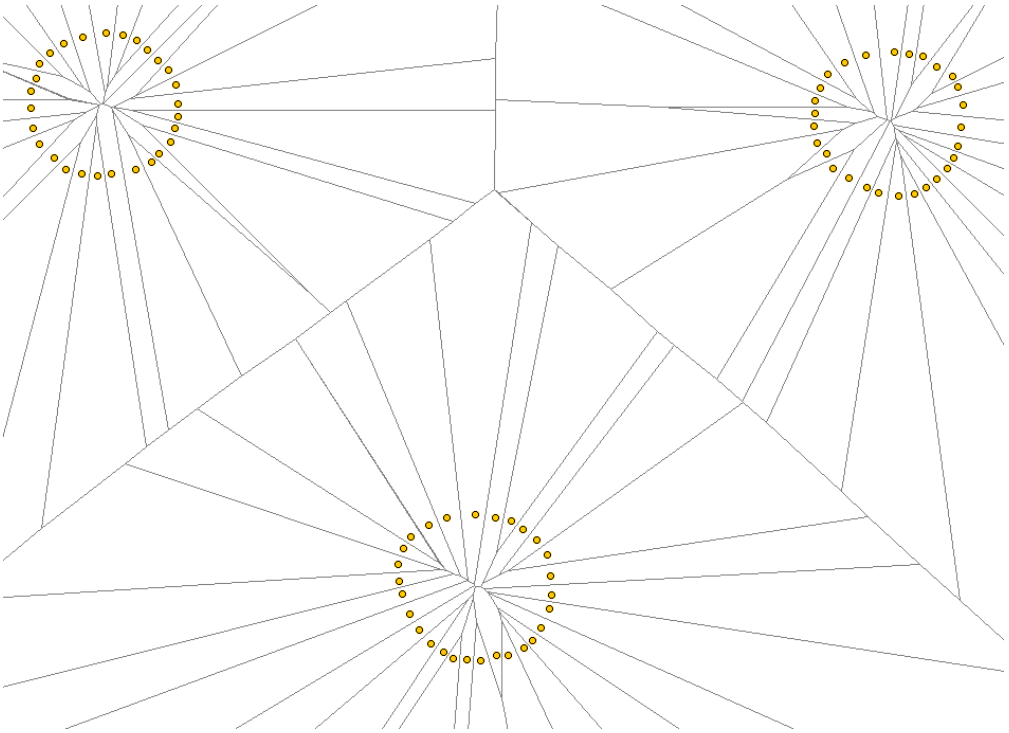


Fig. 5.4: The Voronoi diagram of the points from the right column of Table *5.4*.

5.4 Summary and concluding remarks.

The main aim of this work was to give a topological description of some underground geological formation on a base of exploration data, what is extremely desirable in oil and gas fields prospecting. In the context of this PhD project, this oil-containing geological conformation plays the role of a geometrical object which may have any shape. The obtained during geological exploration experiments data may be viewed as a "cloud of points", and contains both noise and missing information. All the input information at our disposal corresponds to reflected post-explision signals, but, nevertheless, strictly related to the geological formation surveyed in the exploration experiment. Since the sampled data received on Earth surface, in the $3D$ problem the seismic data contains only two of three spatial coordinates of the object under investigation. This is the difference with, for example, medical tomography, where it is possible to scan the geometrical object "from all directions". In order to reconstruct the missing spatial coordinate, we needed to involve techniques based on a special kind of algebraic topology formalism. This implies, in particular, that the experimental data should be considered in terms of methods from computational topology, and useful information can be extracted directly from the experimental, unpro-

cessed exploration data by applying topological methods, notably methods from computational homology.

Construction of an approximation by simplicial complexes creates a topological setting which offer flexible tools for gauging various topological attributes. Here our interest lies in a detection of long-lived homology groups of a constructed similarity simplicial complex during the course of its history which include both addition and removal of simplices. An obvious consequence of such a resilience is that it gives important information about robust quality of the considered topological constructions. The persistence of certain topological attributes assumes also prolonged deficiency in certain topological forms in simplicial complexes, corresponding to the deficiency of certain relations in the geological object, which is indicated by Betti numbers. This dynamical connectivity information cold not be inferred by making use of any conventional methods. As the result, the created "Singular" software has barcodes and persistent Betti numbers of the beforehand created approximation/similarity complex.

This PhD research is devoted to a mathematical part of the ambitious project. I do believe that the combination of this work with the geophysical theory represents an exciting avenue of research and will be a great step towards interpretations of geological exploration data.

Appendices

Appendix A

Basic Notions and Concepts

Following the expositions in the mentioned before classic algebraic topology books, we give here some basic notions and concepts, just in order to give to a reader without this kind of mathematical background a possibility to understand the main regulations of the work.

A topology on a set $\mathbb{X}$ is a system of subsets $\mathcal{X} \subseteq 2^{\mathbb{X}}$ with the following properties:

1) $\emptyset, \mathbb{X} \in \mathcal{X}$;

2) if $\{\mathcal{Y}_i \mid i \in I\} \subseteq \mathcal{X}$, then $\bigcup_{i\in I} \mathcal{Y}_i \in \mathcal{X}$;

3) if $\{\mathcal{Y}_i \mid i \in I,\ I \text{ finite}\} \subseteq \mathcal{X}$, then $\bigcap_{i\in I} \mathcal{Y}_i \in \mathcal{X}$.

The pair $(\mathbb{X}, \mathcal{X})$ is called a topological space. A sets in $\mathcal{X}$ is open sets and the complements of the open sets are closed sets of $\mathbb{X}$. A neighborhood of a point $x \in \mathbb{X}$ is an open set that contains x. A cover is a collection of sets whose union is $\mathbb{X}$. $\mathbb{X}$ is a compact if every cover of $\mathbb{X}$ with open sets has a finite subcover. $\mathbb{X}$ is connected if the only subsets of $\mathbb{X}$ that are both open and closed are $\emptyset$ and $\mathbb{X}$. The subspace topology of $\mathbb{Y} \subseteq \mathbb{X}$ is the system $\mathcal{Y} = \{\mathbb{Y} \cap X \mid X \in \mathcal{X}\}$. The pair $(\mathbb{Y}, \mathcal{Y})$ is called a subspace of the topological space $(\mathbb{X}, \mathcal{X})$.

Suppose that we have topological spaces $\mathbb{X}$ and $\mathbb{Y}$. A function $\varphi\colon \mathbb{X} \to \mathbb{Y}$ is continuous if the preimage of every open set in $\mathbb{Y}$ is open in $\mathbb{X}$. A map is a continuous function. The closure $\bar{A}$ of A is the intersection of all closed sets containing A. The interior $\mathring{A}$ of A is the union of all open sets contained in A. The boundary of A is $\partial A = \bar{A} - \mathring{A}$. A homeomorphism is a bijective map whose inverse is also continuous. $\mathbb{X}$ and $\mathbb{Y}$ are homeomorphic or topologically equivalent or have the same topological type, written $\mathbb{X} \approx \mathbb{Y}$, if there is a homeomorphism between them. This is the most restrictive notion of equivalence in topology.

A homotopy between two maps $f_0, f_1 : \mathbb{X} \to \mathbb{Y}$ is a continuous map

$$F : \mathbb{X} \times [0,1] \to \mathbb{Y} \text{ such that } F(x,0) = f_0(x) \text{ and } F(x,1) = f_1(x) \text{ for all } x \in \mathbb{X},$$

then f_1 and f_2 are said to be homotopic, denoted $f_1 \simeq f_2$, via homotopy F. Given two continuous maps $g : \mathbb{X} \to \mathbb{Y}$ and $h : \mathbb{Y} \to \mathbb{X}$ so that $g \circ h$ and $h \circ g$ are homotopic to to $1_{\mathbb{Y}}$ and $1_{\mathbb{X}}$ respectively, then $\mathbb{X}$ and $\mathbb{Y}$ are homotopy equivalent and have the same homotopy type: $\mathbb{X} \simeq \mathbb{Y}$. A space with the homotopy type of a point is contractible or null-homotopic. Homotopy is a topological invariant since $\mathbb{X} \approx \mathbb{Y} \Rightarrow \mathbb{X} \simeq \mathbb{Y}$.

A covering space of $\mathbb{X}$ is a topological space $\mathbb{Y}$ together with a projection $\mathsf{p} : \mathbb{Y} \to \mathbb{X}$, which satisfies the following property:

- $\forall\, x \in \mathbb{X}$ there is a path-connected neighborhood U so that, for each path-connected component V of $\mathsf{p}^{-1}(U)$, the restriction $\mathsf{p}|_V$ is a homeomorphism.

If $\mathbb{Y}$ is connected then it is universal. Any two universal covering spaces of $\mathbb{X}$ are topologically equivalent.

The d-dimensional Euclidean space is the set of real d-tuples,

$$\mathbb{R}^d = \{\, x = (x_1, x_2, \ldots, x_d) \mid x_i \in \mathbb{R}^d \,\}.$$

The norm of $x \in \mathbb{R}^d$ is $\|x\| = (\sum_{i=1}^{d} x_i^2)^{\frac{1}{2}}$. The distance between points $x, y \in \mathbb{R}^d$ is $d(x, y) = \|x - y\|$. The affine hull of a set of points $T = \{p_0, p_1, \ldots, p_n\}$ in $\mathbb{R}^d$ is

$$\mathsf{aff}(T) \stackrel{\text{def}}{=} \left\{ \sum_{i=0}^{n} \phi_i p_i \;\mid\; \sum_{i=0}^{n} \phi_i = 1 \right\}.$$

T is affinely independent if all $\mathsf{aff}(T)$ is different from the affine hull of every proper subset of T. The convex hull of T is

$$\mathsf{conv}(T) \stackrel{\text{def}}{=} \{\, x \in \mathsf{aff}(T) \mid \phi_i \geq 0 \,\}.$$

Let $T = \{p_0, p_1, \ldots, p_k\}$ be affinely independent. Then $\sigma = \mathsf{conv}(T)$ is a k-simplex with vertixes T and with dimension $\dim(\sigma) = k = \mathsf{card}(T) - 1$, $k \leq d$. A face of σ is a simplex $\tau = \mathsf{conv}(U)$ with $U \subseteq T$, i.e. $\tau \subseteq \sigma$; it is a proper face if U is a proper subset of X. The barycentric coordinates of a point $x \in \sigma$ are the real numbers ϕ_i with

$$\sum_{i=0}^{k} \phi_i p_i = x, \text{ and } \sum_{i=0}^{k} \phi_i = 1.$$

The barycenter of σ is the point $\mathfrak{b}(\sigma)$ with barycentric coordinates $\phi_i = \frac{1}{k+1}$.

A path is a continuous map $\varphi : [0, 1] \to \mathbb{X}$, it joins the initial point, $\varphi(0)$, to the terminal point, $\varphi(1)$. $\mathbb{X}$ is path-connected if every pair of points in $\mathbb{X}$ can be

joined by a path. Two paths are equivalent if they are connected by a homotopy which leaves the common initial and terminal points fixed. The inverse of φ is $\varphi^{-1}(t)=\varphi(1-t)$. The product of two paths φ and ϕ is defined if $\varphi(1)=\phi(0)$:

$$\varphi\cdot\phi=\begin{cases}\varphi(2t) & \text{if } 0\le t\le\frac{1}{2}\\ \phi(2t-1) & \text{if } \frac{1}{2}\le t\le 1\end{cases}.$$

A path φ is a loop if $\varphi(0)=\varphi(1)=x_0$, where x_0 is called a basepoint. A trivial loop $\varphi\cdot\varphi^{-1}$ is equivalent to the constant map $[0,\,1]\to x_0$. The fundamental group of $\mathbb{X}$ at the basepoint x_0, denoted $\pi(\mathbb{X},\,x_0)$, is the equivalence classes of loops based at x_0 together with the product operation. For a path-connected space $\mathbb{X}$ any two groups $\pi(\mathbb{X},\,y_0)$ and $\pi(\mathbb{X},\,z_0)$ are isomorphic, therefore, we have a unique $\pi(\mathbb{X})$ for the entire space. The fundamental group is invariant over homotopy equivalent spaces. If $\mathbb{X}$ is contractible then $\pi(\mathbb{X})$ is trivial; the reverse is not correct, and a good example for this is the d-sphere, where $d\ge 2$.

A topological space may be viewed as an abstraction of a metric space. Similarly, manifolds generalize the connectivity of d-dimensional Euclidean spaces, $\mathbb{R}^d$, by being locally similar, but globally different. A d-dimensional chart at some point $p\in\mathbb{X}$ is a homeomorphism $\varphi\colon U\to\mathbb{R}^d$ onto an open subset of $\mathbb{R}^d$, where U is a neighborhood of p. Loosely, a d-manifold is $\mathbb{X}$ with such a chat φ at every point. So every point of a d-manifold has a neighborhood homeomorphic to $\mathbb{R}^d$. If it exist, the boundary of a d-manifold with boundary is always a $(d-1)$-manifold without boundary, and is a set of $\mathbb{X}$'s points with neighborhoods homeomorphic to $\{\,x=(x_1,\,x_2,\ldots,\,x_d)\in\mathbb{R}^d\mid x_1\ge 0\,\}$. Here we are interest in a compact d-manifold with boundary. A closed surface is a compact *2*-manifold. All manifolds of dimension $d\le 3$ are triangulable.

An ordered k-simplex, $\sigma=[p_0,p_1,\ldots,p_k]$, is a k-simplex together with a permutation of its vertices. Two orderings have the same orientation if they differ by an even permutation. All simplices of dimension *1* and higher have two orientations. The orientation of a $(k-1)$-face induced by σ is

$$\tau=(-1)^i\,[\,p_0,\,p_2,\,\ldots,\,p_{i-1},\,p_{i+1}\ldots,p_k\,],$$

where a leading minus reverses the orientation. An orientation of a k-simplex $\sigma=[\,p_0,\,p_1,\,\ldots,\,p_k\,]$ is an equivalence class of orderings of vertices of σ, where $(\,p_0,\,p_1,\,\ldots,\,p_k\,)\sim(\,p_{\tau_0},\,p_{\tau_1},\,\ldots,\,p_{\tau_k}\,)$ are equvalent if the parity of the permutation τ is even, that is the sign of τ is *1*; denote an oriented simplex as $[\sigma]$. Two k-simplexes sharing a $(k-1)$-face τ are consistently oriented if they induce different orientations of τ. A triangulable d-manifold is orientable if all d-simplices in any of its triangulations can be oriented consistently, i.e. so that all adjacent pairs are consistently oriented. Otherwise, the d-manifold is non-orientable. Non-orientable closed surfaces can not be embedded in $\mathbb{R}^3$.

A simplicial complex, $\mathcal{K}$, is a finite collection of simplices such that:

1) if $\sigma \in \mathcal{K}$ and τ is a face of σ then $\tau \in \mathcal{K}$;

2) if $\sigma, \sigma' \in \mathcal{K}$ then $\sigma \cap \sigma'$ is empty or a face of both.

The multifaceted property – algebraic, topological and combinatorial – of simplicial complexes makes them particularly convenient for a modeling of complex structures and connectedness between different substructures. A subcomplex of $\mathcal{K}$ is a subset of $\mathcal{K}$ which is also a simplicial complex. The k-skeleton $\mathcal{K}^{(\ell)}$ of $\mathcal{K}$ is the subcomplex containing simplices with dimension less than or equal to ℓ. The vertex set is $\mathsf{vert}\,\mathcal{K} = \{\sigma \in \mathcal{K} \mid \dim \sigma = 0\}$. The underlying space of $\mathcal{K}$ is the part of space covered by simplexes in $\mathcal{K}$: $|\mathcal{K}| = \bigcup_{\sigma \in \mathcal{K}} \sigma$. The dimension is $\dim \mathcal{K} = \max\{\dim \sigma \mid \sigma \in \mathcal{K}\}$. A triangulation of a topological space $\mathbb{X}$ is a simplicial complex, $\mathcal{K}$, such that $|\mathcal{K}| \approx \mathbb{X}$. Triangulation enable us to represent topological spaces compactly as simplicial complexes, and $\mathbb{X}$ is triangulable if it has a triangulation. Two simplicial complexes $\mathcal{K}$ and $\mathcal{L}$ are isomorphic, $\mathcal{K} \cong \mathcal{L}$, if $|\mathcal{K}| \approx |\mathcal{L}|$. Usually, the evolution of the complex considers its creation starting from the empty set, hence, the assumption is that simplices are added to the complex in the order of increasing. A filtration of a complex $\mathcal{K}$ is a nested sequence of subcomplexes

$$\emptyset = \mathcal{K}^0 \subseteq \mathcal{K}^1 \subseteq \ldots \subseteq \mathcal{K}^d = \mathcal{K},$$

where superscripts are ranks in a filtration sequence.

Let $\mathcal{K}$ and $\mathcal{L}$ be two simplicial complexes with a map $\varphi : \mathsf{vert}\,\mathcal{K} \to \mathsf{vert}\,\mathcal{L}$ which takes vertices of any simplex in $\mathcal{K}$ to the vertexes of a simplex in $\mathcal{L}$. So, if $\sigma = \{\,\mathsf{conv}\,T \mid T = [p_0, p_1, \ldots, p_k]\,\}$ is a simplex in $\mathcal{K}$, then $\mathsf{conv}\,\varphi(T)$ is a simplex in $\mathcal{L}$. A simplicial map $\phi : |\mathcal{K}| \to |\mathcal{L}|$ is the linear extension of a vertex map φ:

$$\phi(x) = \sum_{i=0}^{k} \mu_i \varphi(p_i), \quad \text{where} \quad \left\{ \begin{array}{l} p \in T,\ x \in \sigma \text{ and } \mu_i \text{ is the barycentric} \\ \text{coordinate of } x \text{ that corresponds to } p_i \in T \end{array} \right. .$$

$\mathcal{K}$ and $\mathcal{L}$ are isomorphic or simplicially equivalent if they permit a bijective vertex map φ. In this case, ϕ is a homeomorphism between $|\mathcal{K}|$ and $|\mathcal{L}|$. There is a standard realization for a k-simplex as follows. The standard k-simplex, Δ^k, is the convex hull of $\{e^i\}_{i \in \{0,1,\ldots,k\}}$, where

$$e^i = \{\,(0, \ldots, 1, \ldots, 0) \mid 1 \text{ is in the } i\text{th position},\ i \in I = \{0, 1, \ldots, k\}\,\}$$

is the ith standard basis vector for $\mathbb{R}^k$. For any indexing set $J \subseteq I$, Δ^J is the face of $\Delta^k = \Delta^I$ spanned by $\{e^j\}_{j \in J}$. The standard simplex may be subdivided using the barycenters of its faces to produce the simplicial complex $\mathcal{K}^k$ with $|\mathcal{K}^k| = \Delta^k$. Each non-empty face Δ^J of Δ^k has an associated vertexes in $\mathcal{K}^k$. Δ^J is triangulated by subcomplex $\mathcal{K}^J \subseteq \mathcal{K}^k$ with $|\mathcal{K}^J| = \Delta^J$.

It is possible to define simplicial complexes as purely combinatorial objects, what is crucial from a computation point of view. An abstract simplicial complex is a pair $(\mathcal{A}, \Sigma)$, where $\mathcal{A}$ is a finite set whose elements are referred to as vertices, and where Σ is a family of non-empty subsets of $\mathcal{A}$ so that $\sigma \in \Sigma$ and $\tau \subseteq \sigma$ implies $\tau \in \Sigma$; the elements of Σ are referred to as faces. The sets in $\mathcal{A} = \bigcup \Sigma$ are called abstract simplexes. If a face $\tau \in \Sigma$ consists of $k+1$ elements of $\mathcal{A}$, then $\tau = \{ p_0, p_1, \ldots, p_k \}$ is a k-simplex of Σ with 0-simplexes $p_0, p_1, \ldots, p_k$ as vertices. The dimensions of τ and Σ are $\dim(\tau) = \mathsf{card}(\tau) - 1 = k$ and $\dim(\Sigma) \stackrel{\text{def}}{=} \max\{\dim(\tau) \mid \tau \in \Sigma\}$. Intuitively, a simplicial complex structure on a space is an expression of the space as a union of points, intervals, triangles, and higher dimensional analogues. Abstract simplicial complexes are purely combinatoric objects, which enables computations of topological invariants. A graph is a 1-dimensional abstract simplicial complex. The nerve of $\mathcal{A}$ is an abstract simplicial complex

$$\mathcal{N}(\mathcal{A}) \stackrel{\text{def}}{=} \{ A \subseteq \mathcal{A} \mid \cap A \neq \emptyset \}.$$

A geometric realization of an abstract simplicial complex $(\mathcal{A}, \Sigma)$ is a map

$$r \colon \mathcal{A} \to \mathbb{R}^d \text{ for which } \mathcal{K} = \{ \mathsf{conv}\, r(X) \mid X \in \mathcal{A} \} \text{ is a simplicial complex,}$$

i.e. r is given by $\bigcup_{\sigma \in \Sigma} \mathcal{K}(\sigma)$, where $\mathcal{K}(\sigma) = \{\mathsf{conv}\ e_{r(s)}\}_{s \in \sigma}$ is a simplicial complexes, and e_i denotes the ith standart basis vector. A realization gives us the familiar low-dimensional k-simplices: vertices, edges, triangles, tetrahedrons etc. Every abstract simplicial complex of dimension k has a geometric realization in $\mathbb{R}^d$ for some large enough d.

There is a strong relationship between the geometric and abstract definitions: every abstract simplicial complex, $(\mathcal{A}, \Sigma)$, is isomorphic to the geometric realization of some simplicial complex $\mathcal{K}$. An approximation of a topological space by simplicial complexes is a combinatorial way to describe one, and the homology can be computed using only linear algebra of finitely generated $\mathbb{Z}$-modules.

The smallest subcomplex of $\mathcal{K}$ which contains an another subcomplex $\mathcal{L} \subseteq \mathcal{K}$ is the closure, $\bar{\mathcal{L}}$, of $\mathcal{L}$. The $\mathsf{star}\,\mathcal{L}$ contains all of the cofaces of $\mathcal{L}$, and

$$\mathsf{link}\,\mathcal{L} \stackrel{\text{def}}{=} \overline{\mathsf{star}}\,\mathcal{L} - \mathsf{star}\,(\bar{\mathcal{L}} - \{\emptyset\})$$

is the boundary of $\mathsf{star}\,\mathcal{L}$. Stars and links corresponds to open sets and boundaries in topological spaces.

Assign to each simplex an arbitrary but fixed ordering of its vertices, i.e. impose a total order on the vertex set $\mathcal{A}$. Denote as

$$\Sigma_k \stackrel{\text{def}}{=} \{ \sigma \in \Sigma \mid \mathsf{card}(\sigma) = k+1 \}$$

the subset of Σ with ordered k-simplices as elements. A chain is a collection of abstract simplices which can be ordered so that $\Sigma_0 \subset \Sigma_1 \subset \ldots \subset \Sigma_k$. A k-chain

is the function $c = c^k : \Sigma_k \to \mathbb{Z}$, and can be written as a formal sum:

$$c^k \stackrel{\text{def}}{=} \sum_{i=1}^{N_k} n_i[\sigma_i], \text{ where } \sigma_i \in \Sigma_k,\ n_i \in \mathbb{Z} \text{ and } N_k \text{ is the cardinlity of } \Sigma_k \text{ in } \mathcal{K}.$$

Define the group of k-chains $C_k = C_k(\mathbb{X})$ in $\mathbb{X}$ as the free abelian group on the set of oriented k-simplices Σ_k. I.e. the group C_k is formed by the set of all k-chains together with the operation of addition. A collection of $(k-1)$-dimensional faces of a k-simplex is a $(k-1)$-chain itself and is the boundary, $\sigma\partial_k$, of σ. The boundary of the k-chain is the sum of the boundaries of the simplices in the chain, i.e. $c^k\partial = \sum_{i=1}^{N_k} n_i(\sigma_i^k\partial_k)$. For a k-chain c every-time we have $c\partial_k\partial_{k-1} = 0$. The boundary operator $d_i\sigma : \Sigma_k \to \Sigma_{k-1}$ (for $0 \le i \le k$) maps an ordered k-simplex to a $(k-1)$-chain:

$$d_i\sigma \stackrel{\text{def}}{=} \sum_{i=0}^{k} (-1)^i\, [\, p_0,\, p_2,\, \ldots,\, p_{i-1},\, p_{i+1},\, \ldots,\, p_k\,].$$

The boundary homomorphism $\partial_k : C_k \to C_{k-1}$ is defined linearly on a chain c by action on any simplex $\sigma = [\, p_0,\, p_2,\, \ldots,\, p_k\,] \in c$:

$$\partial_k \stackrel{\text{def}}{=} \sum_i (-1)^i d_i = \sum_i (-1)^i [\, p_0,\, p_2,\, \ldots,\, p_{i-1},\, p_{i+1},\, \ldots,\, p_k\,].$$

The chain complex is the sequence of chain groups connected by boundary homomorphisms:

$$\ldots \xrightarrow{\partial_{k+2}} C_{k+1} \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \longrightarrow \ldots \longrightarrow C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} \emptyset,$$

with $\partial_k\partial_{k+1} = \emptyset$ for all k. The image and the kernel of a boundary homomorphism are

$$\texttt{Im}\ \partial_k = \{c\partial_k \in C_{k-1} \mid c \in C_k\} \text{ and } \texttt{Ker}\ \partial_k = \{c \in C_k \mid c\partial_k = 0\}, \text{ respectively.}$$

A k-chain c is a k-cicle if it has no boundary, i.e. if $c \in \texttt{Ker}\ \partial_k$. Since the k-cycles constitute the kernel of ∂_k, they form a subgroup of C_k, the kth chain group

$$Z_k \stackrel{\text{def}}{=} \texttt{Ker}\ \partial_k = \{\, c \in C_k \mid c\partial_k = 0\,\}.$$

A k-chain c is a k-boundary if it is the boundary of a $(k+1)$-chain, i.e. if $c \in \texttt{Im}\ \partial_{k+1}$. Another name of a k-boundary is a non-homologous k-cycle. Since the k-boundaries lies in the image of ∂_{k+1}, they form an another subgroup of C_k, the kth boundary group

$$B_k \stackrel{\text{def}}{=} \texttt{Im}\ \partial_{k+1} = \{\, c \in C_k \mid c = c'\partial_{k+1} \text{ for } c' \in C_{k+1}\,\}.$$

Since the boundary of a boundary is always empty, $\partial_{k-1}\partial_k\, c = 0$ for all k and for every $c \in C_k$. Thus, defined subgroups are nested: $B_k \subseteq Z_k \subseteq C_k$. The k-cycles

are the basic topological objects that define the presence of k-dimensional holes in the simplicial complex. Also, many k-cycles may characterize the same hole, and cycles possessing the property that their difference is the boundary are said to be homologous. The kth homology group is an algebraic invariant that expressed as the quotient of the cycle group over the boundary group:

$$H_k \stackrel{\text{def}}{=} Z_k/B_k.$$

If $z_1, z_2 \in Z_k$ are in the same homology class, then they are homologous, denoted $z_1 \sim z_2$, and $z_1 = \{ z_2 + b \mid b \in B_k \}$. Since the groups $C_k(X)$ are equipped with the bases Σ_k, ∂_k can be expressed as matrices $D(k)$ which we describe next. Columns of $D(k)$ are parametrized by Σ_k, rows are parametrized by Σ_{k-1}, and, for $\sigma \in \Sigma_k$ and $\tau \in \Sigma_{k-1}$, the entry $D(k)_{\tau\sigma}$ is 0 if $\tau \not\subseteq \sigma$, and is $(-1)^i$ if $\tau \subseteq \sigma$ and if τ is obtained by removing of the ith member of σ. So homology is algorithmically computable for simplicial complexes. This calculations can be performed by putting matrices constructed out of the $D(k)$'s in Smith normal form (at greater length see [10]).

The homology groups are finitely generated and abelian, and can be computed using only linear algebra of finitely generated $\mathbb{Z}$-modules. The fundamental theorem on such groups implies that

$$H_k = \mathbb{Z}^{\beta_k} \oplus T,$$

where $\beta_k = \text{rank } H_k = \text{rank } Z_k - \text{rank } B_k$ is the kth Betti number of a simplicial complex $\mathcal{K}$, it counts the number of k-dimensional holes in $\mathcal{K}$. T is the torison subgroup of H_k, and can be written as the direct sum of finitely many cyclic groups $\mathbb{Z}_k$. This construction is justified by the fact that H_k is an invariant over all simplicial complexes triangulating the same topological space $\mathbb{X} \approx |\mathcal{K}|$. $H_k = H_k(\mathcal{K}) = H_k(\mathbb{X})$ is functorial, i.e. every continuous map $f : X \to Y$ induces a linear transformation $H_k(f) : H_k(X) \to H_k(Y)$. The homology groups are invariants for $|\mathcal{K}|$ and for homotopy equivalent spaces. Formally,

$$\mathbb{X} \simeq \mathbb{Y} \Rightarrow H_k(\mathbb{X}) \cong H_k(\mathbb{Y}) \text{ for all } k.$$

In particular, $\beta_k(\mathbb{X}) = \beta_k(\mathbb{Y}) = \beta_k(\mathcal{K}) = \beta_k$ is invariant over all triangulations of $\mathbb{X}$.

Since H_k is a finitely-generated group, the standard structure theorem states that it decomposes uniquely into a direct sum

$$\bigoplus_{i=1}^{\beta_k} \mathbb{Z} \oplus \bigoplus_{j=1}^{l} \mathbb{Z}_{t_j}, \text{ where } \beta_k, t_j \in \mathbb{Z}, \ t_j | t_{j+1}, \ \mathbb{Z}_{t_j} = \mathbb{Z}/\, t_j\mathbb{Z}\,.$$

The left sum captures the free subgroup and its rank is the kth Betti number, β_k, of $\mathcal{K}$; the right sum captures the torsion subgroup, and the integers t_j are

the torsion coefficients for the homology group. Over a field $\mathfrak{F}$, a module becomes a vector space and is fully characterized by its dimension, the Betti number, and we get a full characterization for torsion-free spaces in this case. For torsion-free spaces in three-dimensions, the Betti numbers have intuitive meaning as a consequence of the Alexander Duality: β_0 counts the number of connected components of the space, β_1 is the dimension of any basis for the tunnels, β_2 counts the number of enclosed spaces or voids. *For instance, the torus is one connected component, has two tunnels, and encloses one void, correspondently,* $\beta_0 = 1$, $\beta_1 = 2$, *and* $\beta_2 = 1$.

Appendix B

The "Singular" Code

Here is represented a code of three "Singular" programs which was created in the scope of the project. The first one is devoted to the computation of persistence Betti numbers via the correspondent barcode. This is the main result of this work, where the structure of this program was described in detail.

Two others programs are dedicated to a calculation of the Gröbner bases, and are realizations of exact and approximative versions of the Buchberger-Möller algorithm. This was initial "stream" of the research, but was postponed for the sake of the main direction. As explained in the end of this chapter, this software can be used for a further development of the project.

B.1 Computation of persistence Betti numbers of noisy point cloud data.

After a launching of the program, the main procedure `Betti` demands for input the name of the input file with the full directory path. For instance:

```
Betti(``/Users/Oleg/PhD/GreatProgram/InputData.txt'');
```

The program return the barcode and, via it, the persistence Betti numbers which corresponds to the input point cloud data X.

```
//******************* The main procedure (Betti)  ******************

proc Betti(string path)
{
    string qhull="qvoronoi <"+path+" o TO ComplexInput";
```

```
    int d=system("sh",qhull);
    string s=read("./ComplexInput");
    list input=inputData(s);
    int N= input[1];
    list FST=input[2];
    list Bett=input[3];
    int i,m;
    list CD,PN,V,BV,E;
    for(i=1;i<=N;i++)
    {
        V[i]=i;
        PN[i]=0;
        CD[i]=list(i);
        BV[i]=list();
    }
    for(i=1;i<=size(FST);i++)
    {
        V=V+list(FST[i]);
        PN[N+i]=0;
        CD[N+i]=list(N+i);
//      BV[N+i]=BS(N+i,V);
    }
    for(i=1;i<=size(FST);i++)
    {
        BV[N+i]=BS(N+i,V);
    }
    for(i=N+1;i<=N+size(FST);i++)
    {
        E=EBD(i,CD,PN,BV);
        m=E[1];
        CD=E[2];
        if(m!=0)
        {
            PN[i]=m;
            PN[m]=i;
        }
    }
    list Pers=Output(PN, Bett);
    write("outputPN",PN);write("outputBett",Bett);
    return("Barcodes-",PN,"Pesistence-",Pers);
}
```

```
//*************** Eliminate-Boundaries (EBD) ******************

proc EBD(int i,list CD,list PN,list BV)
{
    int m;
    list A,Q,E;
    A=YBD(CD[i],BV,Q);
    m=A[1];
    Q=A[2];
    if(m==0)
    {
        E[1]=m;
        E[2]=CD;
    }
    while(m!=0)
    {
        if(PN[m]==0)
        {
            E[1]=m;
            E[2]=CD;
            return(E);
        }
        else
        {
            CD[i]=ADD(CD[i],CD[PN[m]]);
            A=YBD(CD[PN[m]],BV,Q);
            m=A[1];
            Q=A[2];
        }
    E[1]=m;
    E[2]=CD;
    }
    return(E);
}

//**************** Youngest-Boundary of Cascade (YBD) ****************

proc YBD(list X,list BV,list Q)
{
    int i,m;
```

```
    list A;
    for(i=1;i<=size(X);i++)
    {
        if(size(BV[X[i]])>0)
        {
            Q=ADD(Q,BV[X[i]]);
        }
        else
        {
            A[1]=m;
            A[2]=Q;
            return(A);
        }
    }
    for(i=1;i<=size(Q);i++)
    {
        if(m<Q[i])
        {
            m=Q[i];
        }
    }
    A[1]=m;
    A[2]=Q;
    return(A);
}

//********************** Boundary of Simplex (BS) ******************

proc BS(int i,list V)
{
    int j,k;
    list BV,P;
    for(j=size(V[i]);j>=1;j--)
    {
        P=delete(V[i],j);
        k=ID(P,V);
        if(k>0)
        {
            BV=ADD(BV,k);
        }
    }
```

```
    return(BV);
}

//********************** Identification (ID) ***********************

proc ID(list P,list V)
{
    int i,e,k;
    int j=1;
    int s=size(P);
    for(i=1;i<=s;i++)
    {
        while(j<size(V))
        {
            if(size(V[j])==s)
            {
                for(e=1;e<=s;e++)
                {
                    if(P[e]!=V[j][e])
                    {
                        k=1;
                        break;
                    }
                }
                if(k==0)
                {
                    return(j);
                }
                else
                {
                    k=0;
                    j++;
                }
            }
            else{j++;}
        }
    }
    return(0);
}

//******************** Addition of Lists (ADD) **********************
```

```
proc ADD(list A,list B)
{
    int i,j,k;
    for(i=1;i<=size(B);i++)
    {
        for(j=1;j<=size(A);j++)
        {
            if(B[i]==A[j])
            {
                k=j;
                break;
            }
        }
        if(k==0)
        {
            A[size(A)+1]=B[i];
        }
        else
        {
            A=delete(A,k);
            k=0;
        }
    }
    return(A);
}

//********* Read OFF Format Data of "Qhull" (inputData) ************

proc inputData(string s)
{
    int i,j,q,l;
    list A,L,FST,Bett;
    string V,s1,s2;
    V=s[1];
    execute("int Dim="+V+";");
    A=NLE(s);
    int N=A[1];
    s1=A[2];
    A=NEL(s1);
    int F=A[1];
```

```
    s2=A[2];
    for(i=1;i<=N+1;i++)
    {
        s1=s2[find(s2,newline)+1,size(s2)];
        s2=s1;
    }
    while(j<F)
    {
        V=s1[1,find(s1," ")-1];
        execute("q="+V+";");
        for(i=1;i<=q-1;i++)
        {
            A=NEL(s1);
            l=A[1];
            s1=A[2];
            L=L+list(l+1);
            s2=s1;
        }
        s2=s1[find(s1," ")+1,size(s1)];
        V=s2[1,find(s2,newline)-1];
        s1=s2[find(s2,newline)+1,size(s2)];
        execute("l="+V+";");
        L=L+list(l+1);
        j=j+1;
        FST[j]=L;
        L=list();
        Bett[N+j]=q-1;
    }
    for(i=1;i<=N;i++)
    {
        Bett[i]=0;
    }
    return(list(N,FST,Bett));
}

//*************** New element in a line (NEL) ******************

proc NEL(string s1)
{
    int Val;
    string s2,V;
```

```
    list A;
    s2=s1[find(s1," ")+1,size(s1)];
    V=s2[1,find(s2," ")-1];
    execute("Val="+V+";");
    A[1]=Val;
    A[2]=s2;
    return(A);
}

//***************** New line element (NLE) ********************

proc NLE(string s2)
{
    int f,Val;
    string s1,V;
    list A;
    s1=s2[find(s2,newline)+1,size(s2)];
    f=find(s1," ");
    V=s1[1,f-1];
    execute("Val="+V+";");
    A[1]=Val;
    A[2]=s1;
    return(A);
}

//********* Output Betti numbers persistence (Output) ************

proc Output (list PN, list Bett)
{
    int i;
    list Pers;
    for(i=1;i<=size(PN);i++)
    {
        if(PN[i]!=0)
        {
            if(PN[i]>=i)
            {
                Pers[i]=string("Betti(",Bett[i],") has persistence ",PN[i]-i);
            }
            if(PN[i]<i)
            {
```

```
                Pers[i]=string("Betti(",Bett[i],") has negative persistence ",
                PN[i]-i, ", just marks a ‘‘destroyer’’ for the homology class");
            }
        }
    }
    return(Pers);
}
```

B.2 Computation of the Gröbner basis by a realization of the Buchberger-Möller algorithm.

There is huge amount of literature devoted to the Gröbner basis and to the Buchberger-Möller algorithm for it computation, see for instance [1], [17], [19]. For an introduction to this area see well written [23], [24], and [7].

The main procedure `BMA` stipulates for input a matrix whose rows are points coordinates. For example:

```
matrix P [7] [2] =1,2,3,4,5,6,7,8,9,10,11,12,13,14;
BMA(P);
```

The program return the Gröbner basis which corresponds to the input points. Also for output we have generators.

```
//******************* The main procedure (BMA)  ******************

ring K=(real,30),(x,y),(c,dp);
proc BMA(matrix P)
{
    int n=nvars(basering);
    int s=nrows(P);
    int i,j,k,e,sizem,sizes;
    list L=1;
    list G,Q,AA;
    poly t,h;
    vector V;
    ideal S,A,GG,LL;
    module M;
    while(size(L)!=0)
    {
        t=L[size(L)];
        L=delete(L,size(L));
```

```
V=Plugin(P,t);
AA=Calc(V,M,s);
V=AA[1];
A=AA[2];
if(V==0)
{
    for(i=1;i<=ncols(S);i++)
    {
        h=h+A[i]*S[i];
    }
    G=insert(G,(t-h));h=0;
}
else
{
    M[size(M)+1+sizem]=V;
    if(V==0)
    {
        sizem=1;
    }
    else
    {
        sizem=0;
    }
    for(i=1;i<=ncols(S);i++)
    {
        h=h+(A[i]*S[i]);
    }
    S[size(S)+1+sizes]=t-h;
    if(t-h==0)
    {
        sizes=1;
    }
    else
    {
        sizes=0;
    }
    h=0;
    Q=insert(Q,t);
    for(j=1;j<=size(G);j++)
    {
        GG[j]=lead(G[j]);
```

```
}
for(j=1;j<=size(L);j++)
{
    LL[j]=L[j];
}
attrib(GG,"isSB",1);attrib(LL,"isSB",1);
for(j=1;j<=n;j++)
{
    for(i=1;i<=size(L);i++)
    {
        if((NF(var(j)*t,LL)!=0)&&(NF(var(j)*t,GG)!=0))
        {
            for(k=1;k<=size(L);k++)
            {
                if(L[k]<=(var(j)*t))
                {
                    L=insert(L,var(j)*t,k-1);
                    e=1;
                    break;
                }
            }
            if(e<>1)
            {
                L=insert(L,var(j)*t);
            }
            e=0;
            break;
        }
    }
}
if(size(L)==0)
{
    for(j=1;j<=n;j++)
    {
        if(NF(var(j)*t,GG)!=0)
        {
            e=e+1;
            L[e]=var(j)*t;
        }
    }
    e=0;
```

```
            }
            GG=0;LL=0;
        }
    }
    return(G,Q);
}

//******************  (Plugin)  ******************

proc Plugin(matrix P,poly t)
{
    int n=nvars(basering);
    int s=nrows(P);
    int i,j;
    poly f;
    vector V;
    for(i=1;i<=s;i++)
    {
        f=t;
        for(j=1;j<=n;j++)
        {
            f=subst(f,var(j),P[i,j]);
        }
        V=V+f*gen(i);
    }
    return(V);
}

//******************  (Calc)  ******************

proc Calc(vector V,module M,int s)
{
    int r=size(M);
    int i;
    list W;
    ideal B;
    module N;
    for(i=1;i<=r;i++)
    {
        N[i]=M[i]+gen(s+i);
    }
```

```
    attrib(N,"isSB",1);
    option(redSB);
    vector A=reduce(V,N);
    option(noredSB);
    W[1]=A[1..s];
    for(i=1;i<=r;i++)
    {
        B[i]=-A[s+i];
    }
    W[2]=B;
    return (W);
}
```

B.3 Computation of the Gröbner basis by a realization of the approximative version of the Buchberger-Möller algorithm.

As opposed to the exact Buchberger-Möller algorithm, the approximative variant has another structure, uses so called singular value decomposition*, and, in addition, has for input some approximative parameter which defines an extent of an approximation. For the sake of simplicity, in the represented below program such a parameter is installed inside of a body of the main procedure as the number `eps` with the value is equal to 0.000000007. Of course, the parameter can be easily led out for preliminary input, and, in this case, the input for the main procedure `ABM` look like in the following example:

```
matrix P [7] [2] = 14,13,12,11,10,9,8,7,6,5,4,3,2,1;
number eps=0.000000007;
BMA(P,eps);
```

The program return the approximative Gröbner basis which corresponds to the input points and to the approximative parameter value. Also for output we have generators.

```
//******************** The main procedure (ABM)  ******************

ring K=(real,30),(x,y),(c,dp);
LIB "matrix.lib";
```

*The procedure `SVD` in the program is compiled and installed inside "Singular". `SVD` is the $C++$ procedure from the free open source software, it will be included to the next version of "Singular".

```
LIB "aksaka.lib";
proc ABM(matrix P)
{
    int n=nvars(basering);
    int s=nrows(P);
    int i,j,k,e,sizes;
    number eps=0.000000007;
    number f=eps+1;
    list L=1;
    list G,Q,AA;
    poly t,h;
    ideal S,A,GG,LL;
    vector V;
    matrix M[s][0];
    while(size(L)!=0)
    {
        t=L[size(L)];
        L=delete(L,size(L));
        V=Plugin(P,t);
        if(ncols(M)!=0)
        {
            AA=TLS(V,M,s);
        }
        V=AA[1];
        A=AA[2];
        f=AA[3];
        if(f<=eps)
        {
            for(i=1;i<=ncols(S);i++)
            {
                h=h+A[i]*S[i];
            }
            G=insert(G,(t-h));
            h=0;
            i=0;
            while(i<size(L))
            {
                i++;
                if(L[i]/t!=0)
                {
                    L=delete(L,i);
```

```
                i--;
            }
        }
    }
    else
    {
        if(ncols(M)==0)
        {
            matrix M[s][1]=V;
        }
        else
        {
            M=concat(M,V);
        }
        for(i=1;i<=ncols(S);i++)
        {
            h=h+(A[i]*S[i]);
        }
        S[size(S)+1+sizes]=t-h;
        if(t-h==0)
        {
            sizes=1;
        }
        else
        {
            sizes=0;
        }
        h=0;
        Q=insert(Q,t);
        for(j=1;j<=size(G);j++)
        {
            GG[j]=lead(G[j]);
        }
        for(j=1;j<=size(L);j++)
        {
            LL[j]=L[j];
        }
        attrib(GG,"isSB",1);attrib(LL,"isSB",1);
        for(j=1;j<=n;j++)
        {
            for(i=1;i<=size(L);i++)
```

```
                {
                    if((NF(var(j)*t,LL)!=0)&&(NF(var(j)*t,GG)!=0))
                    {
                        for(k=1;k<=size(L);k++)
                        {
                            if(L[k]<=(var(j)*t))
                            {
                                L=insert(L,var(j)*t,k-1);
                                e=1;
                                break;
                            }
                        }
                        if(e<>1)
                        {
                            L=insert(L,var(j)*t);
                        }
                        e=0;
                        break;
                    }
                }
            }
            if(size(L)==0)
            {
                for(j=1;j<=n;j++)
                {
                    if(NF(var(j)*t,GG)!=0)
                    {
                        e=e+1;
                        L[e]=var(j)*t;
                    }
                }
                e=0;
            }
            GG=0;
            LL=0;
        }
    }
    return(G,Q);
}

//************************ Auxiliary Procedures ******************
```

```
proc Plugin(matrix P,poly t)
{
    int n=nvars(basering);
    int s=nrows(P);
    int i,j;
    poly f;
    vector V;
    for(i=1;i<=s;i++)
    {
        f=t;
        for(j=1;j<=n;j++)
        {
            f=subst(f,var(j),P[i,j]);
        }
        V=V+f*gen(i);
    }
    return(V);
}

//..............................................................

proc TLS(vector V,matrix M,int s)
{
    list W;
    int i,j,k;
    int r=ncols(M);
    number f,f1,f2;
    number epsilon=1/1e2147483647;
    matrix B[s][r+1]=concat(M,-V);
    matrix v[s][1]=V;
    matrix b[r][1];
    matrix u[s][1];
    matrix uu[s][1];
    list L=system("svd",B);
    for(k=1;k<=3;k++)
    {
        for(i=1;i<=nrows(L[k]);i++)
        {
            for(j=1;j<=ncols(L[k]);j++)
            {
```

```
            if(absValue(leadcoef(L[k][i,j]))<epsilon)
            {
                L[k][i,j]=0;
            }
        }
    }
}
matrix l[r+1][1]=(transpose(L[3]))[r+1];
int rr=r+1;
while(absValue(leadcoef(l[r+1,1]))<=epsilon)
{
    rr=rr-1;
    l=(transpose(L[3]))[rr];
}
matrix n[s][1]=(L[1]*L[2])[rr];
for(i=1;i<=r;i++)
{
    b[i,1]=l[i,1]/l[r+1,1];
}
for(i=1;i<=s;i++)
{
    u[i,1]=-n[i,1]/l[r+1,1];
}
for(i=1;i<=s;i++)
{
    uu[i,1]=v[i,1]-(M*b)[i,1];
}
for(i=1;i<=s;i++)
{
    if(leadcoef(u[i,1])>=epsilon)
    {
        k=7;
        break;
    }
}
if(k==7)
{
    for(i=1;i<=s;i++)
    {
        f1=f1+leadcoef(u[i,1]^2);
        f2=f2+leadcoef(V[i]^2);
```

```
        }
        f=(wurzel(f1))/(wurzel(f2));
    }
    else
    {
        f=0;
    }
    k=0;
    f1=0;
    f2=0;
    V=0;
    for(i=1;i<=size(u);i++)
    {
        V=V+u[i,1]*gen(i);
    }
    W[1]=V;
    W[2]=b;
    W[3]=f;
    return(W);
}
```

The Gröbner basis can be considered as a one of outlooks for the farther project development, and represent a direction which gives a fresh perspective as well as a new arsenal of computational tools to attack an old and significant problem in data analysis. The Buchberger-Möller algorithm can be used for a computation of the so called multidimensional persistance; the Gröbner basis enable to reconstruct the entire multidimensional persistence vector space, and provide a convenient way for a computation of the rank invariant. Since this matters are beyond the scope of this work, we refer to [5] for the details (see also [4], page 293).

Appendix C

A Reference Mapping Way and a Representative Graph

I would like to represent here an alternative way of a similarity complex construction, and a method of the complex visual depiction. This techniques was not reflected in the created in the scope of the work software, but, nevertheless, are interesting by themselves.

C.1 Filtering.

We should come up ourselves with additional input information. First of all, we need to define a so called reference map or a filter which is chosen for a partition of the given cloud of points X. This is a real valued continuous function $f\colon X \to Z$ to a fixed reference space, Z, whose dimension will be an upper bound for the dimension of the similarity simplicial complex. This f can be a well known function which reflect geometric properties of the data set, or can be a user defined function which is chosen in order to understand how these properties interact with it.

A polynomial function: it is naturally to choose a function from the polynomial ring $\mathbb{R}[x_1, x_2 \ldots, x_n]$.* Advantages of this choice are possibility to use the well developed theory, and also a comparative simplicity of a treatment with such functions.

*In order to avoid a dependence of the radius-vectors signs in the case of more complicated parametrizations, it is reasonable to take the ring $\mathbb{R}[x_1^2, x_2^2 \ldots, x_n^2]$ instead.

Below is represented few functions which carry interesting geometric information about X in general. All of these functions are rely on the ability to compute distances between points, so it is important to generate filters directly from the metric (see [35]).

Density estimator: a non-negatve function on X, which reflect useful geometrical information about the given data, it can be produced by any density estimator applied to X.

Gaussian kernel:

$$\mathbf{f}_\varepsilon(\mathbf{x}) = \mathbf{C}_\varepsilon \sum_{\mathbf{y}} \left(\frac{\exp(-\mathbf{d^2}(\mathbf{x},\mathbf{y}))}{\varepsilon} \right),$$

where $x, y \in X$, C_ε is a constant such that $\int f_\varepsilon \, dx = 1$, and $\varepsilon > 0$ control the smoothness of the estimation of the density function on X.

Eccentricity:

$$\mathbf{E_p}(\mathbf{x}) = \left(\frac{\sum_{\mathbf{y}\in\mathbf{X}} \mathbf{d}(\mathbf{x},\mathbf{y})^{\mathbf{p}}}{\mathbf{N}} \right)^{\frac{1}{\mathbf{p}}},$$

where $x, y \in X$ and $1 \leq p < +\infty$. Also $E_\infty(x) = \max_{x' \in X} d(x, x')$ for $p = +\infty$. The idea is to identify points which are far from the "center" without identifying an actual center point, and refers to a data depth.

A (normalized) graph Laplacian matrix:

$$\mathbf{L}(\mathbf{x},\mathbf{y}) = \frac{\mathbf{w}(\mathbf{x},\mathbf{y})}{\sqrt{\sum_{\mathbf{z}} \mathbf{w}(\mathbf{x},\mathbf{z})}\sqrt{\sum_{\mathbf{z}} \mathbf{w}(\mathbf{y},\mathbf{z})}},$$

which eigenvectors gives us a set of orthogonal vectors on the vertex set of the graph, which encode interesting geometric information and can be used as filter functions on the data. The vertex set of Graph Laplasian is the set of all points in X, and the weight of the edge between points $x, y \in X$ is $w(x, y) = k(\, d(x, y))$, where k is some "smoothing kernel" such as a Gaussian kernel (at greater length see [40]).

Filters determines reference spaces to which we produce a map. In the simplest case, $Z = \mathbb{R}$ but it can be $\mathbb{R}^2$, the unit circle $\mathbb{S}^1$ in the plane, or any another space where a covering could be constructed relatively easily. Then we have to establish some parameters for the construction method of such a covering.

We start with a finding of the range of the function restricted to the given points. For the sake of simplicity, let us consider $Z = \mathbb{R}$. So suppose that we are given a space equipped with a continuous map $f : X \to \mathbb{R}$ and a covering

$$\mathcal{U} = \{\, \{U_i\}_{i\in I} \mid F \subseteq \cup_{i\in I} U_i,\ F = \mathtt{Im}(f) \,\}$$

with some integer indexing set I. In order to use this construction, one must develop methods for coverings creation. For instance, the simplest cover can be constructed by dividing F into a smaller overlapping intervals, that also gives us a possibility to parameterize the covering by two parameters which, in due course, can be used for a resolution control: the length of the smaller intervals and the percentage of a overlap between successive intervals.

It is natural to represent the covering method as covering of F by open balls

$$B_\varepsilon = \{\, y \in \mathbb{R}^d \mid d(x,y) < \varepsilon \,\}$$

with a positive real radius ε. Below are considered two modes of such constructions for two different input values: distance between balls centers, R, and a positive real radius, ε. These parameters can be interpreted as an "amount of blurring" applied to Z.

- The covering $\mathcal{U}[\,R,\varepsilon\,]$ of $Z=\mathbb{R}$ consist of all intervals of the form

  $$U_i = [\,iR-\varepsilon,\,(i+1)\,R+\varepsilon\,].$$

  Here we have two parameters for a resolution control, and the covering dimension will be 1 while $\varepsilon < \frac{R}{2}$, since there are will be no non-empty threefold overlaps in this case. It is easy to obtain a corresponding covering of $\mathbb{R}^n$ by multiplying the intervals.

- Let an integer $N \geq 2$. The covering $\mathcal{U}[\,N,\varepsilon\,] = \{U_j\}_{0\leq j\leq N}$ of $Z=\mathbb{S}^1$ is defined by the setting

  $$U_i = \left\{ (\cos(x),\,\sin(x)) \mid x \in \left[\, \frac{2\pi i}{N}-\varepsilon,\, \frac{2\pi i}{N}+\varepsilon \,\right] \right\}, \text{ whenever } \varepsilon > \frac{\pi}{N}.$$

In order to use this predefined covering of the reference space for our purposes, we pull back to a covering of X by the set $f^{-1}(\mathcal{U}[\,N,\varepsilon\,])$. Since f is a continuous function, the fibers that form it domain,

$$X_i = f^{-1}(U_i) = \{\, x \in X \mid f(x) \in U_i \,\},$$

also form the corresponding open covering $\{X_i\}_{i\in I}$ of X.

C.2 Clustering.

Now it is necessary to describe a method for a transporting of this construction from the setting of topological spaces to the setting of the points cloud. Since elements X_i of the covering of X might be in several connected components, we need to treat each such a connected component as a separate subset in $\{X_i\}_{i\in I}$.

We use so called clustering in order to avoid this kind of difficulty. For all $i \in I$, let us consider the decomposition

$$f^{-1}(U_i) = \bigcup_{j=1}^{J_i} V_{i,j}$$

of X_i into its path connected components, where J_i is the number of such components in X_i. Each X_i is presented now as the union of the disjoint sets, $V_{i,j}$, treated as the representative points which makes up a vertex set. Finally, we can represent the obtained from $\mathcal{U}$ covering of X as

$$\{X_i\}_{i\in I} = \{V_{i,j}\}_{i\in I,\, j\in[1,\ldots,J_i]} \stackrel{\text{def}}{=} \{Q_\alpha\}_{\alpha\in A}\,, \text{ where } A = \left[1, \ldots, \sum_i i J_i\right].$$

Let us call this subsets of sampled points Q_α as clusters. This construction depend on the filter as well as on values of parameters of the covering of the parameter space.

Clustering refer to the process of a partitioning of a data set into a number of parts which are recognizably distinguishable from each other. So the goal of clustering is to identify high density regions which are separated by low density regions. A finding of a good clustering is a challenging problem and is a fundamental issue in a computing of the similarity simplicial complex. There are a lot of different principles of a clustering construction (e.g. see [21] or diffuse speculations about this in [41]).

Usually, algorithms require parameters to be set before a output is received. Such parameters often designates arbitrarily, but the arbitrariness of various thresholds choices does not go with a lack of robustness. Some work in clustering theory has been done in trying to determine the optimal choice of ε, but it is much more informative to consider the so called hierarchical clustering (see [22]). This kind of clustering combines data objects into clusters, those clusters into larger clusters, and so forth, creating a hierarchy. A tree representing this hierarchy of clusters is a dendrogram which provide a summary of the behavior of clustering under all possible values of the parameter ε at once.

For instance, one can construct data sets which have been thresholded at two different values, and the behavior of clusters under the inclusion of the set with tighter threshhold into the one with the looser threshold is informative about what is happening in the data set. Of course, we can play even a more complicated "game" when we have more then one threshold parameters or when we associate many functions[†] with each data point instead of just one. Individual data objects

[†]As an example of $Z = \mathcal{S}^1$, consider a parameter space defined by two functions f and g which are related such that $f^2 + g^2 = 1$. A very simple covering for such a space is generated by considering of overlapping intervals of equal size. If we used M functions, let $\mathbb{R}^M$ to be our parameter space. After this we would have to find a covering of an M dimensional hypercube which is defined by the ranges of M functions.

are the leaves of the tree, and the interior nodes are nonempty clusters. This allow us to explore the data at different levels of granularity.

Hierarchical clustering methods are based on linkage metrics results in clusters of proper shapes, and are categorized into so called agglomerative (bottom-up) and divisive (top-down) approaches. The agglomerative clustering starts with singleton clusters and recursively merges two or more of the most similar clusters. A divisive clustering starts with a single cluster containing all data points and recursively splits that cluster into appropriate subclusters. The process continues until a stopping criterion is achieved, e.g. as such a criterion could be the requested a number k of clusters.

The advantages of hierarchical clustering:
- *flexibility regarding to the level of granularity;*
- *ease of handling any form of similarity or distance;*
- *applicability to any attribute type.*

The disadvantages of hierarchical clustering:
- *the difficulty of choosing the right stopping criteria;*
- *most hierarchical algorithms do not revisit (intermediate) clusters once they are clustered.*

There are really overwhelming amount of different clustering techniques, algorithms (e.g. BIRCH, AGNES, MST, CHAMELEON and others, see a survey in [22]), and there is even open source clustering software[‡]. Classification of clustering algorithms is neither straightforward nor canonical. In fact, the different classes of algorithms overlap.

After the partitioning of X to subsets which corresponds to elements of the covering $\{X_i\}_{i\in I}$, we can use the interaction of the formed by this way subsets between each other for an approximate representation of the exploration data.

C.3 The cluster complex.

In order to switch from the topological construction to a point cloud, we apply standard clustering algorithms to subsets of the given data and use then the interaction of the partial clusters with each other, just like we did before with elements of the covering.

The cluster complex of the covering $\{X_i\}_{i\in I}$ is the nerve of the covering of X by sets which are path connected components of each $\{X_i\}$. In another

[‡]For example see [38].

words, this is the abstract simplicial complex $\mathcal{C}(\{X_i\}_{i\in I}) = \mathcal{C}(X, \{X_i\}_{i\in I})$ whose vertex set is the defined above indexing set A, and where a family $\{\alpha_0, \alpha_1, \ldots, \alpha_k\}$ spans a k-simplex if and only if

$$Q_{\alpha_0} \cap Q_{\alpha_1} \cap \ldots \cap Q_{\alpha_k} \neq \emptyset.$$

So we should find corresponding clusters $\{Q_\alpha\}_{\alpha\in A}$, each of which we treat as a vertex in our complex whenever $Q_{\alpha_i} \bigcap Q_{\alpha_j} \neq \emptyset$. And then, whenever $\{Q_{\alpha_0}, Q_{\alpha_1}, \ldots, Q_{\alpha_k}\}$ are overlapping, we add a $(k{-}1)$-simplex to the complex.

The set map $A \to I$ yields the map of simplicial complexes

$$\mathcal{C}(\{Q_\alpha\}_{\alpha\in A}) \longrightarrow \mathcal{N}(\{X_i\}_{i\in I}).$$

This is kind of projection, and $\mathcal{C}(\{X_i\})$ is more sensitive then $\mathcal{N}(\{X_i\})$. Actually, $\mathcal{C}(\{X_i\})$ is homeomorphic to X, while $\mathcal{N}(\{X_i\})$ is not.

As mentioned before, our input set is equipped with the Euclidean metric. Let us assume that clusters Q_α is represented by equal balls with a "large enough" radius, i.e. that one covering of X is given by the family

$$B_\varepsilon(X) \stackrel{\text{def}}{=} \{B_\varepsilon(x) \mid x\in X,\, \varepsilon>0,\, Q_\alpha \subseteq B_\varepsilon\}.$$

One can construct the nerve $\mathcal{N}_\varepsilon = \mathcal{N}(B_\varepsilon(X))$. According to the definition, for $\varepsilon \geq 0$ the cluster complex $\mathcal{C}_\varepsilon(X) = \mathcal{C}(B_\varepsilon(X))$ includes the k-simplex $\sigma = [p_0, p_1, \ldots, p_k]$ if and only if $B_\varepsilon(p_i)$ have non-empty common intersections. In our case of Euclidean data, there is the following consequence of the nerve theorem (see [3]).

Theorem C.1. *For a finite set of points in Euclidean space, $X \subset \mathbb{R}^d$, there is a number $\delta > 0$ such that $\mathcal{C}_\varepsilon(X)$ is homotopy equivalent to X whenever $\varepsilon \leq \delta$. Moreover, if Y is sampled from X, and $B_\varepsilon(Y)$ covers and is homotopy equivalent to X, then the subcomplex $\mathcal{C}_\varepsilon(Y) \subseteq \mathcal{C}_\varepsilon(X)$ on the vertices in Y is also homotopy equivalent to X, and therefore $\mathcal{C}_\varepsilon(Y)$ has the same homology as X.*

We finish this section with three simple examples of constructions which produces a multiresolution structure.

Example C.1.1. *If $\{\cup B_{\varepsilon_i} \mid X \subseteq \cup B_{\varepsilon_i},\, i\in I\}$ is a representation of the covering by balls with equal radiuses, then we get a diagram*

$$\mathcal{C}_{\varepsilon_0}(X) \xrightarrow{\lambda_0} \mathcal{C}_{\varepsilon_1}(X) \xrightarrow{\lambda_1} \ldots \xrightarrow{\lambda_{n-1}} \mathcal{C}_{\varepsilon_n}(X).$$

of inclusions of upper complexes into lower complexes, since the upper one corresponds to a smaller parameter value ε then the lower one.

Example C.1.2. *Let consider the covering* $\{\cup B^R_{\varepsilon_i} \mid i\in I\}$ *of* $Z=\mathbb{R}$ *with the integer indexing set, where* R *is the second parameter with the meaning of a distance between balls centers. The identity map on* $\mathbb{Z}$ *for some* $\varepsilon_0\leq\varepsilon_1\leq\ldots\leq\varepsilon_n$ *provides a map of coverings*

$$\cup B^R_{\varepsilon_0}\to\cup B^R_{\varepsilon_1}\to\ldots\to\cup B^R_{\varepsilon_n}$$

which consists of inclusions of intervals into the intervals with the same center but with a lager diameter. Finally, we get the diagram

$$\mathcal{C}(f^{-1}(\cup B^R_{\varepsilon_0}))\xrightarrow{\lambda_0}\mathcal{C}(f^{-1}(\cup B^R_{\varepsilon_1}))\xrightarrow{\lambda_1}\ldots\xrightarrow{\lambda_{n-1}}\mathcal{C}(f^{-1}(\cup B^R_{\varepsilon_n})).$$

Example C.1.3. *For the equal-radius balls covering* $\{\cup B^{R_i}_{\varepsilon} \mid i\in I\}$ *of* $Z=\mathbb{R}$*, consider a map of coverings* $B^R_\varepsilon \xrightarrow{\lambda_{R\to 2R}} B^{2R}_\varepsilon$ *induced by the map of integers* $k\to\lfloor\frac{k}{2}\rfloor$*. This gives us a diagram of simplicial complexes*

$$\ldots\xrightarrow{\lambda_{R/8}}\mathcal{C}(f^{-1}(B^{R/4}_\varepsilon))\xrightarrow{\lambda_{R/4}}\mathcal{C}(f^{-1}(B^{R/2}_\varepsilon))\xrightarrow{\lambda_{R/2}}\mathcal{C}(f^{-1}(B^{R}_\varepsilon)).$$

As farther to the left one moves here, as the coverings of X*, and therefore the resolution of the picture of the object under investigation, becomes more and more refined.*

The complex $\mathcal{C}(Z,R,\varepsilon)$ captures large-scale topology features and ignores small-scale ones. The extent of scale is defined in terms of the parameters which are nested. So natural inclusion maps

$$\mathcal{C}(f^{-1}(\cup B^R_\varepsilon))\to\mathcal{C}(f^{-1}(\cup B^{R'}_{\varepsilon'})),$$

whenever $R\leq R'$ or/and $\varepsilon\leq\varepsilon'$, induce corresponding maps between homology groups. Finally, we have a similarity theorem or heuristic relating $H_k(\Xi)$ to the persistence homology group $\mathtt{Im}\,[\,H_k(Z,R,\varepsilon)\to H_k(Z,R',\varepsilon')\,]$ under reasonable geophisical sampling and for some choice of the parameters.

Persistence homology here study the full system of the homology groups $H_k(Z,R,\varepsilon)$ together with the induced maps between them by varying the nested parameters over a large range. By this way, we decant features which no longer visible at scale R' or/and ε'.

C.4 The Mayer-Vietoris blowup.

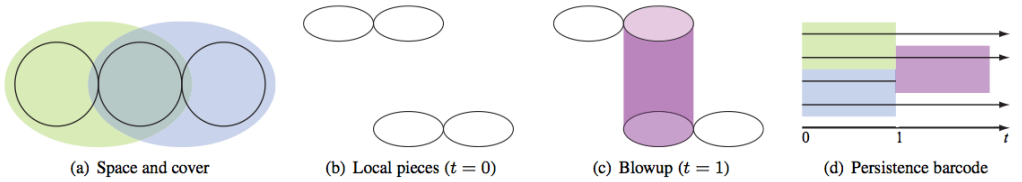
Let $\mathcal{K}^I$ denote the abstract simplicial complex with a vertex set I. For any non-empty $J\subseteq I$, subcomplex $\mathcal{K}^J\subseteq\mathcal{K}^I$ is the face spanned by J. Here we can

define the so called associated to the covering Mayer-Vietoris blowup of X as the subspace

$$\mathfrak{M}(X,\{X_i\}_{i\in I}) \stackrel{\text{def}}{=} \bigcup_{\emptyset\neq J\subseteq I} K^J \times \bigcap_{j\in J} X_j \;\subseteq\; K^I \times X.$$

Here we use those fact that the map $\mathfrak{M}(X,\{X_i\}_{i\in I}) \to X$ is a homotopy equivalence when X has the homotopy class of the finite complex and when all X_i are open sets ([18],[34]).

On the below picture example the graph containing three cycles is covered by two sets which is blown up into two pieces, each with two *1*-cycles. Since the middle cycle of the original space is contained in the intersection of the cover sets, it exists in both local pieces. To recover the global topology, we equate the two copies of the middle cycle by gluing a cylinder to them. The resulting construction, the so called Mayer-Vietoris blowup complex, has the same number of cycles as the original space but also incorporates the geometric cover information within its structure.



(a) Space and cover (b) Local pieces ($t=0$) (c) Blowup ($t=1$) (d) Persistence barcode

Fig. C.1: Given a space equipped with a cover (a), we first blow up the space into local pieces (b) and then glue back the pieces to get the blowup complex (c), giving us a filtration consisting of two complexes at times $t=0$ and $t=1$, respectively. The persistence barcode barcode (d) localizes the topology of the original space with respect to the cover.

In constructing of the blowup complex there are no any tear or gluing manipulations, but only stretching of certain pieces. Therefore, blowup complex has the same topology as the original space. This fact reflected in the below lemma (see [34]).

Lemma C.2. *The projection* $\pi_X\colon \mathfrak{M}(X,\{X_i\}) \to X$ *is a homotopy equivalence in the following cases:*

- $\{X_i\}$ *is an open covering of a normal space, e.g. any subspace of* $\mathbb{R}^d$*;*
- $\{X_i\}$ *is a covering of a simplicial complex by subcomplexes.*

Therefore, π_X *induces an isomorphism at the homology level. That is*

$$\mathfrak{M}(X,\{X_i\}) \simeq X, \text{ and } H_*(\mathfrak{M}(X,\{X_i\})) \cong H_*(X).$$

So the geometry which is contained within the cover can be incorporated into homology by building the blowup complex and computing its persistent homology.

Obtained by this way so called localization reflects the quality of the given cover and gives better description of the cover, that portray the geometry of the space via the attributes location [45].

C.5 The similarity graph.

We start with a simple definition.

A graph , $\mathfrak{G}$, is a subset of $\mathbb{R}^3$ which is made up of a finite collection of points $\{ v_0, v_1, \ldots, v_n \}$, called vertices, together with joins these vertices straight-line segments $\{ e_0, e_1, \ldots, e_m \}$, called edges, and which satisfy the following intersection conditions:

1) the intersection of distinct edges either is empty or consists of exactly one vertex;
2) if edge and a vertex intersect, then the vertex is an endpoint of the edge.

More explicitly, an edge $[v_i, v_j]$ joining vertices v_i and v_j is the set of points

$$\{ x \in \mathbb{R}^3 \mid x = tv_i + (1-t)v_j, \ 0 \leq t \leq 1 \}.$$

A path in $\mathfrak{G}$ is an ordered sequence of edges of the form

$$\{ [v_0, v_1], [v_1, v_2], \ldots, [v_{l-1}, v_l] \}.$$

Since any k-simplicial complex can be embedded in $(2k+1)$-dimensional Euclidean space, any 1-dimensional abstract simplicial complex can be represented as a graph.

One method of a point cloud clustering is the so-called single-linkage clustering, where a graph is constructed whose vertex set is the set of points in the cloud, and where also two such points are connected by an edge if their distance is less or equal then some ε. The parameter ε can be used for a control of resolution. Here shorter edges are required to connect points within each cluster, but relatively longer edges are required to merge the clusters. So the number of clusters is obtained automatically, and it is not necessary to require specifying one beforehand. For instance, implemented in [35] algorithm returns a vector $\mathcal{V} \in \mathbb{R}^{N-1}$ which holds the length of the edge which was added to reduce the number of clusters by one at each step in the algorithm. Of course, it is possible to define the number of clusters first and to obtain the correspondent parameter ε then.

Since the represented in the work complexes contains information about multiresolution structure of the input data, it serve as a source for a construction of

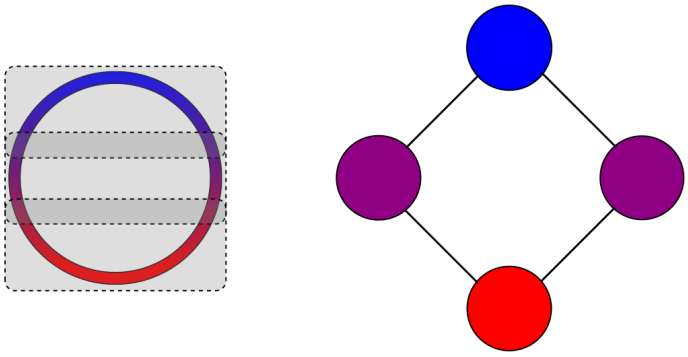


Fig. C.2: The similarity graph created on base of a simple cover of a circle.

the similarity graph. A visualization[§] conducts to better qualitative understanding of the noisy data set, and the graph representation of the higher dimensional approximaton simplicial complex is one of ways for such a qualitative representation of $\Xi \subset \mathbb{R}^3$. Each vertex of this graph, $\mathfrak{G}$, are nodes of the simplicial complex, is corresponded to the clusters, and labeled by color and size. The size of each node is proportional to the cardinality of the cluster complex elements Q_α. The color indicates[¶] the value of the reference map, f, at a representative point in the corresponding set of the covering, $\{X_i\}_{i \in I}$. For example, as the representative point could be taken a barycenter or, perhaps, a suitable average taken over the set of the ponts belonging to each cluster.

It is pretty widespread to model the data points and their distances by a neighborhood graph, just because of a visual obviousness. It is also possible to use the graph for a reflection of object shape changes with the course of time. A clustering can be reduced to standard graph algorithms: in the easiest case, one can simply define clusters as connected components of the graph, and, alternatively, one can try to construct minimal graph cuts which separate the clusters from each other. Anyhow, constructing the similarity graph even for a finite sample from some larger underlying space is not a trivial task (several popular constructions see e.g. in [40], [27]).

In order to be able to merge or split subsets of points rather than individual points, the distance between individual points has to be generalized to the distance between subsets. Such a derived proximity measure is called a linkage metric. Since each node of the graph corresponds to a cluster, and since the original set of points is came from a metric space, we can define a new metric

[§]Quite interesting graph visualization software avaliable at [37].
[¶]E.g. red being high and blue being low.

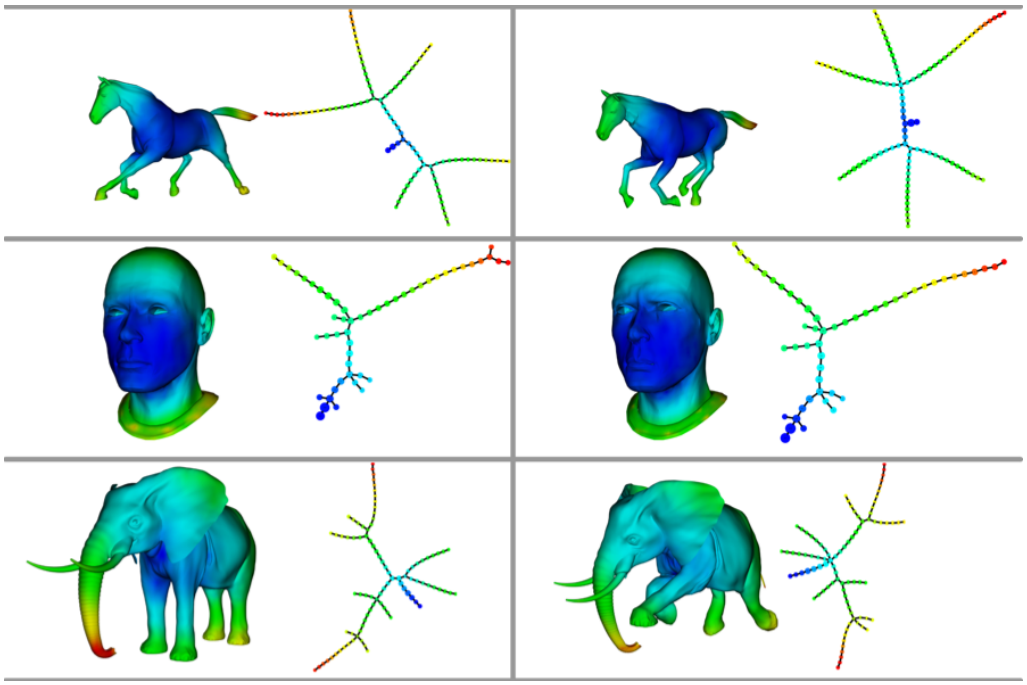


Fig. C.3: Similarity graph representations of the same objects at two different moments of time.

space, $(\{ G_1, G_2, \ldots, G_n \}, D_{\mathfrak{G}})$, on the graph by computing distances between clusters. Let say that the vertices of the graph are G_i, each G_i corresponds to a cluster Q_{α_i} with the cardinality $\mathsf{card}\, Q_{\alpha_i}$, and let us to define the metric as

$$D_H(G_i, G_j) \stackrel{\text{def}}{=} \left\{ \max \left(\frac{\sum_y \min_x d(x,y)}{\mathsf{card}\, Q_{\alpha_j}}, \frac{\sum_x \min_y d(x,y)}{\mathsf{card}\, Q_{\alpha_i}} \right) \mid x \in Q_{\alpha_i},\ y \in Q_{\alpha_j} \right\}.$$

This dissimilarity measure is defined in terms of pairs of nodes, one in each respective cluster. The measure calculates inter-cluster distances and naturally related to the similarity graph. Here every data partition directly corresponds to a graph partition.

This intrinsic graph metric is an alternative choice to the Euclidean metric as, in some situations, it represents the intrinsic geometry of the data much better.

Appendix D

Brief Description of the Project

The theme of my PhD work is "Topological Methods for the Representation and Analysis of Exploration Data in Oil Industry". The main purpose of the research is to apply algebraic topology methods to a description of shapes of undegraund capacities where oil/gas gathering. The motivation of the research is clear: since to drill one oil well is extremely expensive, it is crucial point to have a view how the capacity roughly look like. This information is as important as information about oil fields location.

Briefly, the idea behind this is as follows. Imagine a volume of oil and gas in some reservoir. This volume can be considered as an algebraic surface in three-dimensional space. The surface is embedded in the reservoir rock and captures also faults as well as impermeable layers in the reservoir rock, where as a result can be no oil or gas. In this context, these anomalies can be interpreted as holes of the algebraic surface. It is for establishing and counting of these holes the algorithms from computational algebraic topology with some auxiliary algorithms are implemented in the computer algebra system "Singular" in the scope of this project.

This kind of knowledge stipulate for a processing of huge amount of experimental exploration data which always contain a lot of noise and also has missing information. The data are obtained after a row of explosions and corresponds to times of arrival of post-explosion reflected waves to a network of special sensor detectors.

What was proposed in the work is a distillation of persistent topological features from the noisy changeable input data. Since Betti numbers enable us capture the three tipes of holes characterizing its connectivity (the gaps which separate components, the tunnels which pass through the shape, the voids which are

components of the complement space inaccessible from the outside), the persistent Betti numbers are significant geometric information about the input point cloud and, via it, about the underground geological formation. The Betti numbers can be represented as their geometrical persistence analog, so called barcodes. Of course, this calculations stipulate a condition of a construction of some simplicial complex which approximate the data in topological meaning.

From the point of view of "exploration", i.e. finding oil/gas reservoirs, this approach is new and, therefore, will be assessed against traditional techniques.

References

[1] J. Abbott, A. Bigatti, M. Kreuzer, and L. Robbiano. Computing ideals of points. *Journal of Symbolic Computation*, 30(4):341 – 356, October 2000. [cited at p. 79]

[2] Biondo L. Biondi. *3D Seismic Imaging*. Number 14 in Investigations in Geophysics. Society if Exploration Geophysicists, 2006. [cited at p. 5]

[3] E. Carlsson, G. Carlsson, and V. de Silva. An algebraic topological method for feature identification. *International Journal of Computational Geometry and Applications*, 16(4):291 – 314, 2006. [cited at p. 96]

[4] G. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255 – 308, April 2009. [cited at p. 17, 89]

[5] G. Carlsson, G. Singh, and A. Zomorodian. *Computing Multidimensional Persistence*, volume 5878/2009 of *Lecture Notes in Computer Science*. Springer, Berlin / Heidelberg, December 2009. pages 730 - 739, more complete version available on http://arxiv1.library.cornell.edu/abs/0907.2423. [cited at p. 89]

[6] G. Carlsson, A. Zomorodian, A. Collins, and L. Guibas. Persistence barcodes for shapes. *International Journal of Shape Modeling*, 11(2):149 – 187, 2005. [cited at p. 33]

[7] D. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms*. Undergraduate Texts in Mathematics. Springer, New York, third edition, 2007. [cited at p. 11, 38, 40, 79]

[8] V. de Silva. A weak definition of delaunay triangulation. *submitted*, October 16 2003. [cited at p. 18]

[9] V. de Silva and R. Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7:339 – 358, April 2007. DOI: 10.2140/agt.2007.7.339. [cited at p. 16]

[10] J.G. Dumas, F. Heckenbach, B.D. Saunders, and V. Welker. Computing simplicial homology based on efficient smith normal form algorithms. *Algebra, Geometry, and Software Systems*, pages 177 – 207, 2003. [cited at p. 69]

[11] H. Edelsbrunner. *Algorithms in Combinatorial Geometry*. Springer-Verlag, New York, 1987. [cited at p. 20]

[12] H. Edelsbrunner. The union of balls and its dual shape. *Discrete and Computational Geometry*, 13:415 – 440, 1995. [cited at p. 22]

[13] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete Computational Geometry*, 28:511 – 533, 2000. [cited at p. 29, 44, 49]

[14] H. Edelsbrunner and N. R. Shah. Triangulating topological spaces. *International Journal of Computational Geometry and Applications*, 7:365 – 378, 1997. [cited at p. 10, 23]

[15] D. Eisenbud. *Commutative Algebra with a View Toward Algebraic Geometry*. Graduate Texts in Mathematics. Springer, third printing edition, 1999. [cited at p. 41]

[16] M. Erwig. The graph voronoi diagram with applications. *Networks*, 36(3):156 – 163, 2000. [cited at p. 11]

[17] Claudia Fassino. An approximation of the gröbner basis of ideals of perturbed points, part i. Available at http://arxiv.org/abs/math/0703154v1, March 2007. [cited at p. 79]

[18] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, third edition, 2002. [cited at p. 2, 8, 10, 11, 26, 36, 98]

[19] D. Heldt, M. Kreuzer, S. Pokutta, and H. Poulisse. Approximate computation of zero-dimensional polynomial ideals. *Journal of Symbolic Computation*, 44(11):1566 – 1591, November 2009. In Memoriam Karin Gatermann. [cited at p. 79]

[20] Tomasz Kaczynski, Konstantin Mischaikow, and Marian Mrozek. *Computational Homology*, volume 157 of *Applied Mathematical Sciences*. Springer, 2004. [cited at p. 11]

[21] J. Kogan. *Introduction to Clustering Large and High-Dimensional Data*. Cambridge University Press, Cambridge, 2007. [cited at p. 94]

[22] J. Kogan, Ch. Nicholas, M. Teboulle, et al. *Grouping Multidimensional Data*. Springer-Verlag, Berlin Heidelberg, 2006. [cited at p. 94, 95]

[23] Kreuzer and L. Robbiano. *Computational Commutative Algebra 1*. Springer, Heidelberg, 2000. [cited at p. 79]

[24] Kreuzer and L. Robbiano. *Computational Commutative Algebra 2*. Springer, Heidelberg, 2005. [cited at p. 79]

[25] J. Leray. Sur la forme des espaces topologiques et sur les points fixes des représentations. *J. Math. Pures Appl.*, 24(9):95 – 167, 1945. [cited at p. 13, 21]

[26] A. T. Lundell and S. Weingram. *The Topology of CW Complexes*. Van Nostrand Reinhold Company, New York, 1969. [cited at p. 22]

[27] M. Maier, M. Hein, and U. von Luxburg. Optimal construction of k-nearest neighbor graphs for identifying noisy clusters. *Theoretical Computer Science*, 410:1749 – 1764, 2009. [cited at p. 100]

[28] J. R. Munkres. *Topology: A First Course*. Prentice Hall, Engiewood Cliffs, New Jersey, 1975. [cited at p. 11]

[29] J. R. Munkres. *Elements of Algebraic Topology*. AddisonWesley, Menlo Park, California, 1984. [cited at p. 11, 12]

[30] Atsuyuki Okabe, Barry Boots, Kokichi Sugihara, and Sung Nok Chiu. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley Series in Probability and Statistics. John Wdey & Sons Ltd, Chichester, second edition, 2000. [cited at p. 11]

[31] F. P. Preparata and M. I. Shamos. *Computational geometry: an introduction*. Springer-Verlag, New York, 1985. [cited at p. 12, 20]

[32] J. J. Rotman. *An Introduction to Algebraic Topology*, volume 119 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1988. [cited at p. 11]

[33] John A. Scales. *Theory of seismic imaging*. Springer-Verlag, Berlin, New York, 1995. [cited at p. 5]

[34] G. Segal. Classifying spaces and spectral sequences. *Publications Mathémathiques de l'Institut des Hautes Études Scientifiques*, 34:105 – 112, 1968. [cited at p. 98]

[35] G. Singh, F. Memoli, and G. Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. *Point Based Graphics*, September 2007. Prague. [cited at p. 92, 99]

[36] E. H. Spanier. *Algebraic Topology*. McGraw-Hill Book Co., 1966. [cited at p. 14]

[37] Graphviz. Graph visualization software. `http://www.graphviz.org`. [cited at p. 100]

[38] Open Source Clustering Software. `http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/index.html`. [cited at p. 95]

[39] Qhull. `http://www.qhull.org`. [cited at p. 47]

[40] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395 – 416, 2007. [cited at p. 92, 100]

[41] U. von Luxburg and S. Ben-David. Towards a statistical theory of clustering. In *PASCAL workshop on Statistics and Optimization of Clustering*, London, July 2005. [cited at p. 94]

[42] A. Zomorodian. *Topology for Computing*, volume 14 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, New York, 2005. [cited at p. 8]

[43] A. Zomorodian. *Computational Topology*, volume 2. Algorithms and Theory of Computation Handbook, second edition, 2009. Chapter 3. [cited at p. 49]

[44] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete Computational Geometry*, 33:249 – 274, 2005. [cited at p. 29, 40, 49, 52, 56]

[45] A. Zomorodian and G. Carlsson. Localized homology. *Computational Geometry: Theory and Applications*, 41:126 – 148, November 2008. [cited at p. 49, 99]

List of Symbols and Abbreviations

$\mathbb{X}$	*topological space*
Ξ	*algebraic surface of a geological formation*
$\mathfrak{A}_i$	*amplitude of the ith reflected signal*
X	*points cloud data sampled from* Ξ
$\mathcal{U}$	*covering (of* X*)*
$\mathcal{N}$	*nerve (of* $\mathcal{U}$*)*
$\mathcal{G}$	*geometric realization (of* $\mathcal{N}$*)*
$\sim$	*similarity (homologous) relation*
$\simeq$	*homotopy relation*
$\approx$	*homeomorphism relation*
$\cong$	*isomorphism relation*
$\pi(\mathbb{X})$	*fundamental group*
σ	k*-simplex*
$\mathcal{K}(X)$	*simplicial complex*
$\lvert\mathcal{K}\rvert$	*underlying space of* $\mathcal{K}$
$\check{\mathcal{C}}(X)$	*Čech complex*
ε	*some fixed parameter,* $\varepsilon > 0$
B_ε	*open ball with a radius* ε
$\mathcal{R}_\varepsilon(X)$	*Rips complex*
$\mathcal{L}$	*landmark points*
$\mathcal{W}(X, \mathcal{L}, \varepsilon)$	*strong witness complex*
$\bar{\mathcal{W}}(X, \mathcal{L}, \varepsilon)$	*weak witness complex*
$\mathcal{V}_p$	*Voronoi cell of* $p \in X$
$\mathcal{V}_X$	*Voronoi diagram*
$\mathcal{V}_\Xi$	*Voronoi diagram restricted to* Ξ
$\mathcal{D}_X$	*Delaunay complex (triangulation)*
$\mathcal{D}_\Xi$	*Delaunay triangulation restricted to* Ξ
$\aleph_\varepsilon(X)$	α*-shape complex*
$\mathcal{C}(X)$	*cluster complex*
C_k	*group of* k*-chains*
∂_k	k*-dimensional boundary operator*

Z_k	k*th chain (cycle) group*
B_k	k*th boundary group*
H_k	k*th homology group*
β_k	k*th Betti number*
Z_k^j	k*-th cycle group*
B_k^i	k*th boundary group*
$H_k^{j,p}$	p*-persistent* k*th homology group*
$\beta_k^{i,p}$	p*-persistent* k*th Betti numbers*
M_k	*standard matrix representation of* ∂_k
$\widetilde{M_k}$	*(Smith) normal form of* M_k
$\mathcal{M}$	*persistence module*
$\Gamma(\mathcal{M})$	*graded module*
Σ^α	*shift upward in grading by* α
D	*graded* PID
$\mathfrak{b}$	*barycenter*
$\mathfrak{M}$	*Mayer-Vietoris blowup (complex)*
$\mathfrak{G}$	*graph*
R	*ring*
$\mathfrak{F}$	*field*
PID	*principal ideal domain*
gcd	*greatest common devisor*

List of Figures

List of Tables

Index

Scientific Career

1999	*Diploma, St. Petersburg State University*
1999 -- 2001	*Teacher, Company "Informatization of Education"*
2001 -- 2004	*Teacher, Electrical Engineering Saint-Petersburg State University*
2005	*Research Fellow, A.N. Krylov Research Institute*
2006	*Research Assistant, Hong Kong University of Science and Technology*
2008 -- 2009	*Research Mathematician, EP-Research, Royal Dutch Shell, Rijswijk, The Netherlands*
2006 -- 2010	*Ph.D. Candidate in Mathematics, Technische Universität Kaiserslautern*