MASTERARBEIT | MASTER'S THESIS

Titel | Title

Computational Experiments On Random Chromatic Persistent Homology

verfasst von | submitted by

Sophie Larissa Rosenmeier BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 821
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Mathematik
Betreut von | Supervisor:	Dipl.-Ing. Dr. techn. Herbert Edelsbrunner

Abstract

This thesis explores the six persistence diagrams that arise out of persistent homology for chromatic point clouds. Concretely, for each diagram the expected number and length of their persistence pairs is being studied under the condition of uniformly distributed 2-colored points in the unit square. In a subset of cases we find explicit formulae for these very expectations for large enough numbers of points as the probability distribution of the colors varies.
During the creation of this work, deeper connections to the research area of Euclidean minimum spanning trees were discovered as well, in the sense that the constant giving the asymptotic limit of the (expected) length of the Euclidean minimum spanning tree also presents itself in the computational experiments for the 1-norm in the diagrams.

Abriss

Diese Masterarbeit erforscht die sechs Persistenz-Diagramme, die aus der persistierenden Homologie für gefärbte Punktwolken hervorgehen. Konkret wird für jedes Diagramm die erwartete Anzahl und Länge der Persistenz-Paare unter der Bedingung von gleichmäßig verteilten, zweigefärbten Punkte im Einheitsquadrat untersucht. In einem Teil der Fälle finden wir explizite Formeln für ebendiese Erwartungswerte bei genügend vielen Punkten, während die Wahrscheinlichkeitsverteilung der zwei Farben variiert. Bei der Entstehung dieser Arbeit wurden zudem tiefere Verbindungen zum Forschungsgebiet rund um Euklidische minimale Spannbäume festgestellt.
Dies ist in dem Sinne so zu verstehen, dass die Konstante, die den asymptotischen Grenzwert der (erwarteten) Länge des Euklidischen minimalen Spannbaums bildet, sich auch in den rechnerischen Experimenten zur erwarteten 1-Norm in den Diagrammen zeigt.

Acknowledgements

First of all, I would like to thank my supervisor, Herbert Edelsbrunner, for giving a talk at my university two years ago, which allowed me to see the powerful technique of persistent homology for the first time and sparked my interest in applied algebraic topology and topological data analysis. Thank you for enabling me an internship in your research group, where I was able to find the topic of my thesis, and thank you – alongside Yasuaki Hiraoka – for inviting me to Kyoto, where we had particularly fruitful discussions and first successes in theoretically proving my computational results.

A big thank you further goes out to Sebastiano Cultrera di Montesano, Ondrej Draganov and Morteza Saghafian, who also helped me understanding the theory behind chromatic Alpha complexes and in interpreting the amounts of data I generated. Especially, I want to highlight the enormous support of Ondrej Draganov regarding the computational part of this thesis as well as conversations about proof ideas. Thank you for always answering my countless little questions about code snippets or visualizations.

Last but not least, I want to thank my mom and my grandma for their endless support and their faith in me. Thank you for letting me pursue my dream of studying mathematics and for always standing by my side no matter what.

This thesis was supported by the University of Vienna as part of a short-term grant abroad.

Contents

1. Introduction

We live in a digitally connected and constantly changing world and generate massive amounts of data on a daily basis. To be able to evaluate and analyze these piles of information for various kinds of applications, we need tools that allow us to do so. The still relatively new field of *Topological Data Analysis* (TDA) has gifted us with exactly such theoretical instruments to investigate different data structures. Concretely, the idea of TDA is based on the assumption that the way data sets are shaped contains information itself, which helps with classification. But to get to this information with a computer is a non-trivial task.
Indeed, consider a small 2-dimensional data set like the one depicted in Figure 1.1. Clearly, the human eye instantly recognizes two clusters in this data set, one of which forming an ellipsoidal loop. For a computer, however, this point cloud only consists of 100 individual discrete objects.

Figure 1.1.: Simple example of a 2-dimensional data set that features two outstanding clusters with different structures.

In order to computationally gain information about such structures, we make use of a famous technique in TDA, that is *persistent homology*. With persistent homology we can extract topological features such as holes, cavities or clusters. Essentially, the idea is to grow balls around the points and to use the radius as a parameter to control the growing. As the radius increases, balls may overlap and the union of these results in a new topological space. It is exactly this change that is recorded by persistent homology.

Historically, persistent homology originally emerged in the 1990's, first touched upon in Frosini's work [20] where he introduced size functions. In 1999, Robins studied the topology of attractors in dynamical systems and introduced the concept of persistent Betti numbers [25]. Independently, Edelsbrunner, Letscher and Zomorodian established the framework of persistent homology in combination with a fast algorithm and a corresponding visualization [14]. This achievement was grounded on the previous work of Delfinado and Edelsbrunner on developing an incremental algorithm for Betti numbers [10] as well as a computational implementation of Alpha shapes, which was done in [18] by Edelsbrunner and Mücke. In 2005, Carlsson and Zomorodian gave another viewpoint on persistent homology in algebraic terms [31].

Since then, a lot has been done and persistent homology has found great resonance, in particular, in the fields of material and life sciences [2], [17], [26]. Especially, the steadily growing area of spatial biology seems to give another opportunity to make usage of persistent homology, see for instance [7]. Motivated by this new field, Cultrera et al developed a theory of persistent homology for *chromatic* point sets [5], [8], where each point can be thought of as a cell with a color indicating its type. Out of this process, six different persistence diagrams, the *six-pack*, emerged instead of only one. This thesis explores the expected 0- and 1-norm of these very diagrams with computational experiments.

In chapter 2, we first give a detailed overview on the usual persistent homology for non-chromatic point sets and afterwards, introduce all aforementioned concepts in the colored setting. In the end of this theoretical chapter, we briefly treat Poisson point processes, which is an essential ingredient for our analysis, and mention Euclidean minimum spanning trees and the asymptotic limit of their (expected) length. Chapter 3 serves as a quick guide through the setup of the computational experiments and its implementation. The heart of this thesis lies in chapter 4, where some explicit formulae are theoretically proven and a tabular summary is presented, including further conjectures and still completely unknown cases regarding the expectation. In the appendix the underlying empirical results, that initially sparked the deeper investigation of the already mentioned formulae, can be looked up in a case-by-case fashion. Finally, chapter 5 discusses further open questions and interesting continuations of this thesis.

2. Mathematical Background

This chapter is devoted to introducing all mathematical concepts that are needed for understanding the computational setup in chapter 3 as well as the results in the subsequent chapter 4. In section 2.1 we begin with defining the notions of geometric and abstract simplicial complexes and how they are related to each other. In section 2.2 we encounter wildly known examples of simplicial complexes.
After a compact summary of the most essential definitions and constructions in simplicial homology[1] in section 2.3, we turn to the useful tool of persistent homology in section 2.4. There, we take the geometrical objects introduced in the first two sections, construct filtrations of them and apply homology theory to be able to say something about topological features like connected components or loops. In sections 2.5 and 2.6 everything about Alpha complexes and persistent homology is translated to the chromatic setting, which is of great importance for the computational setup as we deal with colored point clouds. Finally, section 2.7 serves as a quick guide in the topic of Poisson point processes.

2.1. Geometric and Abstract Simplicial Complexes

In this section we encounter the building blocks for defining Alpha complexes and follow the contents of [16] and [24].

Definition 2.1.1. Let $p_i \in \mathbb{R}^d$ $(0 \leq i \leq k)$. Then the set $\{p_0, \dots, p_k\}$ of $k+1$ points is called *affinely independent* if for all $\lambda_i \in \mathbb{R}$ we get

$$\sum_{i=0}^{k} \lambda_i = 0 \quad \text{and} \quad \sum_{i=0}^{k} \lambda_i p_i = 0 \quad \Longrightarrow \quad \lambda_0 = \cdots = \lambda_k = 0.$$

Remark 2.1.2. The above implication is equivalent to requiring that the k vectors $p_i - p_0$ $(1 \leq i \leq k)$ are linearly independent in the usual sense. One consequence of this observation is that there can be at most $d+1$ affinely independent points in $\mathbb{R}^d$.

Definition 2.1.3. The *convex hull* of a point set $\{p_0, \dots, p_k\} \subseteq \mathbb{R}^d$ is defined as the set of convex combinations

$$\mathrm{conv}\{p_0, \dots, p_k\} = \left\{ \sum_{i=0}^{k} \lambda_i p_i \,\middle|\, \sum_{i=0}^{k} \lambda_i = 1,\ \lambda_i \geq 0 \right\}.$$

[1]First, in a more general way, for coefficients in $\mathbb{Z}$ and later on for coefficients in $\mathbb{Z}_2$.

Definition 2.1.4. A (*geometric*) *k-simplex* (or just *simplex*) σ is the convex hull

$$\sigma = \mathrm{conv}\{p_0, \dots, p_k\}$$

of some affinely independent set $\{p_0, \dots, p_k\} \subseteq \mathbb{R}^d$. We also say that σ is spanned by the points $p_0, \dots, p_k$. The *dimension of* σ is $\dim(\sigma) = k$.

In particular, for $0 \leq k \leq 3$, k-simplices have their own names: A 0-simplex is called a *vertex*, a 1-simplex is an *edge*, a 2-simplex is a *triangle* and, finally, a 3-simplex is called a *tetrahedron*. Figure 2.1 shows these very simplices.

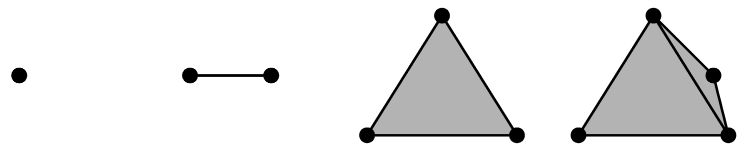

Figure 2.1.: *From left to right:* vertex, edge, triangle and tetrahedron.

Definition 2.1.5. Let $\sigma = \mathrm{conv}\{p_0, \dots, p_k\}$ be a k-simplex in $\mathbb{R}^d$. A *face* τ of σ is the convex hull of a non-empty subset of $\{p_0, \dots, p_k\}$ and it is *proper* if, additionally, the subset is not all of $\{p_0, \dots, p_k\}$.
We write $\tau \leq \sigma$ for τ being a face of σ and $\tau < \sigma$ if it is a proper one.

Remark 2.1.6. (1) Every k-simplex σ in $\mathbb{R}^d$ is compact and convex with respect to the Euclidean topology. Moreover, $\dim(\sigma) = k \leq d$.
(2) It is easily checked that any face τ of a k-simplex σ is again a simplex of dimension of $\dim(\tau) \leq k$.
(3) A k-simplex σ has $2^{k+1} - 1$ faces in total, all of which are proper ones apart from σ itself.

Definition 2.1.7. Let σ be a k-simplex in $\mathbb{R}^d$. The *boundary* of σ is the union of all its proper faces and we denote it by $\mathrm{bnd}(\sigma)$.
Moreover, we define the *interior* of σ to be $\mathrm{int}(\sigma) := \sigma \setminus \mathrm{bnd}(\sigma)$.

Example 2.1.8. Consider a triangle σ in $\mathbb{R}^2$. Its boundary is then given by the union of the points it is spanned by and all edges connecting these points, whereas its interior is precisely what lies inside the triangle.

Lemma 2.1.9. *Let σ be a k-simplex in $\mathbb{R}^d$ spanned by $p_0, \dots, p_k \in \mathbb{R}^d$ and let $x \in \sigma$ be of the form $x = \sum_{i=0}^k \lambda_i p_i$ $(\lambda_i \geq 0)$. Then the following are equivalent:*

(i) x lies in the interior of σ, so $x \in int(\sigma)$.

(ii) $\lambda_i > 0 \quad \forall i \in \{0, \dots, k\}$.

Proof. $(i) \Rightarrow (ii)$. This implication is done by contraposition. Without loss of generality, let $\lambda_k = 0$. Then $x \in \mathrm{conv}\{p_0, \dots, p_{k-1}\}$ since $\sum_{i=0}^{k-1} \lambda_i = 1$. This means that x lies in

the proper face $\operatorname{conv}\{p_0, \ldots, p_{k-1}\}$ of σ, thus $x \in \operatorname{bnd}(\sigma)$.
$(ii) \Rightarrow (i)$. Assume that each coefficient λ_i is positive. Since $\operatorname{bnd}(\sigma)$ consists of all proper faces of σ, x cannot lie in any of them as they are convex hulls of proper subsets of $\{p_0, \ldots, p_k\}$. Hence, $x \in \operatorname{int}(\sigma)$. □

After introducing simplices we can now use them as our smallest building blocks to define more sophisticated structures following a specific recipe.

Definition 2.1.10. A (*geometric*) *simplicial complex* K in $\mathbb{R}^d$ is a finite collection of simplices that obey the following two rules:

(a) $(\sigma \in K$ and $\tau \leq \sigma) \Rightarrow \tau \in K$.

(b) $\sigma_1, \sigma_2 \in K \Rightarrow (\sigma_1 \cap \sigma_2 = \emptyset$ or $\sigma_1 \cap \sigma_2 \leq \sigma_j$ for both $j = 1$ and $j = 2)$.

In other words, (a) requires the closedness of K under taking faces and (b) states that there must not be improper intersections.

Definition 2.1.11. Let K be a simplicial complex in $\mathbb{R}^d$. The *dimension of* K is given by

$$\dim(K) = \max_{\sigma \in K} \{\dim(\sigma)\} \leq d.$$

The *underlying space* $|K|$ of K is the union of all of its simplices viewed as a topological space lying in $\mathbb{R}^d$ and equipped with the trace topology.

Lemma 2.1.12. *Given a simplicial complex K, then its underlying space $|K| \subseteq \mathbb{R}^d$ is compact with respect to the Euclidean topology.*

Proof. Boundedness is clear since K is a finite collection of simplices that are compact by remark 2.1.6/(1). For closedness of $|K|$, first observe that

$$|K| = \bigcup_{\sigma \in K} \sigma.$$

Since any finite union of closed sets is closed again, the claim follows. □

Example 2.1.13. Consider Figure 2.2 and imagine the drawings to lie in $\mathbb{R}^3$. In the first collection of simplices the rightmost vertex is missing, which clearly violates condition (a) from definition 2.1.10. Therefore, it cannot be a simplicial complex. Also, the particular formation with two hollow tetrahedra in the middle of figure 2.2 does not show a simplicial complex, because their intersection gives a proper face for one of them but not for both. However, the third drawn example shows a simplicial complex of dimension 3.

Definition 2.1.14. Let K be a simplicial complex. Then $L \subseteq K$ is called a *subcomplex of* K if L is a simplicial complex itself.

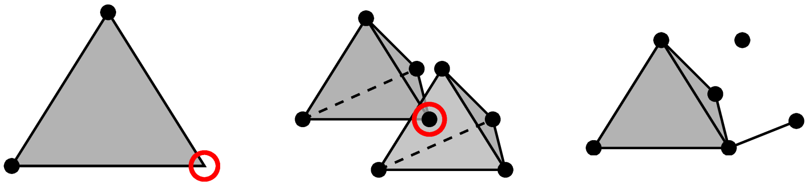

Figure 2.2.: The first two collections of simplices are no simplicial complexes, where the respective violation is highlighted by a red circle. The last one, however, is a simplicial complex.

Example 2.1.15. For any simplicial complex K we can always construct its *j-skeletons*

$$K^{(j)} = \{\sigma \in K \mid \dim(\sigma) \leq j\}, \quad 0 \leq j \leq \dim(K),$$

which are subcomplexes of K. Indeed, property (a) holds by construction and property (b) is inherited from K. In particular, the 0-skeleton $K^{(0)}$ is the set of vertices that are contained in K and we will refer to it as the *vertex set of* K.

So far, our definition of a simplicial complex crucially relies on the ambient Euclidean space $\mathbb{R}^d$ as there cannot be any such complex with underlying space $|K| \subseteq \mathbb{R}^d$ of dimension greater than d. To allow for more freedom, we are going to define a more abstract notion of a simplicial complex.

Definition 2.1.16. A finite collection of non-empty finite sets A is called an *abstract simplicial complex* if the property

$$(\alpha \in A \text{ and } \emptyset \neq \beta \subseteq \alpha) \Rightarrow \beta \in A$$

is fulfilled. An element $\alpha \in A$ is then named a *simplex* with *dimension*

$$\dim(\alpha) = \operatorname{card}(\alpha) - 1,$$

where $\operatorname{card}(\alpha)$ denotes the *cardinality* of the set α, and the *dimension of* A is defined to be

$$\dim(A) = \max_{\alpha \in A} \{\dim(\alpha)\}.$$

Definition 2.1.17. Let A be an abstract simplicial complex and let $\alpha \in A$ be some simplex. A *face of* α is a non-empty subset $\beta \subseteq \alpha$ and it is called *proper* if, additionally, $\beta \neq \alpha$. Moreover, $B \subseteq A$ is called a *subcomplex of* A if B itself is an abstract simplicial complex.

Example 2.1.18. Let $A = \big(\mathcal{P}(\{a_0, a_1, a_2\}) \cup \mathcal{P}(\{a_0, a_3\}) \cup \mathcal{P}(\{a_4\})\big) \setminus \{\emptyset\}$, where every a_i $(0 \leq i \leq 4)$ is viewed as some abstract mathematical object. It is easily seen that A defines an abstract simplicial complex. Simplices of A with different dimensions are, for example, $\{a_0, a_1, a_2\}, \{a_1, a_2\}$ or $\{a_4\}$. Setting $B = \big(\mathcal{P}(\{a_0, a_3\}) \cup \mathcal{P}(\{a_2\})\big) \setminus \{\emptyset\}$ gives us a subcomplex of A.

Intuitively, we can regard a simplex of an abstract simplicial complex as a collection of points that will, later on, represent the vertices of a geometric simplex of a geometric simplicial complex in some well enough chosen ambient Euclidean space. It is, therefore, not surprising to come up with the following definition.

Definition 2.1.19. Let A be an abstract simplicial complex. The *vertex set of* A, denoted by $\mathrm{Vert}(A)$, is the union of all simplices of A, so

$$\mathrm{Vert}(A) = \bigcup_{\alpha \in A} \alpha.$$

Example 2.1.20. Considering A and B to be as in the previous example 2.1.18, we obtain $\mathrm{Vert}(A) = \{a_0, a_1, a_2, a_3, a_4\}$ as well as $\mathrm{Vert}(B) = \{a_0, a_2, a_3\}$.

Up until now, we have introduced two seemingly different notions of simplicial complexes – geometric and abstract ones. In the upcoming lemma we will already see some connection between these constructions as every geometric simplicial complex can be turned into an abstract one by discarding all of its simplices apart from the 0-dimensional ones.

Lemma 2.1.21. *Let K be a geometric simplicial complex of dimension* $\dim(K) = k$ *in* $\mathbb{R}^d$*, then the set*

$$A = \{\{p_{m_0}, \dots, p_{m_i}\} \mid \operatorname{conv}\{p_{m_0}, \dots, p_{m_i}\} \in K,\, 0 \le m_0 < \dots < m_i \le k\}$$

is an abstract simplicial complex and we call it a vertex scheme *of K.*

Proof. This is obviously the case, since K is a geometric simplicial complex. □

A natural question to ask is whether some converse statement can be formulated as well. To be more specifc, is it possible to transfer an abstract simplicial complex A with dimension $\dim(A) = n$ into a suitable ambient space $\mathbb{R}^d$ such that one can construct a geometric simplicial complex K out of it with vertex set $K^{(0)} = \{\{a\} \mid a \in \mathrm{Vert}(A)\}$?
It turns out, such a *geometric realization of* A does, indeed, exist for $d = 2n + 1$ and in the following we sometimes denote it by $\underline{A}$. However, in order to examine this claim, we need one more definition.

Definition 2.1.22. Let $p_i \in \mathbb{R}^d$ $(0 \le i \le k)$. We say that $\{p_0, \dots, p_k\}$ is in *general position* if no $(d+2)$-many points of this set lie on a common $(d-1)$-sphere.

Remark 2.1.23. Considering definition 2.1.22 we observe that for $k \le d$ the set $\{p_0, \dots, p_k\}$ is always in general position.

Theorem 2.1.24 (Geometric Realization Theorem)**.** *Every abstract simplicial complex A with dimension* $\dim(A) = n < \infty$ *has a geometric realization* $\underline{A}$ *in* $\mathbb{R}^{2n+1}$.

Proof. First notice that $\mathrm{Vert}(A) = \{a_i\}_{i \in I}$ where I is a finite index set. Now construct an injection $f : \mathrm{Vert}(A) \to \mathbb{R}^{2n+1}$ whose image is a set of points in general position. Observe that for any simplex $\alpha \in A$ we have $\alpha \subseteq \mathrm{Vert}(A)$, hence the notion $f(\alpha)$ is valid. We claim that

$$K = \{\mathrm{conv}\{f(\alpha)\} \mid \alpha \in A\}$$

is a geometric simplicial complex with vertex set

$$K^{(0)} = \{\mathrm{conv}\{f(\alpha)\} \mid \alpha \in A, \dim(\alpha) = 0\} = \{\{f(a_i)\} \mid a_i \in \mathrm{Vert}(A)\}.$$

For this to show let $\alpha_1, \alpha_2 \in A$, then by assumption $\dim(\alpha_1) \leq n$ and $\dim(\alpha_2) \leq n$. We have

$$\mathrm{card}(\alpha_1 \cup \alpha_2) = \mathrm{card}(\alpha_1) + \mathrm{card}(\alpha_2) - \mathrm{card}(\alpha_1 \cap \alpha_2) \leq 2n + 2.$$

Consequently, all points in $f(\alpha_1 \cup \alpha_2)$ are affinely independent, which implies that we can form their convex hull and that every $x \in \mathrm{conv}\{f(\alpha_1 \cup \alpha_2)\}$ has unique coefficients. Now, let $\sigma_1, \sigma_2 \in K$ be of the form

$$\sigma_1 = \mathrm{conv}\{f(\alpha_1)\},\ \sigma_2 = \mathrm{conv}\{f(\alpha_2)\}$$

and suppose $\sigma_1 \cap \sigma_2 \neq \emptyset$. Then

$$x \in \sigma_1 \cap \sigma_2 \Leftrightarrow x \in \mathrm{conv}\{f(\alpha_1 \cap \alpha_2)\} =: \tau,$$

where τ is a face of both σ_1 and σ_2. This proves condition (b). For condition (a), take $\sigma = \mathrm{conv}\{f(\alpha)\} \in K$, then any face of σ has to be of the form $\mathrm{conv}\{f(\beta)\}$ where $\emptyset \neq \beta \subseteq \alpha$ is a face of α and the claim follows. Therefore, $K = \underline{A}$ is the desired geometric realization of A. □

Because of the Geometric Realization Theorem in most cases it is negligible whether to have a geometric or an abstract simplicial complex in hand. We, thus, sometimes simply speak about *simplicial complexes*, unless we explicitly work with one or the other definition.

In the introductory chapter 1 we already touched upon the superficial concept behind persistent homology. We briefly mentioned growing balls and how this idea eventually changes the homology when taking the union and increasing the radius. Certainly, we will go through this process step-by-step in the following sections. However, an important theoretical result is needed for all of the following to work out. This is what we want to talk about now.

Definition 2.1.25. Let S be a finite collection of sets. We define the *nerve of S*, denoted by $\mathrm{Nrv}(S)$, to be the set of all non-empty subcollections of S whose elements have non-empty common intersection. More compactly written, we have

$$\mathrm{Nrv}(S) = \{T \subseteq S \mid T \neq \emptyset, \bigcap_{X \in T} X \neq \emptyset\}.$$

Remark 2.1.26. The nerve of S is an abstract simplicial complex since for $T \in \operatorname{Nrv}(S)$ and $\emptyset \neq T' \subseteq T$ we obtain

$$\bigcap_{X \in T'} X \supseteq \bigcap_{X \in T} X \neq \emptyset,$$

hence $T' \in \operatorname{Nrv}(S)$.

Example 2.1.27. (1) Take $S = \{X_0, X_1, X_2\}$ to be the set consisting of three overlapping closed disks in $\mathbb{R}^2$ as depicted in the top left of figure 2.3. Then we get

$$\operatorname{Nrv}(S) = \{T_0, T_1, T_2, T_{01}, T_{02}, T_{12}, T_{012}\},$$

where

$$\begin{aligned} T_0 &= \{X_0\}, \quad T_1 = \{X_1\}, \quad T_2 = \{X_2\}, \\ T_{01} &= \{X_0, X_1\}, \quad T_{02} = \{X_0, X_2\}, \quad T_{12} = \{X_1, X_2\}, \\ T_{012} &= \{X_0, X_1, X_2\}. \end{aligned}$$

Since $\operatorname{Nrv}(S)$ is an abstract simplicial complex, we can use an injection f, like in the proof of theorem 2.1.24, to get its geometric realization

$$\underline{\operatorname{Nrv}(S)} = \{\operatorname{conv}\{f(T_i)\} \mid i \in I\}$$

with index set $I = \{0, 1, 2, 01, 02, 12, 012\}$. More concretely, for $i \in \{0, 1, 2\}$, this gives us the vertices $p_i = \operatorname{conv}\{f(T_i)\} = f(T_i)$, then come the edges for $i \in \{01, 02, 12\}$ and, lastly, $i = 012$ yields a triangle that is spanned by p_0, p_1 and p_2.

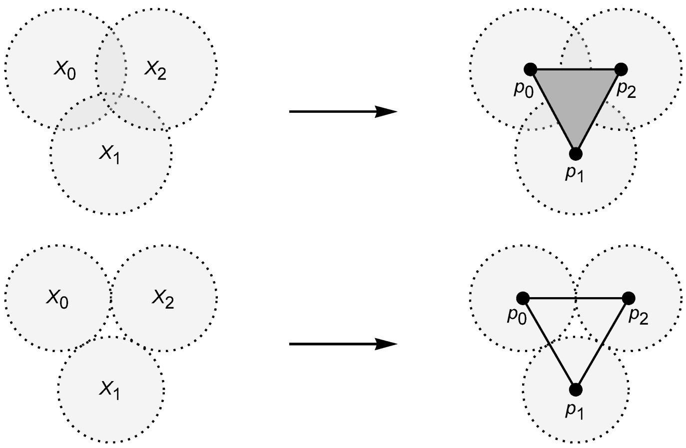

Figure 2.3.: *Top:* Three disks with common intersection, resulting in a solid triangle as nerve. *Bottom:* Three disks with merely pairwise intersection, resulting in an empty triangle as nerve.

(2) Now take $S = \{X_0, X_1, X_2\}$ to be the set consisting of three pairwise overlapping cloesd disks in $\mathbb{R}^2$ as shown in the bottom left of figure 2.3. We again compute the nerve

and obtain

$$\mathrm{Nrv}(S) = \{T_0, T_1, T_2, T_{01}, T_{02}, T_{12}\}$$

with

$$\begin{aligned} T_0 = \{X_0\}, \quad T_1 = \{X_1\}, \quad T_2 = \{X_2\}, \\ T_{01} = \{X_0, X_1\}, \quad T_{02} = \{X_0, X_2\}, \quad T_{12} = \{X_1, X_2\}. \end{aligned}$$

We see that there is no set $T_{012} = \{X_0, X_1, X_2\}$ in the nerve of S anymore, since here the intersection of all disks is empty. When considering the geometric realization $\underline{\mathrm{Nrv}(S)}$ of $\mathrm{Nrv}(S)$, this observation results in a, loosely speaking, unfilled triangle.

From the above example we get the feeling that somehow the union of the elements in S is in some correspondence to the geometric realization of the nerve of S as both sets look alike from the perspective of a topologist. Recall that in the language of topology, this similarity corresponds to being homotopy equivalent.

Definition 2.1.28. Let X, Y be topological spaces, then two continuous functions $f, g : X \to Y$ are called *homotopic* if there exist continuous maps $H_t(x) : X \to Y$, $t \in [0, 1]$, such that $H_0 = f$, $H_1 = g$ and such that the map

$$H : X \times [0, 1] \to Y, \quad H(x, t) = H_t(x)$$

is continuous. Then H is called a *homotopy joining f and g* and we write $f \simeq_H g$ or just $f \simeq g$.

Definition 2.1.29. The topological spaces X, Y are said to be *homotopy equivalent* if there exist continuous maps $f : X \to Y$ and $g : Y \to X$ such that

$$g \circ f \simeq \mathrm{id}_X, \quad f \circ g \simeq \mathrm{id}_Y$$

and we then write $X \simeq Y$.

In fact, we can generalize the insights we gained from example 2.1.27 and arrive at the powerful Nerve Theorem.

Theorem 2.1.30 (Nerve Theorem)**.** *Let $S = \{X_i\}_{i \in I}$ be a finite collection of closed, convex sets in $\mathbb{R}^d$. Then*

$$\underline{\mathrm{Nrv}(S)} \simeq \bigcup_{i \in I} X_i.$$

There are many more versions of the nerve theorem, some of which are even less restrictive with the assumptions on the given sets. A hodgepodge on proofs for the nerve theorem can be found in [4].

2.2. Delaunay and Alpha Complexes

Now that we understand the nerve of a finite family of sets S, let us restrict ourselves again to the special case of closed balls. We orientate ourselves on [8] and [16].

Definition 2.2.1. Given a finite set of closed balls $B_c(r)$ in $\mathbb{R}^d$ with center at $c \in \mathbb{R}^d$ and radius $r \geq 0$ and let C be the collection of all the centers. The *Čech complex of C and r* is then

$$\check{\mathrm{C}}\mathrm{ech}(r, C) = \{\alpha \subseteq C \mid \bigcap_{c \in \alpha} B_c(r) \neq \emptyset\}.$$

Remark 2.2.2. The Čech complex can be regarded as the nerve of $S = \{B_c(r) \mid c \in C\}$ but with elements being subsets $\alpha \subseteq C$ of the centers of the balls instead of subsets $T \subseteq S$ of the balls themselves. Hence,

$$\check{\mathrm{C}}\mathrm{ech}(r, C) \cong \mathrm{Nrv}(S)$$

and $\check{\mathrm{C}}\mathrm{ech}(r, C)$ is an abstract simplicial complex. Also, we have the nice property that

$$\check{\mathrm{C}}\mathrm{ech}(r, C) \subseteq \check{\mathrm{C}}\mathrm{ech}(R, C) \text{ for } r \leq R,$$

which is visualized in Figure 2.4.

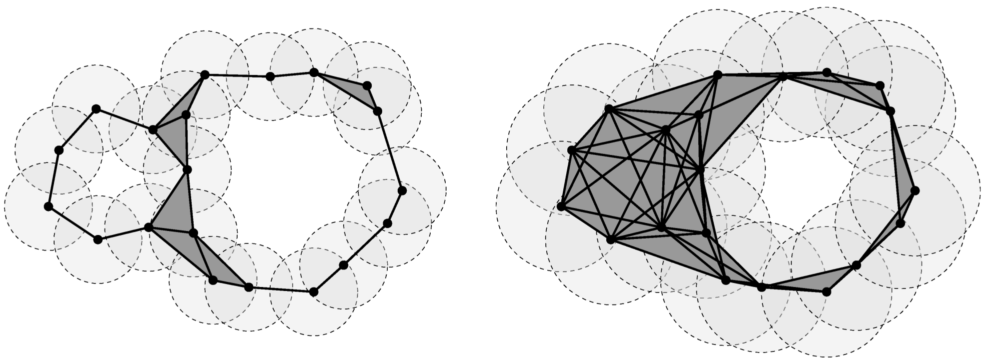

Figure 2.4.: Two Čech complexes of the same point set but with different radii. As the balls are growing, new simplices appear.

The Čech complex is a very simple but powerful discrete object for monitoring the union of growing balls. The important connection to the nerve and the fact that balls are always convex allow us to apply the Nerve Theorem 2.1.30 for each $r \geq 0$, which yields

$$\underline{\check{\mathrm{C}}\mathrm{ech}(r, C)} \simeq \bigcup_{c \in C} B_c(r). \tag{2.1}$$

The downside of the Čech complex is that algorithms for computing $\check{\mathrm{C}}\mathrm{ech}(r, C)$ are rather time consuming as the radius r or the number of points in C increase. Imagining a point cloud with hundreds or thousands of points, this soon gets us into trouble as there are too many simplices forming we have to keep track of. An alternative abstract simplicial complex that is, in some sense, a simplification of the Čech complex is given by the Vietoris-Rips complex.

Definition 2.2.3. Given a finite set of closed balls $B_x(r)$ in $\mathbb{R}^d$ with center at $x \in \mathbb{R}^d$ and radius $r \geq 0$ and let C be the collection of all the centers. The *Vietoris-Rips complex of C and r* is

$$\mathrm{VR}(r, C) = \{\alpha \subseteq C \mid \mathrm{diam}(\alpha) \leq 2r\},$$

where we denote by $\mathrm{diam}(\alpha)$ the maximal Euclidean distance of two points in α.

A major disadvantage of this construction is, however, that, in general, we do not get an analogous homotopy equivalence as in (2.1). The following simple example shows this intuitively.

Example 2.2.4. Consider again example 2.1.27/(2) and assume that all X_i $(i = 0, 1, 2)$ are unit disks, so $r = 1$. This implies that the centers x_i of these disks are exactly at distance $2r = 2$ from each other based on the fact that the pairwise overlapping of the disks here only happens in one point, respectively. Setting $C = \{x_0, x_1, x_2\}$, this gives us that $\check{\mathrm{C}}\mathrm{ech}(1, C)$ is exactly the unfilled triangle as before.
Speaking now of C instead of $T_{012} = \{X_0, X_1, X_2\} = S$, we get that $\mathrm{VR}(1, C)$ corresponds to the filled triangle since C also has $\mathrm{diam}(C) = 2r$.

Nevertheless, the Vietoris-Rips complex serves as a good approximation of the Čech complex and is computationally way better to use, especially for higher dimensions. [30]

We can construct yet another abstract simplicial complex, whose geometric realization is homotopy equivalent to the union of growing balls, as compared to the Vietoris-Rips complex, and which does not have that many simplices within its construction like the Čech complex does. But first, we need one more concept.

Definition 2.2.5. Let $C \subseteq \mathbb{R}^d$ be a finite set. The *Voronoi domain at $c \in C$* is given by

$$V(c, C) = \{x \in \mathbb{R}^d \mid \|x - c\|_2 \leq \|x - a\|_2\,, \forall a \in C\}.$$

This definition tells us that $V(c, C)$ consists of all points in $\mathbb{R}^d$ which are closest to c.

Example 2.2.6. In order to obtain a better feeling for Voronoi domains, let us go through an example. Take $C = \{a, b, c\}$ in the plane $\mathbb{R}^2$. We get $V(a, C), V(b, C), V(c, C)$ by considering the bisectors of a and b, b and c as well as a and c, respectively, and cutting them off at their first point of intersection.

Generalizing the construction in example 2.2.6 to higher dimensions yields that a Voronoi domain is the intersection of finitely many closed half-spaces and, thus, a closed convex polyhedron. Moreover, observe that the topological interiors of the Voronoi domains are pairwise disjoint and that the family of all Voronoi domains builds a covering of $\mathbb{R}^d$.

Both of these properties can be seen well again in Figure 2.5. Therefore, the following definition comes to no surprise.

Definition 2.2.7. Let $C \subseteq \mathbb{R}^d$ be a set of finitely many points, then the collection of Voronoi domains $\{V(c, C)\}_{c \in C}$ is called the *Voronoi tessellation* and we write $\mathrm{Vor}(C)$.

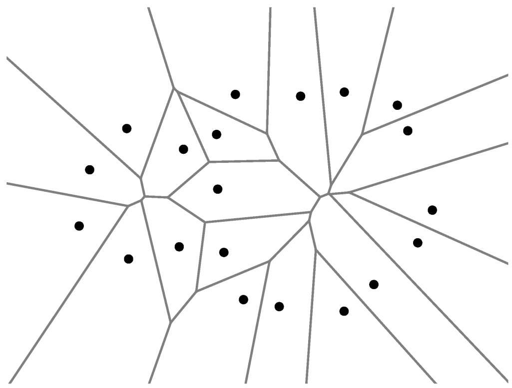

Figure 2.5.: The Voronoi domains of a point set. Together, all Voronoi domains form the Voronoi tessellation.

Having a finite family of closed, convex sets – in our case a Voronoi tessellation – we are able to build its nerve. This results in an abstract simplicial complex that is being very important throughout this thesis.

Definition 2.2.8. Given a finite set of points $C \subseteq \mathbb{R}^d$, define the *Delaunay complex of* C to be

$$\mathrm{Del}(C) = \{\alpha \subseteq C \mid \bigcap_{c \in \alpha} V(c, C) \neq \emptyset\}.$$

Remark 2.2.9. (1) If, additionally, C is in general position, then $\mathrm{Del}(C)$ has a geometric realization in $\mathbb{R}^d$ itself.
(2) As before with the Čech complex in remark 2.2.2, the definition of the Delaunay complex yields a bijection to the nerve of the Voronoi tessellation, thus

$$\mathrm{Del}(C) \cong \mathrm{Nrv}(\mathrm{Vor}(C)).$$

There is another way of viewing Delaunay complexes, which will help us to get the idea behind the concepts introduced in section 2.5.

Definition 2.2.10. Given a finite set of points $C \subseteq \mathbb{R}^d$. We say that a $(d-1)$-sphere in $\mathbb{R}^d$ is *empty* if every $c \in C$ either lies on or outside the sphere. If this is the case, the sphere *passes through* the points that lie on the sphere.

Lemma 2.2.11. *Let $C \subseteq \mathbb{R}^d$ be a finite set of points and let $\alpha \subseteq C$. The following are equivalent:*

(i) $\alpha \in \mathrm{Del}(C)$, *meaning that* α *is a simplex of the Delaunay complex.*

(ii) *There is an empty* $(d-1)$*-sphere that passes through every point in* α.

Proof. Observe that for $\alpha \in \mathrm{Del}(C)$ we get $\emptyset \neq \bigcap_{c\in\alpha} V(c,C) \ni x$. This is equivalent to saying that x has the same distance to all points in α and at least this distance to all other points in C. Taking x to be the center of a $(d-1)$-sphere, we obtain that this sphere is empty and it passes through all points in α. Since the last implication can be reversed as well, the claim $(i) \Leftrightarrow (ii)$ follows. $\square$

Unlike in the cases of Čech and Vietoris-Rips complexes, there is no varying parameter r in the Delaunay complex. However, we can establish one by modifying our Voronoi domains a little bit.

Definition 2.2.12. Given a finite set of points $C \subseteq \mathbb{R}^d$. For each $c \in C$, let $B_c(r)$ be the closed ball in $\mathbb{R}^d$ of center c and radius $r \geq 0$. We call $B_c^V(r,C) = B_c(r) \cap V(c,C)$ the *clipped Voronoi domain of* c *with radius* r and define the *Alpha complex of* C *and* r to be

$$\mathrm{Alpha}(r,C) = \{\alpha \subseteq C \mid \bigcap_{c\in\alpha} B_c^V(r,C) \neq \emptyset\}.$$

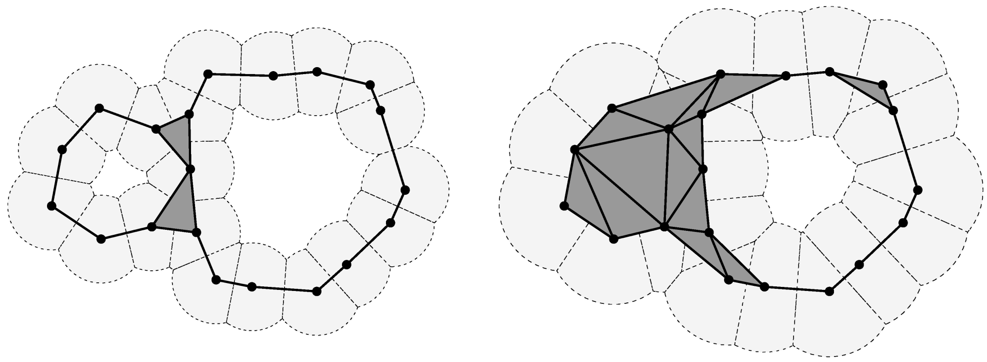

Figure 2.6.: Two Alpha complexes of the same point set but with different radii. As the clipped Voronoi balls are growing, new simplices appear.

Remark 2.2.13. (1) Clearly, all $B_c^V(r,C)$ $(c \in C)$ are closed and convex and by definition of the Alpha complex we get

$$\mathrm{Alpha}(r,C) \cong \mathrm{Nrv}(\{B_c^V(r,C)\}_{c\in C}).$$

In particular, the Alpha complex is an abstract simplicial complex.
(2) Since $B_c^V(r,C) \subseteq V(c,C)$, the Alpha complex is a subcomplex of the Delaunay complex and since $B_c^V(r,C) \subseteq B_c(r)$, the Alpha complex is a subcomplex of the Čech complex. This can be noticed by comparing Figure 2.4 with Figure 2.6. Moreover, we have

$$\mathrm{Alpha}(r,C) \subseteq \mathrm{Alpha}(R,C) \text{ for } r \leq R,$$

which is a property that is visualized in Figure 2.6 as well.
(3) Just as before, we get a geometric realization for $\operatorname{Alpha}(r, C)$ in $\mathbb{R}^d$ if C is in general position. By (1) it follows that

$$\underline{\operatorname{Alpha}(r,C)} \simeq \bigcup_{c\in C} B_c^V(r,C) = \left(\bigcup_{c\in C} B_c(r)\right) \cap \left(\bigcup_{c\in C} V(c,C)\right) = \bigcup_{c\in C} B_c(r), \qquad (2.2)$$

since the Voronoi tessellation covers $\mathbb{R}^d$. Notice that we obtain a homotopy equivalence between $\operatorname{Alpha}(r, C)$ and $\operatorname{Čech}(r, C)$ for free out of equations (2.1) and (2.2).
(4) For $r \leq R$ the diagram

$$\begin{tikzcd} \bigcup_{c\in C} B_c(r) \arrow[r, hook] \arrow[d, "\simeq"'] & \bigcup_{c\in C} B_c(R) \arrow[d, "\simeq"] \\ \underline{\operatorname{Alpha}(r,C)} \arrow[r, hook] & \underline{\operatorname{Alpha}(R,C)} \end{tikzcd}$$

naturally commutes. This might not seem very exciting at first. Nevertheless, it carries the important information that unions of growing balls and Alpha complexes share the same persistences of "holes" which will become clear in section 2.4.[2]

Again, there is an equivalent description via empty spheres. For this to work, we observe that an empty sphere, like in lemma 2.2.11, does not have to be unique. However, there exists a smallest such.

Definition 2.2.14. Let $C \subseteq \mathbb{R}^d$ be a finite set of points and let $\alpha \in \operatorname{Del}(C)$. We denote by S_α the *smallest empty* $(d-1)$*-sphere* that passes through all points of α, which has the radius $r_\alpha \geq 0$. On top of that, we define a map

$$\operatorname{Rad} : \operatorname{Del}(C) \to \mathbb{R}_0^+, \quad \operatorname{Rad}(\alpha) = r_\alpha,$$

which we refer to as the *radius function on* $\operatorname{Del}(C)$.

Lemma 2.2.15. *Let $C \subseteq \mathbb{R}^d$ be a finite set of points, then*

$$\operatorname{Alpha}(r, C) = \operatorname{Rad}^{-1}([0, r]).$$

Proof. $\boxed{\subseteq}$ First, let $\alpha \in \operatorname{Alpha}(r, C)$. Choose $x \in \bigcap_{c\in\alpha} B_c^V(r, C) \neq \emptyset$ such that it is the center of S_α. This we can require because of lemma 2.2.11. Then the radius of this sphere is $r_\alpha \leq r$, hence $\alpha \in \operatorname{Rad}^{-1}([0, r])$.
$\boxed{\supseteq}$ Conversely, let $\alpha \in \operatorname{Del}(C)$ such that its smallest empty sphere S_α has radius $r_\alpha \leq r$. The center x of S_α has equal distance (namely r_α) to all of the points in α and, by definition of empty spheres, the same or larger distance to any other point in C. Hence, $x \in \bigcap_{c\in\alpha} V(c, C)$ and, simultaneously, $x \in \bigcap_{c\in\alpha} B_c(r_\alpha)$. Therefore, $\alpha \in \operatorname{Alpha}(r_\alpha, C) \subseteq \operatorname{Alpha}(r, C)$. ☐

[2]The same holds for the Čech complex as well, but we will not focus on that.

2.2.1. About the Radius Function

We follow [3], [8] and the introduction given in [19], in order to talk about the radius function in more detail, as these considerations turn out to be essential in section 2.7.1.

Definition 2.2.16. Let A be a simplicial complex and consider a function $f : A \to \mathbb{R}$. If $f(\beta) \leq f(\alpha)$ whenever β is a face of $\alpha \in A$, we say that f is *monotonic*. If, additionally, f is constant on the set $[\beta, \alpha] = \{\gamma \in A \mid \beta \subseteq \gamma \subseteq \alpha\}$, we call $[\beta, \alpha]$ an *interval of* f. Such an interval is *maximal* if there is no strictly larger interval that contains it.
We, further, call a maximal interval $[\beta, \alpha]$ *singular* if $\beta = \alpha$ and *non-singular* otherwise. In the singular case, the only simplex α we get is said to be *critical*. Any other simplex that is not critical is simply called *non-critical*.

Definition 2.2.17. Let A be a simplicial complex and let $f : A \to \mathbb{R}$ be a monotonic function. If the collection of all maximal intervals of f yields a partition of A, we say that f is a *generalized discrete Morse function*.

Proposition 2.2.18. *Given a finite point set $C \in \mathbb{R}^d$ that is in general position. Then radius function* $\mathrm{Rad} : \mathrm{Del}(C) \to \mathbb{R}_0^+$ *is generalized discrete Morse. In particular,* Rad *is monotonic.*

In [8], a proof is given for this statement in an even more general fashion that involves chromatic point sets, to which we will come in section 2.5.

2.3. Simplicial Homology

Homology theory is a main branch of algebraic topology and is covered in many textbooks. Here, we mainly focus on simplicial homology that works with simplicial complexes as the underlying geometrical object and refer to [22] and [29] as used literature.

2.3.1. Chains and Homology Classes

Definition 2.3.1. Let σ_k be a k-simplex in $\mathbb{R}^d$ spanned by $p_0, \dots, p_k \in \mathbb{R}^d$ and consider now the $(k+1)!$-many ways to order the vertices of σ_k.
Take two permutations $\tau_1, \tau_2 \in S(k+1)$ and define the equivalence relation

$$(p_{\tau_1(0)}, \dots, p_{\tau_1(k)}) \sim (p_{\tau_2(0)}, \dots, p_{\tau_2(k)}) \Leftrightarrow \exists \rho \in S_{\text{even}}(k+1) \text{ such that } \tau_2 = \rho \circ \tau_1,$$

where $S_{\text{even}}(k+1)$ denotes the set of even permutations. The equivalence class of some ordering $(p_{\tau(0)}, \dots, p_{\tau(k)})$ is written as $[p_{\tau(0)}, \dots, p_{\tau(k)}]$ and we call it an *orientation of* σ_k. Moreover, a k-simplex σ_k with a fixed orientation is called an *oriented k-simplex*.

Remark 2.3.2. (1) Usually, we fix the orientation $(p_0, \dots, p_k)$ and, in short, this reads

$$\sigma_k = [p_0, \dots, p_k].$$

(2) For $k \geq 1$, σ_k has two orientations, whereas for $k = 0$, we only get one orientation since σ_0 defines a single point.

Definition 2.3.3. Let σ_k be a k-simplex in $\mathbb{R}^d$ with $k \geq 1$. If $\sigma_k = [p_0, \dots, p_k]$, then

$$\sigma_k^{-1} = [p_1, p_0, \dots, p_k]$$

is called the *opposite oriented k-simplex of σ_k*.

Definition 2.3.4. Let $(G, +)$ be an abelian group. A subset $B \subseteq G$ is a *basis of G* if any $0 \neq g \in G$ has the (up to ordering of the summands) unique representation of the form

$$g = \sum_{b \in B} n_b\, b, \quad n_b \in \mathbb{Z},$$

where $n_b \neq 0$ for only finitely many $b \in B$.
We call G a *free abelian group* if it possesses a basis. The cardinality of such a basis B is named the *rank of G*.

Definition 2.3.5. Let K be a simplicial complex in $\mathbb{R}^d$. For $k \in \mathbb{Z}$ construct the *k^{th} (simplicial) chain group $C_k(K)$ of K* by obeying the following rules:

(a) For $k < 0$ and $k > \dim(K)$, $C_k(K)$ equals the trivial group, so $C_k(K) = 0$.

(b) For $k = 0$, $C_0(K)$ is the free abelian group generated by the vertices of K.

(c) For $0 < k \leq \dim(K)$, $C_k(K)$ is the free abelian group generated by all oriented k-simplices σ_k of K modulo the relation that opposite oriented k-simplices are viewed as being inverse to each other in a group theoretical sense.

The elements of $C_k(K)$ are called *(simplicial) k-chains in K*.

The set $C_k(K)$, indeed, is a free abelian group by construction as the following proposition summarizes.

Proposition 2.3.6. *Let K be a simplicial complex in $\mathbb{R}^d$. For every k-simplex in K fix an orientation and write them as $\sigma_k^{(1)}, \dots, \sigma_k^{(m_k)}$, where $m_k \in \mathbb{N}_0$ is the number of k-simplices in K. Then any k-chain $c \in C_k(K)$ has a unique representation as a formal sum*

$$c = \sum_{l=1}^{m_k} n_l\, \sigma_k^{(l)}, \quad n_l \in \mathbb{Z}.$$

Consequently, $C_k(K)$ is a free abelian group of rank m_k with basis $\{\sigma_k^{(1)}, \dots, \sigma_k^{(m_k)}\}$.

Remark 2.3.7. As a group, $C_k(K)$ has the neutral element $0 = \sum_{l=1}^{m_k} 0 \cdot \sigma_k^{(l)}$ and by definition 2.3.5(c) we get that $(\sigma_k^{(l)})^{-1} = (-1)\, \sigma_k^{(l)}$ for every $l = 1, \dots, m_k$.

Definition 2.3.8. Let K be a simplicial complex in $\mathbb{R}^d$ and let $\sigma_k = [p_0, \dots, p_k]$ be an oriented k-simplex in K. The *boundary of* σ_k is the following $(k-1)$-chain in K:

$$\partial_k\, \sigma_k = \begin{cases} \sum_{i=0}^{k} (-1)^i\, [p_0, \dots, \hat{p_i}, \dots, p_k], & 0 < k \leq \dim(K), \\ 0 & k = 0. \end{cases}$$

Here, $\hat{p_i}$ means that the point p_i is missing, so that $[p_0, \dots, \hat{p_i}, \dots, p_k]$ is an oriented $(k-1)$-simplex in K.
The *boundary of a k-chain* $c = \sum_{l=1}^{m_k} n_l\, \sigma_k^{(l)} \in C_k(K)$ then is defined as

$$\partial_k\, c = \sum_{l=1}^{m_k} n_l\, \partial_k\, \sigma_k^{(l)} \in C_{k-1}(K).$$

This gives a group homomorphism $\partial_k : C_k(K) \to C_{k-1}(K)$, which is named the k^{th} *boundary operator*. In particular, for $k < 0$ and $k > \dim(K)$, ∂_k is set to be the zero map.

Definition 2.3.9. Let K be a simplicial complex in $\mathbb{R}^d$.

(i) A k-chain $c \in C_k(K)$ is a *k-cycle* if $\partial_k\, c = 0$. The subgroup

$$Z_k(K) = \{c \in C_k(K) \mid \partial_k\, c = 0\} = \ker(\partial_k)$$

of $C_k(K)$ is called the k^{th} *cycle group of* K.

(ii) A k-chain $c \in C_k(K)$ is a *k-boundary* if there exists a $d \in C_{k+1}(K)$ such that $\partial_{k+1}\, d = c$. The subgroup

$$B_k(K) = \{c \in C_k(K) \mid \exists d \in C_{k+1}(K) : \partial_{k+1}\, d = c\} = \operatorname{im}(\partial_{k+1})$$

of $C_k(K)$ is called the k^{th} *boundary group of* K.

(iii) Two k-chains $b, c \in C_k(K)$ are said to be *homologous* if $b - c \in B_k(K)$. This defines an equivalence relation and the equivalence class

$$\{c\} = \{b \in C_k(K) \mid b - c \in B_k(K)\} = c + B_k(K)$$

is called the *homology class of* c.

Definition 2.3.10. A *chain complex* is a system $G = (G_k, f_k)_{k \in \mathbb{Z}}$ of abelian groups G_k and group homomorphisms $f_k : G_k \to G_{k-1}$ such that

$$f_k \circ f_{k+1} \equiv 0 \quad \forall k \in \mathbb{Z}.$$

Proposition 2.3.11. *Let K be a simplicial complex in $\mathbb{R}^d$. Then we have*

$$\partial_k \circ \partial_{k+1} \equiv 0 \quad \forall k \in \mathbb{Z}.$$

In particular, the system $C(K) = (C_k(K), \partial_k)_{k \in \mathbb{Z}}$ is a chain complex and $B_k(K)$ is a normal subgroup of $Z_k(K)$ for any $k \in \mathbb{Z}$.

Definition 2.3.12. Let K be a simplicial complex in $\mathbb{R}^d$. We define

$$H_k(K) = Z_k(K)/B_k(K) = \{\{z\} \mid z \in Z_k(K)\}$$

to be the k^{th} *(simplicial) homology group of* K.

Remark 2.3.13. (1) Since the group $C_k(K)$ is abelian and finitely generated as every simplicial complex K is a finite object, so are $Z_k(K)$, $B_k(K)$ and $H_k(K)$.
(2) The group $H_k(K)$ is non-trivial if and only if there is a k-cycle that is not a k-boundary. Thus, the k^{th} homology group measures how many "essential" k-cycles there are in K.

2.3.2. Intermezzo: From Coefficients in $\mathbb{Z}$ to $\mathbb{Z}_2$

In computational topology, it is often more useful and intuitive to work with homology with coefficients in $\mathbb{Z}_2 = \{0, 1\}$ instead of $\mathbb{Z}$. In order to sustain the powerful machinery of algebraic topology behind it, we are going to derive the k^{th} homology group $H_k(K; \mathbb{Z}_2)$ and – before that – the k^{th} chain group $C_k(K; \mathbb{Z}_2)$ with $\mathbb{Z}_2$-coefficients from a more general concept, where we want an arbitrary abelian group $(G, +)$ to be our coefficient set. But first we need some more background knowledge on tensor products of abelian groups.

Lemma 2.3.14. *Let* $X \neq \emptyset$ *and set*

$$\mathcal{F}(X) = \{f : X \to \mathbb{Z} \mid f(x) \neq 0 \textit{ for only finitely many } x \in X\}.$$

Then $(\mathcal{F}(X), +)$ *defines a free abelian group with addition being the pointwise operation of maps.*

Proof. Obviously, $\mathcal{F}(X)$ is an abelian group. It is left to show that it possesses a basis $B \subseteq \mathcal{F}(X)$. For fixed $x \in X$, let $\mathbb{1}_x : X \to \mathbb{Z}$ be the indicator function of x and set $n_x = f(x) \in \mathbb{Z}$. Then we can write any $f \in \mathcal{F}(X)$ uniquely in the form

$$f = \sum_{x \in X} n_x \, \mathbb{1}_x,$$

thus $B = \{\mathbb{1}_x\}_{x \in X}$ gives us a basis of $\mathcal{F}(X)$. □

Definition 2.3.15. Let $X \neq \emptyset$ be a set. We call $\mathcal{F}(X)$ from lemma 2.3.14 the *from X freely generated abelian group* and instead of $\mathbb{1}_x \in \mathcal{F}(X)$ we write just $x \in X$, so that all elements of $\mathcal{F}(X)$ look like $\sum_{x \in X} n_x \, x$.

Definition 2.3.16. Let $(G_1, +), (G_2, +)$ be abelian groups and consider $\mathcal{F}(G_1 \times G_2)$. We define $\mathcal{R}(G_1 \times G_2) \subseteq \mathcal{F}(G_1 \times G_2)$ to be the subgroup that is generated by the elements

$$(g_1 + g_1', g_2) - (g_1, g_2) - (g_1', g_2) \text{ and } (g_1, g_2 + g_2') - (g_1, g_2) - (g_1, g_2') \tag{2.3}$$

with $g_1, g_1' \in G_1$, $g_2, g_2' \in G_2$. The quotient group

$$G_1 \otimes G_2 = \mathcal{F}(G_1 \times G_2)/\mathcal{R}(G_1 \times G_2)$$

is the *tensor product of* G_1 *and* G_2.

Lemma 2.3.17. *Let* $(G_1, +), (G_2, +)$ *be abelian groups. In* $G_1 \otimes G_2$ *we have the equalities*

$$(g_1 + g_1') \otimes g_2 = g_1 \otimes g_2 + g_1' \otimes g_2 \;\; \textit{and} \;\; g_1 \otimes (g_2 + g_2') = g_1 \otimes g_2 + g_1 \otimes g_2'.$$

It immediately follows then that any element in $G_1 \otimes G_2$ *can be written as a (not uniquely determined) sum with summands of the form* $g_1 \otimes g_2$.

Proof. The claim is due to (2.3). □

Definition 2.3.18. Let K be a simplicial complex in $\mathbb{R}^d$ and let $(G, +)$ be an abelian group. We define the k^{th} *(simplicial) chain group with coefficients in* G as

$$C_k(K; G) = C_k(K) \otimes G.$$

Example 2.3.19. (1) First, observe that for any abelian group $(A, +)$ we can build a group isomorphism $A \cong A \otimes \mathbb{Z}$ via the mapping $a \mapsto a \otimes 1$ and with inverse $a \otimes k \mapsto ka$. Thus, we may identify a with $a \otimes 1$ and write $A = A \otimes \mathbb{Z}$. Hence, setting $A = C_k(K)$ we end up with

$$C_k(K; \mathbb{Z}) = C_k(K) \otimes \mathbb{Z} = C_k(K),$$

for $G = \mathbb{Z}$, which gives us the ordinary k^{th} chain group that was constructed in definition 2.3.5.

(2) Let $(A, +)$ be a free abelian group of rank m and let $\{a_i\}_{i \in I}$ be a basis of A (with $\mathrm{card}(I) = m$). It can be shown that for any abelian group $(G, +)$ we obtain

$$A \otimes G \cong \bigoplus_{i \in I} G.$$

Therefore, any $z \in A \otimes G$ has the unique representation $z = \sum_{i \in I} a_i \otimes g_i$ where only finitely many $g_i \neq 0$.

In particular, for $A = C_k(K)$ and $G = \mathbb{Z}_2$ this yields that any element $c \in C_k(K; \mathbb{Z}_2)$ looks like

$$c = \sum_{l=1}^{m_k} n_l \, (\sigma_k^{(l)} \otimes g_l) \overset{2.3.17}{=} \sum_{l=1}^{m_k} \sigma_k^{(l)} \otimes (n_l \, g_l) \quad (n_l \in \mathbb{Z}),$$

where $g_l = 0$ or $g_l = 1$. Since then $n_l \, g_l \in \mathbb{Z}_2$ and $\sigma_k^{(l)} \otimes 0 = 0$ as well as $\sigma_k^{(l)} \otimes 1 = \sigma_k^{(l)}$ by (1), this leads us to

$$c = \sum_{l=1}^{m_k'} \sigma_k^{(l)} \;\text{ with }\; m_k' \leq m_k$$

after some possible rearrangement of indices.
From remark 2.3.7 we, moreover, conclude that

$$(\sigma_k^{(l)})^{-1} = (-1)\,\sigma_k^{(l)} = \sigma_k^{(l)}$$

for $G = \mathbb{Z}_2$, which tells us that we do not have to care about orientation anymore. The corresponding k^{th} boundary operator $\partial_k : C_k(K;\mathbb{Z}_2) \to C_{k-1}(K;\mathbb{Z}_2)$ now reads

$$\partial_k\, c = \sum_{l=1}^{m'_k} \partial_k\, \sigma_k^{(l)},$$

where the boundary of some k-simplex is defined to be the sum of all its $(k-1)$-dimensional faces. For $k < 0$ and $k > \dim(K)$, ∂_k is again the zero map.
With those tools in hand, we can define the k^{th} cycle group $Z(K;\mathbb{Z}_2)$, the k^{th} boundary group $B(K;\mathbb{Z}_2)$ and the k^{th} homology group $H(K;\mathbb{Z}_2)$ with coefficients in $\mathbb{Z}_2$ in a similar fashion as already done with integer coefficients.

Remark 2.3.20. Let A be an abelian group and let $(F,+,\cdot)$ be a field. Then the tensor product $A \otimes F$ equipped with the additional operation of multiplication with elements in F

$$x'(a \otimes x) = a \otimes (x' \cdot x) \quad (a \in A,\ x, x' \in F)$$

forms a vector space over F. If, further, A is a free abelian group of rank m and with basis $\{a_i\}_{i\in I}$, then $A \otimes F$ – seen as a vector space – is of dimension m and has basis $\{a_i \otimes 1\}_{i\in I}$. Here, 1 denotes the multiplicative neutral element of F.
In the special case $A = C_k(K)$ and $F = \mathbb{Z}_2$, we immediately get that $C_k(K;\mathbb{Z}_2)$ is a vector space and so are $Z_k(K;\mathbb{Z}_2)$, $B_k(K;\mathbb{Z}_2)$ and $H_k(K;\mathbb{Z}_2)$. The dimension (or rank) of $H_k(K;\mathbb{Z}_2)$ is called the k^{th} *Betti number*

$$\beta_k = \beta_k(K;\mathbb{Z}_2) = \operatorname{rank}(H_k(K;\mathbb{Z}_2)).$$

Moreover, the group homomorphism ∂_k becomes a linear map.

Henceforth, we shall only deal with the case $G = \mathbb{Z}_2$ to exploit the vector space nature of homology groups. For the sake of simplicity and better readability, we will make a slight abuse of notation and write $C_k(K), H_k(K), \ldots$ instead of $C_k(K;\mathbb{Z}_2), H_k(K;\mathbb{Z}_2), \ldots$.

2.3.3. Simplicial Maps and Induced Homomorphisms

Definition 2.3.21. Let K, L be two simplicial complexes. A map $\varphi : K \to L$ is a *simplicial map* if the following properties hold:

(a) φ maps vertices of K to vertices of L.

(b) If $K \ni \sigma_k = \operatorname{conv}\{p_0, \ldots, p_k\}$, then $L \ni \varphi(\sigma_k) = \operatorname{conv}\{\varphi(p_0), \ldots, \varphi(p_k)\}$.

If, furthermore, φ is bijective, we call φ an *isomorphism* and we say that K and L are *isomorphic*.

Definition 2.3.22. Let K, L be two simplicial complexes and let $\varphi : K \to L$ be a simplicial map. For any $k \in \mathbb{Z}$ we define a group homomorphism

$$(\varphi_k)_\bullet : C_k(K) \to C_k(L)$$

between the k^{th} chain groups of K and L by requiring the below settings:

(a) For $k < 0$ or $k > \dim(K)$, set $(\varphi_k)_\bullet \equiv 0$.

(b) For $0 \leq k \leq \dim(K)$, let $\sigma_k = \operatorname{conv}\{p_0, \ldots, p_k\} \in K$ and set

$$(\varphi_k)_\bullet(\sigma_k) = \begin{cases} \operatorname{conv}\{\varphi(p_0), \ldots, \varphi(p_k)\} & \text{if } \varphi(p_i) \neq \varphi(p_j) \text{ for } i \neq j, \\ 0 & \text{otherwise.} \end{cases}$$

(c) For $0 \leq k \leq \dim(K)$, let $c = \sum_{l=1}^{m'_k} \sigma_k^{(l)} \in C_k(K)$ and set

$$(\varphi_k)_\bullet(c) = \sum_{l=1}^{m'_k} (\varphi_k)_\bullet(\sigma_k^{(l)}).$$

We say that $(\varphi_k)_\bullet$ is the *induced homomorphism of the k^{th} chain groups.*

Definition 2.3.23. Let $G = (G_k, f_k)_{k\in\mathbb{Z}}$, $G' = (G'_k, f'_k)_{k\in\mathbb{Z}}$ be two chain complexes. A *chain map* $\Lambda : G \to G'$ is a system $\Lambda = (\Lambda_k)_{k\in\mathbb{Z}}$ of group homomorphisms $\Lambda_k : G_k \to G'_k$ such that

$$f'_k \circ \Lambda_k = \Lambda_{k-1} \circ f_k \quad \forall k \in \mathbb{Z}.$$

Proposition 2.3.24. *Let K, L be two simplicial complexes, let $\varphi : K \to L$ be a simplicial map and consider the chain complexes $C(K) = (C_k(K), \partial_k^{(K)})_{k\in\mathbb{Z}}$ and $C(L) = (C_k(L), \partial_k^{(L)})_{k\in\mathbb{Z}}$. Then we have*

$$\partial_k^{(L)} \circ (\varphi_k)_\bullet = (\varphi_{k-1})_\bullet \circ \partial_k^{(K)} \quad \forall k \in \mathbb{Z}. \qquad \begin{tikzcd} C_k(K) \arrow[r, "\partial_k^{(K)}"] \arrow[d, "(\varphi_k)_\bullet"'] & C_{k-1}(K) \arrow[d, "(\varphi_{k-1})_\bullet"] \\ C_k(L) \arrow[r, "\partial_k^{(L)}"] & C_{k-1}(L) \end{tikzcd}$$

In particular, $\varphi_\bullet : C(K) \to C(L)$ is a chain map and any $(\varphi_k)_\bullet$ maps k-cycles (resp. k-boundaries) in K to k-cycles (resp. k-boundaries) in L.

Definition 2.3.25. Let K, L be two simplicial complexes and let $\varphi : K \to L$ be a simplicial map. For any $k \in \mathbb{Z}$ we define the group homomorphism

$$(\varphi_k)_* : H_k(K) \to H_k(L), \quad (\varphi_k)_*(\{z\}) = \{(\varphi_k)_\bullet(z)\}.$$

We call $(\varphi_k)_*$ the *induced homomorphism of the k^{th} homology groups.*

Example 2.3.26. (1) Given the identity map $\varphi = \mathrm{id}_K : K \to K$, which is obviously a simplicial map, we easily obtain

$$(\varphi_k)_\bullet = \mathrm{id}_{C_k(K)}, \quad (\varphi_k)_* = \mathrm{id}_{H_k(K)}.$$

(2) Let $L \subseteq K$ be a subcomplex of K. Clearly, the inclusion map $\iota : L \hookrightarrow K$ is a simplicial map.
(3) Let $\varphi : K \to L$, $\psi : L \to M$ be simplicial maps, then

$$\begin{aligned}(\psi_k \circ \varphi_k)_\bullet &= (\psi_k)_\bullet \circ (\varphi_k)_\bullet : C_k(K) \to C_k(M),\\ (\psi_k \circ \varphi_k)_* &= (\psi_k)_* \circ (\varphi_k)_* : H_k(K) \to H_k(M).\end{aligned}$$

(4) Let $\varphi : K \to L$ be an isomorphism, then $(\varphi_k)_\bullet$ and $(\varphi_k)_*$ are group isomorphisms.

2.3.4. Relative Simplicial Homology Groups

Definition 2.3.27. Let $G = (G_k, f_k)_{k\in\mathbb{Z}}$ be a chain complex and let, for any $k \in \mathbb{Z}$, $H_k \subseteq G_k$ be a subgroup such that $f_k(H_k) \subseteq H_{k-1}$. Then the system $H = (H_k, f_k|_{H_k})$ is a chain complex and it is called a *chain subcomplex*. We write $H \subseteq G$ for short and we say that (G, H) is a *pair of chain complexes*.

Definition 2.3.28. Let (G, H) be a pair of chain complexes, then

$$G/H = (G_k/H_k, \overline{f}_k)_{k\in\mathbb{Z}}$$

is the *quotient chain complex* with

$$\overline{f}_k : G_k/H_k \to G_{k-1}/H_{k-1}, \quad \overline{f}_k(g_k + H_k) = f_k(g_k) + H_{k-1}.$$

Definition 2.3.29. Let K be a simplicial complex and let $L \subseteq K$ be a simplicial subcomplex of K. Clearly, $C_k(L) \subseteq C_k(K)$ and the quotient group

$$C_k(K, L) = C_k(K)/C_k(L) = \{\overline{c} = c + C_k(L) \mid c \in C_k(K)\}$$

is called the k^{th} *relative chain group of* K *relative* L and its elements are called k-*chains of* K *relative* L.

Remark 2.3.30. It is easy to see that $(C(K), C(L))$ is a pair of chain complexes. Thus, $C(K, L) = (C_k(K, L), \overline{\partial}_k)_{k\in\mathbb{Z}}$ is the corresponding quotient chain complex with k^{th} *relative boundary operator*

$$\overline{\partial}_k : C_k(K, L) \to C_{k-1}(K, L), \quad \overline{\partial}_k(c + C_k(L)) = \partial_k\, c + C_{k-1}(L).$$

Definition 2.3.31. Let K be a simplicial complex and let $L \subseteq K$ be a simplicial subcomplex of K.

(i) The k^{th} *cycle group of K relative L* is defined to be

$$Z_k(K, L) = \ker(\overline{\partial}_k)$$

and its elements are the *k-cycles relative L*. To be more specific,

$$\overline{c} = c + C_k(L) \in Z_k(K, L) \quad \Leftrightarrow \quad \partial_k\, c \in C_{k-1}(L).$$

Thus, if $c \in C_k(K)$ has the property $\partial_k\, c \in C_{k-1}(L)$, we simply say that c is a k-cycle relative L, even though we actually mean its equivalence class $\overline{c}$.

(ii) The k^{th} *boundary group of K relative L* is defined to be

$$B_k(K, L) = \operatorname{im}(\overline{\partial}_{k+1})$$

and its elements are the *k-boundaries relative L*.

(iii) Two relative k-chains $\overline{b}, \overline{c} \in C_k(K, L)$ are said to be *homologous relative L* if $\overline{b - c} \in B_k(K, L)$. This defines an equivalence relation and the equivalence class

$$\{\overline{c}\} = \{\overline{b} \in C_k(K, L) \mid \overline{b - c} \in B_k(K, L)\} = \overline{c} + B_k(K, L)$$

is called the *relative homology class of $\overline{c}$*.

(iv) By remark 2.3.30, $C(K, L)$ is a chain complex, so we can define the k^{th} *(simplicial) homology group of K relative L* as

$$H_k(K, L) = Z_k(K, L)/B_k(K, L) = \{\{\overline{z}\} \mid \overline{z} \in Z_k(K, L)\}.$$

Example 2.3.32. In the following, let K always be a simplicial complex with simplicial subcomplex $L \subseteq K$.

(1) Given $L = \emptyset$, then

$$H_k(K, \emptyset) = H_k(K) \quad \forall k \in \mathbb{Z}.$$

(2) Given $L = K$, then

$$H_k(K, K) = 0 \quad \forall k \in \mathbb{Z},$$

where 0 indicates that $H_k(K, K)$ is the trivial group.

(3) Let $\dim(L) < k - 1$, then

$$H_k(K, L) = H_k(K).$$

Indeed, for $j \in \{k - 1, k, k + 1\}$ we obtain $C_j(L) = 0$ by definition. Hence, $C_j(K, L) = C_j(K)$ and $\overline{\partial}_j = \partial_j$. Since these are the only relevant ones for construction $H_k(K, L)$, the claim follows.

Definition 2.3.33. Given a sequence of abelian groups G_k ($k \in \mathbb{Z}$) and group homomorphisms $f_k : G_k \to G_{k-1}$, so

$$\cdots \longrightarrow G_{k+1} \xrightarrow{f_{k+1}} G_k \xrightarrow{f_k} G_{k-1} \longrightarrow \ldots$$

This sequence is said to be *exact at G_k* if

$$\operatorname{im}(f_{k+1}) = \ker(f_k).$$

If this is true for any $k \in \mathbb{Z}$, the sequence is *exact* or we say that this is a *long exact sequence*. If the exact sequence is of the form

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0,$$

we call it a *short exact sequence*. In addition, if there is a group isomorphism $B \cong A \oplus C$ then the short exact sequence *splits*.

Proposition 2.3.34. *Let A, B, C be abelian groups such that*

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0,$$

is exact. If C is free abelian, then this short exact sequence splits.

Theorem 2.3.35. *Let K be a simplicial complex and let $L \subseteq K$ be a simplicial subcomplex of K. Then there exists the long exact sequence*

$$\cdots \longrightarrow H_{k+1}(K, L) \xrightarrow{(\partial_{k+1})_*} H_k(L) \xrightarrow{(\iota_k)_*} H_k(K) \xrightarrow{(q_k)_*} H_k(K, L) \longrightarrow \ldots$$

where we define the group homomorphisms

$$\begin{aligned}
&(\iota_k)_* : H_k(L) \to H_k(K), \quad (\iota_k)_*(\{z\}_L) = \{z\}_K, \\
&(q_k)_* : H_k(K) \to H_k(K, L), \quad (q_k)_*(\{z\}_K) = \{\overline{z}\}, \\
&(\partial_k)_* : H_k(K, L) \to H_{k-1}(L), \quad (\partial_k)_*(\{\overline{z}\}) = \{\partial\, z\}_L.
\end{aligned}$$

2.3.5. Connection to Singular Homology

Obviously, not every topological space is a simplicial complex. However, we would also like to have some notion of homology for those spaces. *Singular homology* gives us a generalization of simplicial homology to arbitrary topological spaces. In the following, we briefly describe the most important concepts that are relevant for this thesis.

Definition 2.3.36. Let X be a topological space. A *singular k-simplex in X* is a continuos map $\sigma_k : \Delta_k \to X$, where Δ_k denotes the *standard k-simplex* which is the convex hull of the unit points

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad e_{k+1} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \tag{2.4}$$

of $\mathbb{R}^{k+1}$.

Remark 2.3.37. In a similar fashion to simplicial homology we can define the k^{th} *singular chain group of* X, denoted by $S_k(X)$, as well as the k^{th} *singular homology group of* X, denoted by $H_k(X)$. In addition, for a subspace $A \subseteq X$, there is $H_k(X, A)$, the k^{th} *relative singular homology group of* (X, A).

Theorem 2.3.38. *Let K be a simplicial complex with its underlying space $|K|$. Then there exists an isomorphism between the simplicial homology group $H_k(K)$ and the singular homology group $H_k(|K|)$.*

Theorem 2.3.39. *Let K be a simplicial complex and let $L \subseteq K$ be a simplicial subcomplex. Then for any $k \in \mathbb{Z} \setminus \{0\}$ there exists the isomorphism*

$$H_k(K, L) \cong H_k(|K|, |L|) \cong H_k(|K|/|L|).$$

For the special case $k = 0$ we get[3]

$$H_k(K, L) \oplus \mathbb{Z}_2 \cong H_k(|K|/|L|).$$

2.4. Persistent Homology

We now come to the part where we want to capture the topological features like clusters or holes, that appear and disappear within Alpha complexes as the radius increases. This we will do by studying filtrations of simplicial complexes that give rise to sequences of vector spaces. After explaining the theory about this method in detail, we give a simple example on how to apply it, and throughout this section we follow [15] and [16].

Definition 2.4.1. Let K be a simplicial complex. A *filtration of* K is a collection of subcomplexes $(K_i)_{i \in I}$, where I is a totally ordered index set, and such that $K_i \subseteq K_j$ whenever $i \leq j$.

Example 2.4.2. By the remarks 2.2.2 and 2.2.13, both the families $\big(\check{\mathrm{C}}\mathrm{ech}(r, C)\big)_{r \geq 0}$ and $\big(\mathrm{Alpha}(r, C)\big)_{r \geq 0}$ are filtrations. In particular, the Alpha complexes build a filtration of the Delaunay complex of a given finite point set $C \subseteq \mathbb{R}^d$.

Remark 2.4.3. (1) Another way of viewing a filtration of a simplicial complex K is by considering a monotonic function $f : K \to \mathbb{R}$ and its sublevel sets $K(r) = f^{-1}(-\infty, r]$. All $K(r)$ are subcomplexes of K and $\big(K(r)\big)_{r \in \mathbb{R}}$ becomes a filtration of K. In this sense, we can also argue that $\big(\mathrm{Alpha}(r, C)\big)_{r \geq 0}$ is a filtration as every Alpha complex is the sublevel set of the radius function Rad, which is monotonic by proposition 2.2.18.
(2) Since we always regard a simplicial complex K to be a finite object, a filtration, thus, can only consist of finitely many subcomplexes. More specifically, letting $m \geq 1$ be the total number of simplices in K, we get at most m-many non-empty different

[3] Beware that we work with the coefficient group $G = \mathbb{Z}_2$. For $G = \mathbb{Z}$ the additional $\mathbb{Z}_2$ for $k = 0$ changes to $\mathbb{Z}$.

subcomplexes that are part of a filtration and we arrange them as a nested increasing sequence

$$\emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K \quad \text{for some } n \leq m. \tag{2.5}$$

Here, we added the empty set as the starting point of the filtration. This will be useful in the next step when applying homology theory on such a structure.

Indeed, we are not so much interested in a sequence of subcomplexes but rather in their topological evolution. Thus, taking $\iota^{i,j} : K_i \hookrightarrow K_j$ to be the inclusion map for any $0 \leq i \leq j \leq n$, we can, for every integer[4] $k \geq 0$, view the corresponding induced homomorphism $\big(\iota_k^{i,j}\big)_* : H_k(K_i) \to H_k(K_j)$ and build a sequence of homology groups

$$0 = H_k(K_0) \longrightarrow H_k(K_1) \longrightarrow \cdots \longrightarrow H_k(K_{n-1}) \longrightarrow H_k(K_n) = H_k(K)$$

connected by the aforementioned homomorphisms, which even are linear maps by what was stated in remark 2.3.20. This sequence is a *persistence module*, which essentially is a sequence of vector spaces connected by linear maps. As we go from one subcomplex in the filtration to the next, we might gain new homology classes when non-trivial cycles are added, and we might lose some when they become trivial or merge. We then collect all homology classes that are "born" at or before a given threshold and "die" after another threshold in groups.

Definition 2.4.4. Let K be a simplicial complex and let $(K_i)_{i=0}^n$ be a filtration of K like in (2.5). The k^{th} *persistent homology group* is given by

$$H_k^{i,j}(K) = \operatorname{im}\big(\big(\iota_k^{i,j}\big)_*\big) = Z_k(K_i)/(B_k(K_j) \cap Z_k(K_i)) \quad (0 \leq i \leq j \leq n)$$

and the rank of $H_k^{i,j}(K)$ is called the k^{th} *persistent Betti number*

$$\beta_k^{i,j} = \operatorname{rank}\big(H_k^{i,j}(K)\big).$$

Remark 2.4.5. (1) In the special case $j = i$, the definition yields $H_k^{i,i}(K) = H_k(K_i)$.
(2) At first glance, the definition of persistent homology groups seems a bit mysterious, so let us be a bit more concrete about the classes counted by $H_k^{i,j}(K)$.
Given $\{z\} \in H_k(K_i)$, we say it is *born* at K_i if $\{z\} \notin H_k^{i-1,i}(K)$. Now let $\{z\} \in H_k(K_i)$ be born at K_i, then it *dies* when entering K_{j+1} if $\big(\iota_k^{i,j}\big)_*(\{z\}) \notin H_k^{i-1,j}(K)$ and if there exists a $\{z'\} \in H_k(K_{i-1})$ such that $\big(\iota_k^{i,j+1}\big)_*(\{z\}) = \big(\iota_k^{i-1,j+1}\big)_*(\{z'\})$, so if we have $\big(\iota_k^{i,j+1}\big)_*(\{z\}) \in H_k^{i-1,j+1}(K)$. Put differently, $\{z\}$ has *merged* to $\{z'\}$ – a class that has been living longer than $\{z\}$ – and $\{z'\}$ survives. This is formally known as the *Elder Rule*.
In summation, we observe that $H_k^{i,j}(K)$ contains all homology classes of K_i that are still alive in K_j.

[4]For $k < 0$ all homology groups are trivial, so we can ignore them for the sake of simplicity.

(3) Given $\{z\} \in H_k(K_i)$ that is born at K_i and dies entering K_{j+1}, we call $j+1-i$ its *index persistence*. If the filtration complexes K_i stem from sublevel sets $f^{-1}((-\infty, r_i])$ of a monotonic function $f : K \to \mathbb{R}$, the difference $r_{j+1} - r_i$ is the *persistence of* $\{z\}$. If $\{z\}$ is born at K_i but lives forever, we set its (index) persistence to infinity.

After this rather lengthy explanation of $H_k^{i,j}(K)$ the construction is still very technical and abstract. Thus, we would like to have a good visualization of what is going on for some simplicial complex K and a fixed $k \geq 0$, in order to keep track of the (persistent) Betti number(s) $\beta_k^{i,j}$ and β_k.

Definition 2.4.6. Given a filtration $(K_i)_{i=0}^n$ of the simplicial complex K via a monotonic function $f : K \to \mathbb{R}$ and an integer $k \geq 0$. The k^{th} *persistence diagram* $\mathrm{Dgm}_k(f)$ is the multiset in the extended plane $\overline{\mathbb{R}}^2 = (\mathbb{R} \cup \{\pm\infty\})^2$ that consists of points (r_i, r_{j+1}), each uniquely corresponding to a homology class that is born at K_i and dies entering K_{j+1}. The multiplicity of (r_i, r_{j+1}) we denote by $\mu_k^{i,j+1}$. Often we combine all k^{th} persistence diagrams to a big one and we call it the *persistence diagram* $\mathrm{Dgm}(f)$.

Remark 2.4.7. Because of $i \leq j$, we see that $r_i \leq r_j < r_{j+1}$ and, hence, for $r_i \geq 0$ all points (r_i, r_{j+1}) lie above the diagonal in the first quadrant. The persistence $r_{j+1} - r_i$ is then the vertical distance to the diagonal.

Lemma 2.4.8 (Fundamental lemma of persistent homology)**.** *Given a filtration $(K_i)_{i=0}^n$ of the simplicial complex K via a monotonic function $f : K \to \mathbb{R}$ and an integer $k \geq 0$. Then*

$$\beta_k^{l,m} = \sum_{i \leq l} \sum_{j > m} \mu_k^{i,j}.$$

What lemma 2.4.8 tells us is how to find $\beta_k^{l,m}$ in the persistence diagram. Recall that $\beta_k^{l,m}$, specifically, is the number of homology classes that are born at K_l or before and that are not yet dead in K_m (but might be in K_{m+1}). Therefore, such a homology class has the point representation (r_i, r_j) with $i \leq l$ and $j > m$. Counting all these classes is like summing up all $\mu_k^{i,j}$ with $i \leq l$ and $j > m$. We have, thus, proven lemma 2.4.8 along the way of interpreting it and conclude that $\beta_k^{l,m}$ is the number of points in the semi-open[5] rectangle with lower right corner point (r_l, r_m).

Before turning to a simple example that makes all these concepts and definitions more tangible, we quickly introduce 0- and 1-norms on the multiset $\mathrm{Dgm}_k(f)$.

Definition 2.4.9. Let K be a simplicial complex and let $f : K \to \mathbb{R}$ be a monotonic function that generates a filtration of K. Fix $k \geq 0$. The 0*-norm of* $\mathrm{Dgm}_k(f)$ then is defined to be the number of all persistence pairs

$$\|\mathrm{Dgm}_k(f)\|_0 = \mathrm{card}(\mathrm{Dgm}_k(f)).$$

[5] It is closed along its vertical right side and open along its horizontal lower one.

Moreover, the 1-*norm of* $\mathrm{Dgm}_k(f)$ is the sum of all its persistences

$$\|\mathrm{Dgm}_k(f)\|_1 = \sum_{(r_i,r_j)} r_j - r_i.$$

Remark 2.4.10. Clearly, $\|\mathrm{Dgm}_k(f)\|_0$ is always finite, but $\|\mathrm{Dgm}_k(f)\|_1$ can be infinite as soon as we have a pair of the form (r_i, ∞) and at least for $k = 0$, this is true in any case. To avoid this undesired artifact, we work with a *cut-off* $\delta \in \mathbb{R}$ that replaces infinity and is at least the largest finite death value amongst all k^{th} persistence pairs.

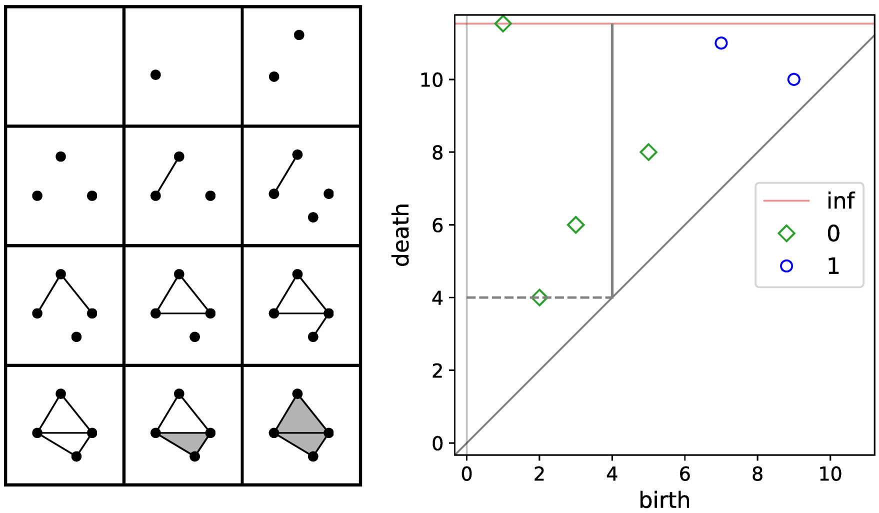

Figure 2.7.: A simple filtration consisting of 12 nested simplicial complexes and its persistence diagram. The unit, in which both birth and death values are given, corresponds to one filtration step.

Example 2.4.11. Consider the filtration given in Figure 2.7. It consists of 12 nested simplicial complexes – the 0^{th} one being the empty set and the 11^{th} one being the full simplicial complex we want to have in the end. Here, at each step one more simplex joins. If we apply the theory of persistent homology on this filtration we obtain its persistence diagram, which can also be found in Figure 2.7. Each point above the diagonal represents a homology class that is born in a specific filtration step and dies entering in another such.
In the 0-dimensional case, we have different 0-dimensional holes throughout the filtration – not appearing all at once. These correspond to the number of connected components. For instance, $\beta_0^{4,4} = 2$ since there are two points inside the semi-open rectangle that is drawn gray inside the diagram. Looking at the 2^{nd} cell in the 2^{nd} row of the filtration grid, this matches with the two connected components that are present there. A special connected component is the one that is born first and lives forever due to the Elder Rule. Similar oberservations can be made for the 1-dimensional case, where 1-dimensional holes

are the cycles that are formed by edges. Such a cycle dies whenever it is fully filled by triangles.
Finally, we can also compute the 0- and 1-norms of each persistence diagram of dimension $k = 0, 1$. Using, for instance, $\delta = 5$ as a cut-off, we get

$$\|\mathrm{Dgm}_0(f)\|_0 = 4, \quad \|\mathrm{Dgm}_1(f)\|_0 = 2, \quad \|\mathrm{Dgm}_0(f)\|_1 = 8, \quad \|\mathrm{Dgm}_1(f)\|_1 = 5.$$

Example 2.4.12. Viewing the same simplicial complex from the last step 11 in Figure 2.7, we notice that it represents a simple Delaunay complex when embedded into $\mathbb{R}^2$. Running an Alpha complex filtration through it with f being the radius function Rad, we get a different persistence diagram, which is illustrated in Figure 2.11 as the *codomain* diagram.

Turning back to the wish of modelling unions of balls in a discrete fashion, we already know that the Alpha complex does the job for each snapshot of time as the radius r continuously grows. However, we are also interested in the topological changes during a filtration and so, the question pops up whether the persistences of the Alpha complexes equal the ones of the unions of balls. The answer to this question is "yes". We show this by recalling the commutative diagram from remark 2.2.13/(4). Since homotopy equivalent spaces have isomorphic homology groups, for $r \leq R$ we obtain the new commutative diagram

$$\begin{tikzcd}
H_k(\bigcup_{c \in C} B_c(r)) \arrow[r] \arrow[d, "\cong"'] & H_k(\bigcup_{c \in C} B_c(R)) \arrow[d, "\cong"] \\
H_k(\underline{\mathrm{Alpha}(r, C)}) \arrow[r] & H_k(\underline{\mathrm{Alpha}(R, C)})
\end{tikzcd}$$

where the horizontal arrows symbolize the respective induced homomorphisms. The desired result follows now from the Persistence Equivalence Theorem stated next.

Theorem 2.4.13 (Persistence Equivalence)**.** *Let $(V_i)_{i=0}^n$ and $(W_i)_{i=0}^n$ be two persistence modules and let $(\psi_i)_{i=0}^n$ be a collection of isomorphisms $\psi_i : V_i \to W_i$ such that the diagram*

$$\begin{tikzcd}
V_0 \arrow[r] \arrow[d, "\cong"'] & V_1 \arrow[r] \arrow[d, "\cong"] & \cdots \arrow[r] & V_{n-1} \arrow[r] \arrow[d, "\cong"] & V_n \arrow[d, "\cong"] \\
W_0 \arrow[r] & W_1 \arrow[r] & \cdots \arrow[r] & W_{n-1} \arrow[r] & W_n
\end{tikzcd}$$

commutes. Then the persistence diagram defined by $(V_i)_{i=0}^n$ is the same as the one defined by $(W_i)_{i=0}^n$.

The last topic we want to touch upon in this section is about *stability.* Most of the time, data does not consist of true values but of slightly different ones due to uncertainties in measurements or rounding errors. Thus, it is essential to know whether pertubations of data points change outcomes significantly or not.
This important question also arises for the concept of persistence and it turns out that

it is stable under such small changes in the data. For further details on this we refer to [16].

2.4.1. Connection to Minimum Spanning Trees

In the appendix of [9] a connection between the 1-norms of 0^{th} persistence diagram and the length of Euclidean minimum spanning trees is explained, which plays a crucial role for providing theoretical proofs of some of the equations in chapter 4. Thus, we briefly want to mention what a minimum spanning tree is and, later on in section 2.7.2, what we know about its expected length in a probabilistic setting in particular.

Definition 2.4.14. Given a graph $G = (V, E)$ with vertex set V and edge set E. We define a *spanning tree* $G_T = (V_T, E_T)$ to be a tree and a subgraph of G such that $V_T = V$, which means that G_T spans over all vertices in G.

Corollary 2.4.15. *Every connected graph $G = (V, E)$ contains a spanning tree.*

A short proof of this statement can be found in [6].

Definition 2.4.16. A graph $G = (V, E)$ is called *weighted* if there exists a function $\mathcal{W} : E \to \mathbb{R}$ assigning each edge $e \in E$ a *weight* $\mathcal{W}(e)$. We then write $G = (V, E; \mathcal{W})$ and call $\sum_{e \in E} \mathcal{W}(e)$ its *total weight*.

Definition 2.4.17. Let $G = (V, E; \mathcal{W})$ be a connected, weighted graph. A *minimum spanning tree of* G is a spanning tree $G_{MST} = (V, E_{MST})$ that has minimal total weight, so

$$\sum_{e \in E_{MST}} \mathcal{W}(e) = \min\{\sum_{e \in E_T} \mathcal{W}(e) \mid T \text{ spanning tree of } G\}$$

In the following we will restrict our attention to the special case of a *Euclidean minimum spanning tree* which is a minimum spanning tree of a weighted graph $G = (V, E; \mathcal{W})$ with $V \subseteq \mathbb{R}^d$ and

$$\mathcal{W} : E \to \mathbb{R}_0^+, \quad \mathcal{W}(e) = \|v_{e,1} - v_{e,2}\|_2 \tag{2.6}$$

being the Euclidean distance between the two endpoints $v_{e,1}, v_{e,2} \in V$ of $e \in E$. In this setting, we write $|\mathrm{EMST}(V)|$ for its total weight and say that this is the *length of the Euclidean minimum spanning tree*.

Proposition 2.4.18. *Let $C \in \mathbb{R}^d$ be a finite point set and take the radius function $\mathrm{Rad} : \mathrm{Del}(C) \to \mathbb{R}_0^+$ to produce the Alpha complex filtration of $\mathrm{Del}(C)$. We then obtain for the 1-norm of the 0^{th} persistence diagram the equality*

$$\|\mathrm{Dgm}_0(\mathrm{Rad})\|_1 = \frac{1}{2}|\mathrm{EMST}(C)| + \delta,$$

where δ denotes the cutoff we fix for computing the 1-norm.

Proof. The construction of the Delaunay complex produces a connected graph as 1-skeleton, whose edges can be equipped with the radius they map to through the radius function. For critical 1-simplices this radius is always half their Euclidean length which easily follows from definition 2.2.14. Moreover, the connected components in the filtration all have the same birth time $r = 0$, where they are only vertices, and they die the moment they are first connected by an edge. The persistence of a connected component, hence, is half the Euclidean length of the death-giving edge[6].
On the other hand, the Euclidean minimum spanning tree of C is clearly given by all the death-giving edges we encounter in the filtration. Summing up all persistences yields the length of $\mathrm{EMST}(C)$ plus the cutoff we chose beforehand, that serves as an alternative for the infinite persistence of the one connected component that lives forever. □

2.5. From Black to Colorful: Chromatic Alpha Complexes

Up to now, we have successfully constructed a way to check point clouds for topological features like holes or voids. With persistence diagrams we can extract these information in a compact form and interpret the initial data set. All of this is well-known in the field of topological data analysis and there are many great examples exploiting this technique. Next, we want to study point clouds with two or more colors, each of which may correspond to a specified attribute, so that we can distinguish different kinds of points. We will start by defining chromatic versions of the Delaunay and Alpha complexes. All definitions and constructions in this section can also be looked up in [8] as well as in [5], where these more general concepts beyond two colors were first established.

Definition 2.5.1. Let C be a finite set of points in $\mathbb{R}^d$. A *chromatic point set* is a map

$$\chi : C \to \omega$$

where $\omega = \{0, 1, \dots, s\}$ is a set of $(s+1)$-many *colors* ($s \in \mathbb{N}_0$). For a fixed color $j \in \omega$, the preimage $C_j = \chi^{-1}(\{j\})$ is the subset of C containing all j-colored points.

Before proceeding with the next definition, let us mention here that the notions of C, ω, χ and C_j stay fixed in the following if not stated otherwise.

Definition 2.5.2. Let $\chi : C \to \omega$ be a chromatic point set. The *chromatic Voronoi tessellation* is

$$\mathrm{Vor}(\chi) = \bigcup_{j \in \omega} \mathrm{Vor}(C_j) = \{V(c, C_{\chi(c)})\}_{c \in C}.$$

Remark 2.5.3. Recall that a Voronoi tessellation is always a partition of $\mathbb{R}^d$. In the chromatic setting, however, overlappings of the interiors of differently colored Voronoi domains may happen (see Figure 2.8).

[6] Death-giving edges are part of critical edges.

Definition 2.5.4. Given a chromatic point set $\chi : C \to \omega$, define the *chromatic Delaunay complex of* χ to be

$$\mathrm{Del}(\chi) = \{\alpha \subseteq C \mid \bigcap_{c \in \alpha} V(c, C_{\chi(c)}) \neq \emptyset\}.$$

Remark 2.5.5. Like with the Delaunay complex, its chromatic variant also is an abstract simplicial complex and by its construction

$$\mathrm{Del}(\chi) \cong \mathrm{Nrv}(\mathrm{Vor}(\chi)).$$

Definition 2.5.6. An ω*-stack* (or short: *stack*) in $\mathbb{R}^d$ is a family $\{S_j\}_{j \in \omega}$ of $(s+1)$-many concentric $(d-1)$-spheres, where for each color in ω there is one. The *center* of a stack is the common center of all S_j and its radius is defined to be the largest radius out of all S_j.
We say that a stack is *empty* if, for any color $j \in \omega$, S_j is empty of points in C_j in the sense of definition 2.2.10. Lastly, a stack *passes through* some $\alpha \subseteq C$ if, for any color $j \in \omega$, the sphere S_j passes through the points in $\alpha \cap C_j$. On the right of Figure 2.8, an empty stack is depicted which passes through four points.

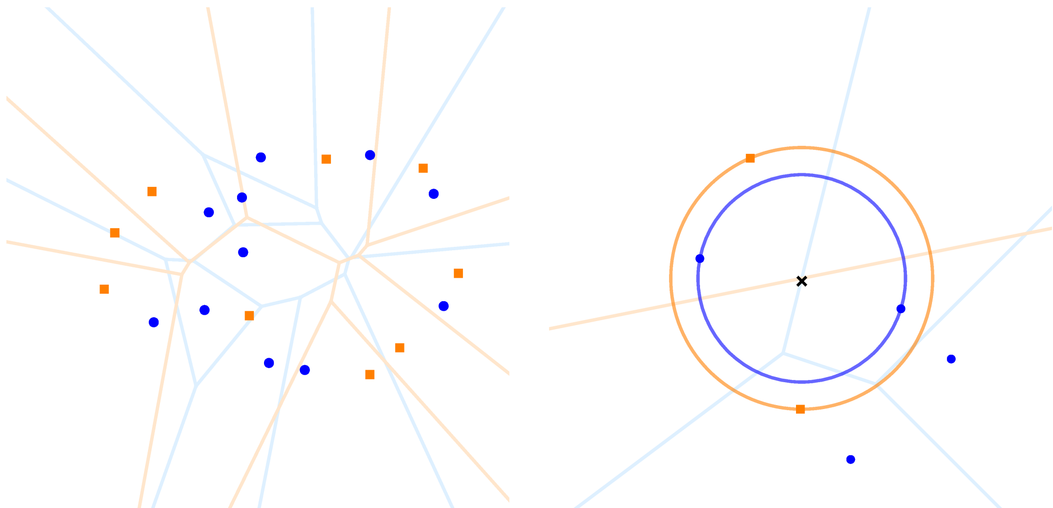

Figure 2.8.: On the *left* the chromatic Voronoi tessellation for a 2-colored point set is illustrated and on the *right* an empty stack through two *blue* points and two *orange* points is shown. In fact, this empty stack of four points is even the *smallest* one there is by definition 2.5.8.

Lemma 2.5.7. *Let $\chi : C \to \omega$ be a chromatic point set in $\mathbb{R}^d$ and let $\alpha \subseteq C$. The following are equivalent:*

(i) $\alpha \in \mathrm{Del}(\chi)$, meaning that α is a simplex of the chromatic Delaunay complex.

(ii) There exists an empty ω-stack $\{S_j\}_{j \in \omega}$ that passes through α.

Proof. For $\alpha \in \mathrm{Del}(\chi)$ we get $\bigcap_{c\in\alpha} V(c, C_{\chi(c)}) \neq \emptyset$. In particular, this means that (i) is equivalent to

$$\bigcap_{c\in\alpha\cap C_j} V(c, C_j) \neq \emptyset, \quad \forall j \in \omega.$$

By lemma 2.2.11, for each $j \in \omega$ there exists an empty $(d-1)$-sphere S_j that passes through $\alpha \cap C_j$, which yields (i) $\Leftrightarrow$ (ii). □

Similarly to the non-colored case, chromatic Alpha complexes are established, but this time we initially define them through a radius function.

Definition 2.5.8. Given a chromatic point set $\chi : C \to \omega$, we denote by $\{S_{j,\alpha}\}_{j\in\omega}$ the *smallest empty stack* that passes through $\alpha \in \mathrm{Del}(\chi)$ and write $r_\alpha \geq 0$ for its radius. The *chromatic radius function on* $\mathrm{Del}(\chi)$ then is

$$\mathrm{Rad}_\chi : \mathrm{Del}(\chi) \to \mathbb{R}_0^+, \quad \mathrm{Rad}_\chi(\alpha) = r_\alpha.$$

Definition 2.5.9. Let $\chi : C \to \omega$ be a chromatic point set in $\mathbb{R}^d$. We call

$$\mathrm{Alpha}(r, \chi) = \mathrm{Rad}_\chi^{-1}([0, r])$$

the *chromatic Alpha complex of* χ *and* r.

Remark 2.5.10. From our definition of a chromatic Alpha complex we immediately see its potential for a filtration of the chromatic Delaunay complex. Indeed, for a fixed chromatic point set $\chi : C \to \omega$ and $r \geq 0$, $\mathrm{Alpha}(r, \chi)$ is an abstract simplicial complex as $\alpha \in \mathrm{Alpha}(r, \chi)$ implies $\alpha \in \mathrm{Del}(\chi)$ with radius of the smallest empty stack $\{S_{j,\alpha}\}_{j\in\omega}$ being at most r. Then any $\emptyset \neq \beta \subseteq \alpha$ is an element of $\mathrm{Del}(\chi)$ as well and, obviously, $\{S_{j,\beta}\}_{j\in\omega}$ cannot have radius bigger than r, therefore $\beta \in \mathrm{Alpha}(r, \chi)$. To see that $\big(\mathrm{Alpha}(r, \chi)\big)_{r\geq 0}$ defines a filtration of $\mathrm{Del}(r, \chi)$ notice that Rad_χ is a monotonic function by the argument in the previous sentence. In fact, Rad is even generalized discrete Morse like its non-chromatic counterpart, given a chromatic version of general position. For the quite technical proof, see [8].

Another immediate consequence of this definition is formulated in the below lemma.

Lemma 2.5.11. *Given a chromatic point set* $\chi : C \to \omega$ *in* $\mathbb{R}^d$ *and take a subset of colors* $\omega' \subseteq \omega$.

(i) Let $\eta : \omega \to \omega'$ *be a merging of colors. Then*

$$\mathrm{Alpha}(r, \eta \circ \chi) \subseteq \mathrm{Alpha}(r, \chi), \quad \forall r \geq 0.$$

(ii) Let $\chi' : \chi^{-1}(\omega') \to \omega'$ *be the restriction of* χ *to colors in* ω'. *Then*

$$\mathrm{Alpha}(r, \chi') \subseteq \mathrm{Alpha}(r, \chi), \quad \forall r \geq 0.$$

Proof. (i) Take $\alpha \in \mathrm{Alpha}(r, \eta \circ \chi)$. We obtain its smallest empty ω'-stack $\{S_{i,\alpha}\}_{i\in\omega'}$ with radius at most r. Then $\{S_{\eta(j)}\}_{j\in\omega}$ is an empty ω-stack of α with radius at most r, so the radius of the smallest such empty ω-stack of α cannot exceed r. Therefore, $\alpha \in \mathrm{Alpha}(r, \chi)$.
(ii) Take $\alpha \in \mathrm{Alpha}(r, \omega')$. Consider again its smallest empty ω'-stack $\{S_{i,\alpha}\}_{i\in\omega'}$ with radius at most r. Then we can build an empty ω-stack with radius at most r by adding $(d-1)$-spheres for each color in $\omega \setminus \omega'$ with zero radius. Therefore, $\alpha \in \mathrm{Alpha}(r, \chi)$. $\square$

Remark 2.5.12. Notice that lemma 2.5.11 yields

$$\mathrm{Alpha}(r, C) \subseteq \mathrm{Alpha}(r, \chi) \text{ and } \mathrm{Alpha}(r, C_j) \subseteq \mathrm{Alpha}(r, \chi) \quad \forall j \in \omega$$

by considering the maps $\eta : \omega \to \{0\}$ and $\chi'_j : C_j \to \{j\}$, respectively.

By now, we are aware of the "filtration-type" nature of $\mathrm{Alpha}(r, \chi)$, but equally important to us is wanting to view it as a discrete space homotopy equivalent to the union of balls emanating from the points in C that grow as r increases. Recall that we even defined Alpha complexes this way in section 2.2.
However, we must be careful as with colors there comes more structure and information that we want to preserve. In fact, we are not only interested in the individual differently colored topological spaces but also how they are in relation to one another.

Definition 2.5.13. Let $\chi : C \to \omega$ be a chromatic point set in $\mathbb{R}^d$. We define the *chromatic clipped Voronoi domain at $c \in C$ of radius $r \geq 0$* as

$$B_c^V(r, \chi) = B_c(r) \cap V(c, C_{\chi(c)}).$$

Lemma 2.5.14. *Given a chromatic point set $\chi : C \to \omega$ and $r \geq 0$, then we have*

$$\mathrm{Alpha}(r, \chi) = \{\alpha \subseteq C \mid \bigcap_{c\in\alpha} B_c^V(r, \chi) \neq \emptyset\}.$$

Proof. $\boxed{\subseteq}$ Let $\alpha \in \mathrm{Alpha}(r, \chi)$, then $\{S_{j,\alpha}\}_{j\in\omega}$ has radius $r_\alpha \leq r$ and, thus, the center of the smallest empty stack has distance at most r_α to any of the points in α, resulting in $x \in \bigcap_{c\in\alpha} B_c(r) \neq \emptyset$. Also, $\alpha \in \mathrm{Del}(\chi)$ which yields $\bigcap_{c\in\alpha} V(c, C_{\chi(c)}) \neq \emptyset$ by definition. In the end we get

$$\bigcap_{c\in\alpha} B_c^V(r, \chi) = \left(\bigcap_{c\in\alpha} B_c(r)\right) \cap \left(\bigcap_{c\in\alpha} V(c, C_{\chi(c)})\right) \neq \emptyset. \tag{2.7}$$

$\boxed{\supseteq}$ Let $\alpha \subseteq C$ such that $\bigcap_{c\in\alpha} B_c^V(r, \chi) \neq \emptyset$. Especially, $\bigcap_{c\in\alpha} V(c, C_{\chi(c)}) \neq \emptyset$ by (2.7), which means that $\alpha \in \mathrm{Del}(\chi)$. Construct an empty stack that passes through α by taking the common center to be $x \in \bigcap_{c\in\alpha} B_c^V(r, \chi)$, then $x \in \bigcap_{c\in\alpha} B_c(r)$, so its radius is at most r. We conclude that its smallest empty stack, therefore, has radius $r_\alpha \leq r$ and $\alpha \in \mathrm{Alpha}(r, \chi)$. $\square$

Remark 2.5.15. All chromatic clipped Voronoi domains are closed and convex and by lemma 2.5.14 it follows

$$\mathrm{Alpha}(r,\chi) \cong \mathrm{Nrv}(\{B_c^V(r,\chi)\}_{c\in C}).$$

Moreover, by the Nerve Theorem 2.1.30

$$\underline{\mathrm{Alpha}(r,\chi)} \simeq \bigcup_{c\in C} B_c^V(r,\chi) = \left(\bigcup_{c\in C} B_c(r)\right) \cap \left(\bigcup_{c\in C} V(c, C_{\chi(c)})\right) = \bigcup_{c\in C} B_c(r)$$

where the last equality is due to the fact that for each color $j \in \omega$ the Voronoi tessellation $\mathrm{Vor}(C_j)$ is a covering $\mathbb{R}^d$.
Finally, we observe that $\mathrm{Alpha}(r,\chi) \simeq \mathrm{Alpha}(r,C)$ by remark 2.2.13/(3).

Because of the usage of different colors, everthing becomes more sophisticated. This is certainly true for the commutative diagram we had encountered first in remark 2.2.13/(4). The following lemma gives the chromatic analogue and a proof can be found in [8].

Lemma 2.5.16. *Let $\chi : C \to \omega$ be a chromatic point set in $\mathbb{R}^d$ and fix some color $j \in \omega$ and some radii $0 \leq r \leq R$. Then we have the commutative diagram*

$$\begin{tikzcd}
\bigcup_{c\in C_j} B_c(r) \arrow[rr, hook] \arrow[dr, hook] \arrow[dd, "\simeq"'] & & \bigcup_{c\in C} B_c(r) \arrow[dr, hook] \arrow[dd, "\simeq"] & \\
& \bigcup_{c\in C_j} B_c(R) \arrow[rr, hook] \arrow[dd, "\simeq"] & & \bigcup_{c\in C} B_c(R) \arrow[dd, "\simeq"] \\
\underline{\mathrm{Alpha}(r,C_j)} \arrow[rr, hook] \arrow[dr, hook] & & \underline{\mathrm{Alpha}(r,\chi)} \arrow[dr, hook] & \\
& \underline{\mathrm{Alpha}(R,C_j)} \arrow[rr, hook] & & \underline{\mathrm{Alpha}(R,\chi)}
\end{tikzcd}$$

2.5.1. Lifting

Computing the chromatic Delaunay complex and, out of this, the chromatic Alpha complexes, first appears to be very complicated, but there is a neat way how to obtain the former which gives another way of viewing chromatic complexes at the same time. The trick we use is to *lift* all points of the same color to an affine subspace of appropriate dimension in a higher-dimensional Euclidean space. More precisely, let us embed the points in C of a chromatic point set $\chi : C \to \omega$ in $\mathbb{R}^d$ into $\mathbb{R}^{d+s} = \mathbb{R}^d \times \mathbb{R}^s$, where $s \in \mathbb{N}_0$ indicates how many colors $\omega = \{0, 1, \dots, s\}$ we have. Now, for each $j \in \omega$ lift the subset $C_j \subseteq C$ to the affine subspace $\mathbb{R}^d \times \{e_j\}$ where e_j denotes the j^{th} unit point in $\mathbb{R}^s$ for $j \neq 0$ (see also equation 2.4) and e_0 represents the origin in $\mathbb{R}^s$. Figure 2.9 shows the lifting of two distinct 2-colored point clouds in $\mathbb{R}^2$ to $\mathbb{R}^3$ that were already illustrated in Figure 2.8. In order to differentiate C_j from its lifted version, we borrow the notation in [8] and use $C_j^{\triangle}$ for the latter as well as $C^{\triangle} = C_0^{\triangle} \sqcup \dots \sqcup C_s^{\triangle}$. Because of the extra dimensions we can just compute the regular Delaunay complex $\mathrm{Del}(C^{\triangle})$ and establish an isomorphism between $\mathrm{Del}(\chi)$ and $\mathrm{Del}(C^{\triangle})$. To see this, recall from lemma 2.2.11 that a simplex lies in the Delaunay complex of $C^{\triangle} \subseteq \mathbb{R}^{d+s}$ if and only if there exists an empty $(d+s-1)$-sphere that passes through every point of the simplex. With the help of this characterization, we build the bridge to the existence of an empty stack of $(s+1)$-many concentric $(d-1)$-spheres of the same simplex projected back to $\mathbb{R}^d$, which yields an element in $\mathrm{Del}(\chi)$. On the right in Figure 2.9 a simple example is drawn to give even more intuition for this correspondence between empty spheres and empty stacks.

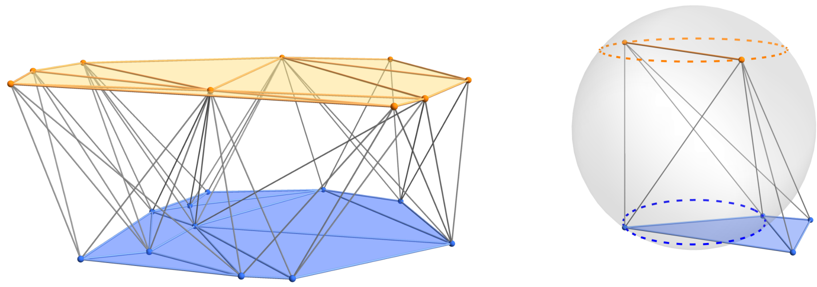

Figure 2.9.: The lifting of the chromatic point clouds of Figure 2.8. On the *right* an empty sphere is drawn that passes through four lifted points and that encapsulates the information of the empty stack as in Figure 2.8 and vice versa. Since four points uniquely determine a 2-sphere, we even get the *smallest* empty one.

2.6. Chromatic Persistent Homology

Persistent homology describes the lifespan of holes within a filtration of some simplicial complex. For colored point clouds it is of further interest to also know how one color affects those persistences of a different color. This is the moment where more than one filtration joins the game. Again, we refer to [8] for further background.

Let $(K_i)_{i=0}^n$ be a filtration of a simplical complex K that arises from a monotonic function $f_K : K \to \mathbb{R}$. Let, furthermore, $L \subseteq K$ be a simplicial subcomplex. Then the restriction $f_L = f_K|_L : L \to \mathbb{R}$ is again monotonic and gives a filtration $(L_i)_{i=0}^n$ of L such that $L_i \subseteq K_i$. From the work done in section 2.4 we obtain two persistence modules

$$\begin{tikzcd}
0 = H_k(L_0) \arrow[r] \arrow[d, "(\jmath_0)_*"'] & H_k(L_1) \arrow[r] \arrow[d, "(\jmath_1)_*"] & \cdots \arrow[r] & H_k(L_{n-1}) \arrow[r] \arrow[d, "(\jmath_{n-1})_*"] & H_k(L_n) = H_k(L) \arrow[d, "(\jmath_n)_*"] \\
0 = H_k(K_0) \arrow[r] & H_k(K_1) \arrow[r] & \cdots \arrow[r] & H_k(K_{n-1}) \arrow[r] & H_k(K_n) = H_k(K)
\end{tikzcd}$$

Notice that the vertical maps $(\jmath_i)_*$ indicate the induced homomorphisms of the inclusions $\jmath_i : L_i \hookrightarrow K_i$. The top one we call the k^{th} *domain persistence module* and the bottom we call the k^{th} *codomain persistence module*. Moreover, we can construct a persistence module out of relative homology groups

$$H_k(K_0, L_0) \longrightarrow H_k(K_1, L_1) \longrightarrow \cdots \longrightarrow H_k(K_{n-1}, L_{n-1}) \longrightarrow H_k(K_n, L_n)$$

which we refer to as the k^{th} *relative persistence module*. From the induced homomorphisms $(\jmath_i)_*$ we further get

$$\begin{array}{ccccccccc}
\ker(\jmath_0)_* & \longrightarrow & \ker(\jmath_1)_* & \longrightarrow & \cdots & \longrightarrow & \ker(\jmath_{n-1})_* & \longrightarrow & \ker(\jmath_n)_* \\
\mathrm{im}(\jmath_0)_* & \longrightarrow & \mathrm{im}(\jmath_1)_* & \longrightarrow & \cdots & \longrightarrow & \mathrm{im}(\jmath_{n-1})_* & \longrightarrow & \mathrm{im}(\jmath_n)_* \\
\mathrm{coker}(\jmath_0)_* & \longrightarrow & \mathrm{coker}(\jmath_1)_* & \longrightarrow & \cdots & \longrightarrow & \mathrm{coker}(\jmath_{n-1})_* & \longrightarrow & \mathrm{coker}(\jmath_n)_*
\end{array}$$

The top sequence is called the *kernel persistence module*, the one in the middle is the *image persistence module* and the bottom one we call the *cokernel persistence module*. In summation, we arrive at six persistence modules and on each we apply the theory of persistent homology to get the six persistence diagrams of kernel, relative, cokernel, domain, image and codomain, which we refer to as the *six-pack*.

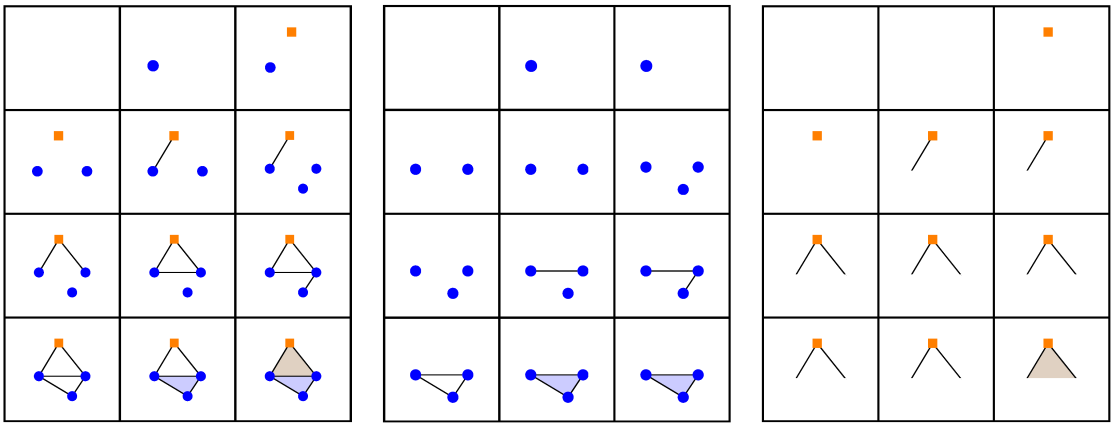

Figure 2.10.: Filtrations induced by f_K (*left*), by f_L (*middle*) and by $f_{K,L}$ (*right*). Notice that the last one is no filtration of a simplicial complex.

Example 2.6.1. Take the filtration of example 2.4.11 but now use colors to indicate which subcomplex $L \subseteq K$ to choose. In this case, let the *blue* points always represent vertices of L and let all other vertices located in $K \setminus L$ be *orange*. By restricting our filtration function f_K to L, we end up with the filtration of the domain. Figure 2.10 shows the colored filtration of the codomain K on the left as well as its restriction on the domain L right next to it. Apart from these two, the other four persistence modules do not necessarily stem from a filtration of a simlicial complex. Thus, it is not so simple to imagine what happens for the other ones. In the hope of giving a good understanding of the relative case in particular, we nevertheless draw a filtration of a collection of simplices that visualize the situation of $H_k(K_i, L_i)$ for any $i = 0, \dots, 11$. This is depicted on the right in Figure 2.10. Notice that this filtration corresponds to restricting f_K to $K \setminus L$ which we write as $f_{K,L}$.

Let us now go through each of the diagrams in the six-pack separately (see Figure 2.11), starting with the codomain. This is the simplest one since it represents exactly the same information we got for the non-chromatic example in section 2.4. Going to the domain, we see that its persistence diagram comes to no surprise either. In the same fashion as for the codomain we read off the lifespans of k-dimensional holes for $k = 0, 1$. The relative diagram appears to be more interesting. From its filtration we extract one 0-dimensional hole that is born in step 2. Its time of death is not noticeable so easily, on the other hand. In fact, it dies the moment we add the first half-edge to the vertex, making it the boundary of exactly this half-edge. The 1-dimensional hole is even more hidden. Observe in step 6 that we have two half-edges connected by the orange point. Naming the respective edges in the codomain e_1 and e_2, we get that

$$\overline{\partial}_1(\overline{e_1 + e_2}) = \partial(e_1 + e_2) + C_0(L) = 2v_o + v_{b,1} + v_{b,2} + C_0(L) = 2v_o + C_0(L) = C_0(L),$$

where $v_{b,1}, v_{b,2}$ denote the blue boundary vertices of e_1 respective e_2 and where v_o is the shared orange boundary vertex. We conclude that $\overline{e_1 + e_2}$ defines a relative 1-cycle and it gets filled with a half-triangle in step 11. Since this half-triangle has $\overline{e_1 + e_2}$ as its boundary, the 1-dimensional hole dies then.
The image diagram is again easier to explain. The first observation we make is that it cannot have persistences of holes other than the ones that are in the codomain diagram because $\mathrm{im}(\jmath_i)_* \subseteq H_k(K_i)$. Secondly, by the very definition, the image reflects all holes of the domain that are still holes in the codomain. The best way to digest this in our example is to pick the specific blue vertex that gets born at step 3 and dies entering step 7 within the domain. However, when embedding the simplicial complex from step 6 in the domain filtration into the one in the codomain, we notice that it is already connected to the orange vertex. It already has died in the codomain at this point. This results in the death in the image diagram. Since death of a homology class always means becoming trivial in some way, the dead 0-dimensional in the image diagram suddenly pops up in the kernel diagram. Only when merging with the color of its own, this 0-dimensional hole disappears completely – both in the kernel diagram and in the one of the domain since $\ker(\jmath_i)_* \subseteq H_k(L_i)$.
Finally, the cokernel diagram gives all persistences of holes in the codomain that do not appear in the image. In our case, this corresponds to the single orange vertex and the mixed-colored triangle.

Remark 2.6.2. The more complex the point cloud, the harder it gets to still understand what is going on in the relative diagram since we have to compute relative homology. But there is another way how to view it. This is done through the isomorphism we got in theorem 2.3.39. It tells us that we can also look at the singular homology group of the quotient $H_k(|K_i|/|L_i|)$ to determine $H_k(K_i, L_i)$.

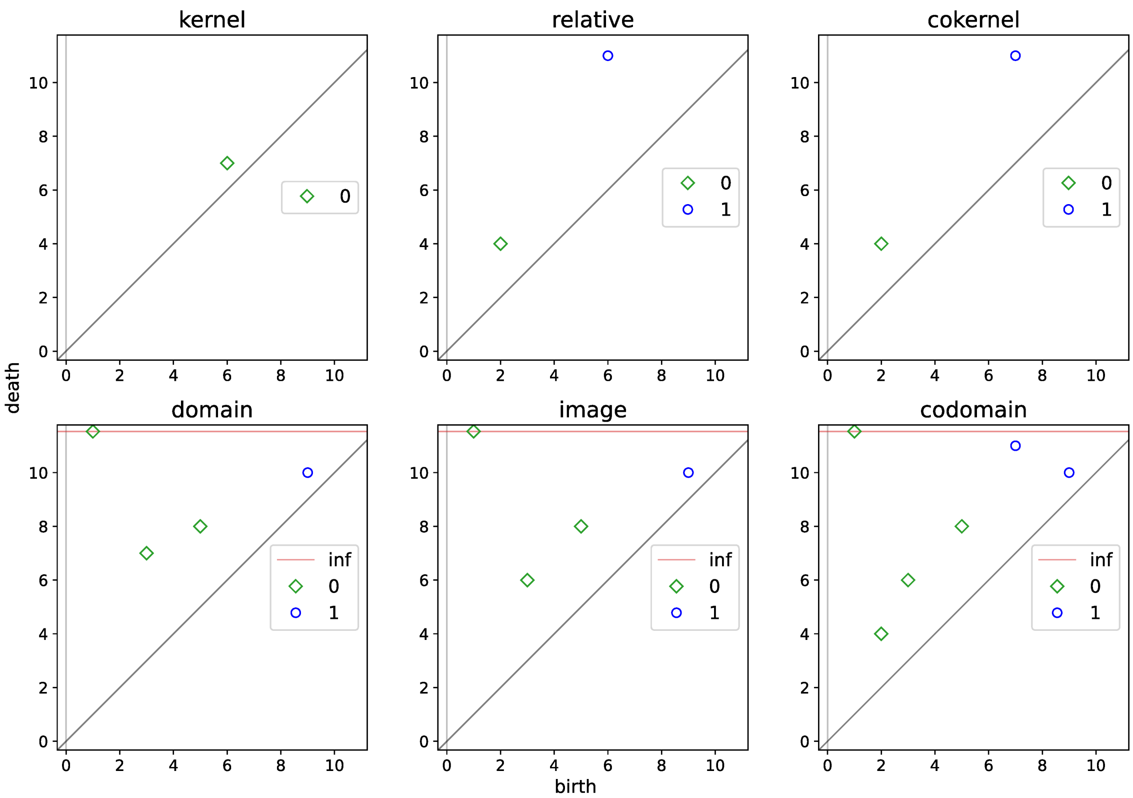

Figure 2.11.: The six-pack of a simple chromatic point cloud with filtrations of codomain K and domain L like in Figure 2.11.

Recall the inclusion maps $\jmath_i : L_i \hookrightarrow K_i$, then $(\jmath_i)_* : H_k(L_i) \to H_k(K_i)$ denoted the induced homomorphism for a fixed $k \in \mathbb{Z}$. Now we want to vary this parameter in order to reveal relations between the different diagrams, so we rename those maps to $\jmath_{i,k}$ for the sake of readability.

Lemma 2.6.3. *Let $(K_i)_{i=0}^n$ be a filtration of a simplicial complex K with monotonic function $f_K : K \to \mathbb{R}$. Moreover, let $L \subseteq K$ be a simplicial subcomplex and generate a filtration $(L_i)_{i=0}^n$ of L by restricting f_K to L. Then for each $i = 0, \dots, n$ and $k \in \mathbb{Z}$, there exist the short exact sequences*

$$0 \longrightarrow \ker \jmath_{i,k} \longrightarrow H_k(L_i) \longrightarrow \operatorname{im} \jmath_{i,k} \longrightarrow 0 \tag{2.8}$$

$$0 \longrightarrow \operatorname{im} \jmath_{i,k} \longrightarrow H_k(K_i) \longrightarrow \operatorname{coker} \jmath_{i,k} \longrightarrow 0 \tag{2.9}$$

$$0 \longrightarrow \operatorname{coker} \jmath_{i,k} \longrightarrow H_k(K_i, L_i) \longrightarrow \ker \jmath_{i,k-1} \longrightarrow 0 \tag{2.10}$$

Proof. The top sequence (2.8) is obviously exact. For (2.9) in the middle, exactness follows because of $\operatorname{coker} \jmath_{i,k} = H_k(K_i)/\operatorname{im} \jmath_{i,k}$. Let us turn to the last sequence (2.10). Notice that $H_k(K_i)$ is a vector space, so $\operatorname{im} \jmath_{i,k}$ is too, and in particular, it is a free abelian group. By proposition 2.3.34, (2.8) yields $H_k(L_i) \cong \ker \jmath_{i,k} \oplus \mathrm{im} \jmath_{i,k}$ and we can replace $H_k(L_i)$ in the long exact sequence from theorem 2.3.35 by $\ker \jmath_{i,k} \to 0 \to \mathrm{im} \jmath_{i,k}$. The same procedure can be done for (2.9), where we exchange $H_k(K_i)$ in the long exact sequence with $\mathrm{im} \jmath_{i,k} \to 0 \to \mathrm{coker} \jmath_{i,k}$. In the end, this leaves us with

$$\cdots \to \ker \jmath_{i,k} \to 0 \to \mathrm{im} \jmath_{i,k} \to \mathrm{im} \jmath_{i,k} \to 0 \to \mathrm{coker} \jmath_{i,k} \to H_k(K,L) \to \ker \jmath_{i,k-1} \to 0 \to \dots$$

containing the desired last short exact sequence. ☐

Theorem 2.6.4. *Let K be a simplicial complex with simplicial subcomplex $L \subseteq K$. Moreover, let $f_K : K \to \mathbb{R}$ be a monotonic function with f_L respective $f_{K,L}$ being the restrictions of f_K to L respective $K \setminus L$. Fix a cut-off $\delta \geq 0$. Then for any dimension $k \in \mathbb{Z}$ we have the equalities*

$$\|\mathrm{Dgm}_k(f_L)\|_1 = \|\mathrm{Dgm}_k(\ker f_L \to f_K)\|_1 + \|\mathrm{Dgm}_k(\mathrm{im}\, f_L \to f_K)\|_1 \tag{2.11}$$

$$\|\mathrm{Dgm}_k(f_K)\|_1 = \|\mathrm{Dgm}_k(\mathrm{im}\, f_L \to f_K)\|_1 + \|\mathrm{Dgm}_k(\mathrm{coker}\, f_L \to f_K)\|_1 \tag{2.12}$$

$$\|\mathrm{Dgm}_k(f_{K,L})\|_1 = \|\mathrm{Dgm}_k(\mathrm{coker}\, f_L \to f_K)\|_1 + \left\|\mathrm{Dgm}_{k-1}(\ker f_L \to f_K)\right\|_1 \tag{2.13}$$

Proof. We only prove the first equality 2.11. The other two can be shown analogously. Let $r_0 < 0 \leq r_1 \leq r_2 \leq \cdots \leq r_n$ be the values such that $K_i = f_K^{-1}([0, r_i])$. Set the cut-off to be $\delta = r_{n+1} \geq r_n$ and let $L_i = f_L^{-1}([0, r_i])$. Observe that for any $r \in [r_i, r_{i+1})$ we still have $L_i = f_L^{-1}([0, r])$. Therefore, the ranks of $H_k(L_i)$, $\ker \jmath_{i,k}$ and $\mathrm{im}\, \jmath_{i,k}$ all stay constant within such a half-open interval and

$$\|\mathrm{Dgm}_k(f_L)\|_1 = \sum_{i=0}^{n} (r_{i+1} - r_i)\, \mathrm{rank}(H_k(L_i)) \tag{2.14}$$

$$\|\mathrm{Dgm}_k(\ker f_L \to f_K)\|_1 = \sum_{i=0}^{n} (r_{i+1} - r_i)\, \mathrm{rank}(\ker \jmath_{i,k}) \tag{2.15}$$

$$\|\mathrm{Dgm}_k(\mathrm{im} f_L \to f_K)\|_1 = \sum_{i=0}^{n} (r_{i+1} - r_i)\, \mathrm{rank}(\mathrm{im}\, \jmath_{i,k}) \tag{2.16}$$

Because the short exact sequence (2.8) even splits by the proof of lemma 2.6.3, the rank of $H_k(L_i)$ is the sum of the ranks of kernel and image and the result follows. ☐

2.6.1. Filtration of the Chromatic Delaunay Complex

We have thoroughly introduced the concept of chromatic persistent homology and proved some useful connections between the 1-norms of persistence diagrams. In example 2.6.1 we constructed very simple filtrations for K and $L \subseteq K$ and analyzed its six-pack. Now we want to apply our knowledge of chromatic Delaunay and Alpha complexes to this setting. Recall from remark 2.5.10 that $\left(\mathrm{Alpha}(r, \chi)\right)_{r \geq 0}$ defines a filtration of $\mathrm{Del}(\chi)$ induced by the chromatic radius function Rad_χ. Moreover, by remark 2.5.12 we know that for every $r \geq 0$ and $j \in \omega$ we have $\mathrm{Alpha}(r, C_j) \subseteq \mathrm{Alpha}(r, \chi)$, which results in

$\mathrm{Del}(C_j) \subseteq \mathrm{Del}(\chi)$ in the limit as r increases.
Setting $f_K = \mathrm{Rad}_\chi$ and choosing the regular radius function $\mathrm{Rad} : \mathrm{Del}(C_j) \to \mathbb{R}_0^+$ for some $j \in \omega$ to be f_L, we are able to produce a six-pack. Let us quickly go through an example to demonstrate this.

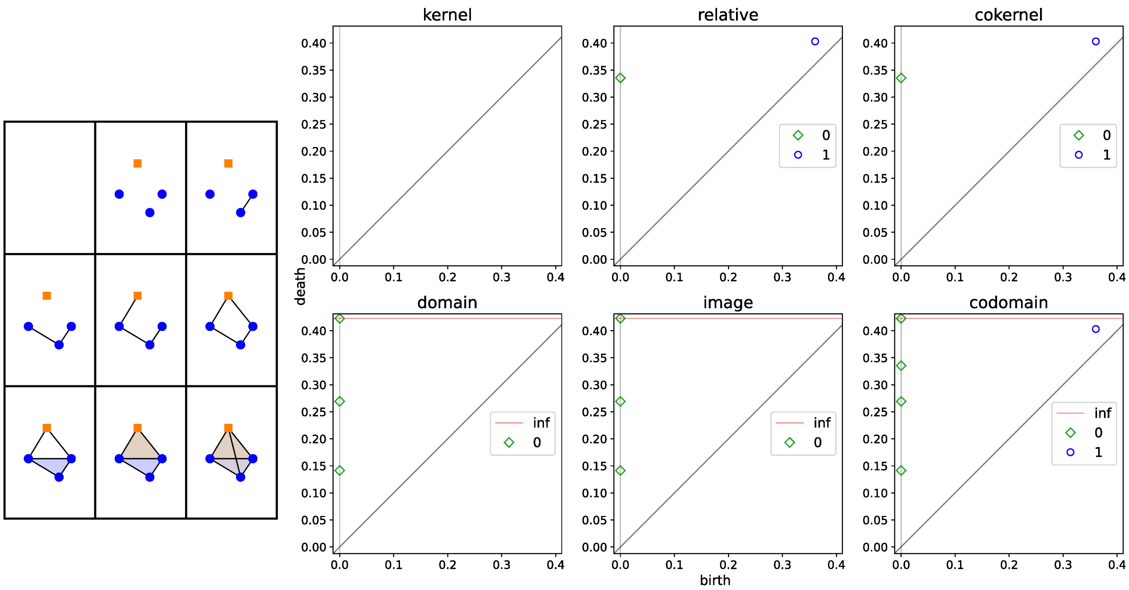

Figure 2.12.: Filtration induced by $f_K = \mathrm{Rad}_\chi$ for a simple 2-colored point cloud (*left*) and the six-pack we get for the choice that L is the Delaunay complex of all blue points $\mathrm{Del}(C_0)$ (*right*).

Example 2.6.5. Take the same chromatic point set as in example 2.6.1 with coordinates $(0.1, 0.3)$, $(0.6, 0.1)$, $(0.8, 0.3)$ and $(0.4, 0.9)$, where the first three are colored in *blue* and the last one in *orange*. Let $K = \mathrm{Del}(\chi)$ and choose $L = \mathrm{Del}(C_0)$ which corresponds to the Delaunay complex of the blue points. The first difference to example 2.6.1 we notice is clearly about the individual filtration steps. Since edges form the moment the respective clipped chromatic Voronoi domains of the boundary points meet, the filtrations in this example depend on Euclidean distances. Thus, it does not come as a surprise that the steps in Figure 2.12 look the way they do. Nevertheless, we want to highlight three special features. First of all, observe that after step 0 all points in step 1 are born at the same "time" $r = 0$, which is best visible in the codomain diagram. This is always the case for filtrations of the (chromatic) Delaunay complex by definition. Secondly, cycles that form the boundary of obtuse triangles are filled in the same instance as the hypotenuse appears. This is an artifact of the underlying geometry.

Finally, in the last step of the codomain filtration a lot happens all at once. A mixed-colored edge appears, which seems to cross over another one. In the non-chromatic setting this would not be possible. However, we have the ability to lift the orange point into 3-dimensional space, where the overlapping dissolves. Moreover, although not that well noticeable in the drawing, simultaneously with the edge two more triangles and the tetrahedron spanned by all four points get born. In the six-pack this happening is not recognizable because the simplicial complexes in step 7 and 8 are homotopy equivalent.

Beware that for a 2-colored chromatic point set in $\mathbb{R}^2$ it may happen that the 2^{nd} persistent homology of the relative filtration is non-trivial, by which we mean that there may appear 2-holes or voids in the respective persistence diagram. A great example of such a case can be found in [8].

2.7. A Spoonful of Poisson Point Processes

The last ingredient we need for the computational part is a bit of probability theory as we want to investigate the six-pack for randomly distributed colored point clouds. In particular, we are interested in stationary Poisson point processes. Since the definition of such one requires some measure theoretical background, we first provide a quick overview of the most necessary terms and follow the content of [27].

Definition 2.7.1. A *measure space* is a triplet (Ω, Σ, μ) consisting of a set Ω, a σ-algebra Σ and a measure $\mu : \Sigma \to [0, \infty]$ on Ω. If μ is defined on the *Borel σ-algebra* $\Sigma = \mathcal{B}(\Omega)$ of Ω, then μ is called a *Borel measure*. If $\mu(\Omega) = 1$, then μ is a *probability measure* and we write P instead of μ. The triplet (Ω, Σ, P) is then a *probability space*. A *measurable space* is merely tuple (Ω, Σ), where Ω is a set, Σ is a σ-algebra for which a measure exists.

Definition 2.7.2. Let $(X, \mathcal{A}), (Y, \mathcal{B})$ be two measurable spaces. A function $f : X \to Y$ is called $(\mathcal{A} - \mathcal{B})$*-measurable* if

$$f^{-1}(B) \in \mathcal{A}, \quad \forall B \in \mathcal{B}.$$

In case it is clear which σ-algebras are meant, we drop the prefix $(\mathcal{A} - \mathcal{B})$ and say that f is *measurable*. If $\mathcal{B}$ equals the Borel σ-algebra $\mathcal{B}(Y)$, we say that f is *Borel-measurable*. In the special case where f has values in $\mathbb{R}$, f is usually named *random variable*.

Definition 2.7.3. Let $(X, \mathcal{A}), (Y, \mathcal{B})$ be two measurable spaces and let $f : X \to Y$ be a measurable function. For every possible measure $\mu : \mathcal{A} \to [0, \infty]$ on $\mathcal{A}$ we define the set function

$$f(\mu) = f_*\mu = \mu \circ f^{-1} : \mathcal{B} \to [0, \infty],$$

which turns out to be a measure on $\mathcal{B}$ and is called the *pushforward measure of μ and f*.

For the remainder of this section, let E denote a locally compact, second countable, Hausdorff topological space and turn it into a measurable space $(E, \mathcal{B}(E))$ with its Borel σ-algebra $\mathcal{B}(E)$.

Definition 2.7.4. We define $M(E)$ to be the set of all Borel measures μ on E that are *locally finite*, which means that $\mu(C) < \infty$ for any compact subset $C \subseteq E$.

Remark 2.7.5. (1) From the above definition and because E is Hausdorff we conclude that any $\mu \in M(E)$ is σ-finite.
(2) We equip $M(E)$ with the σ-algebra $\mathcal{M}(E)$ that is generated by evaluation functions

$$M(E) \to [0, \infty],\ \mu \mapsto \mu(A) \quad (A \in \mathcal{B}(E))$$

to obtain the measurable space $(M(E), \mathcal{M}(E))$.

Definition 2.7.6. We define $N(E)$ to be the set of all counting Borel measures μ on E, which means that $\mu(\mathcal{B}(E)) \subseteq \mathbb{N}_0 \cup \{\infty\}$.

Remark 2.7.7. By definition, such a counting measure is always locally finite, hence $N(E) \subseteq M(E)$ and one can even show that $N(E) \in \mathcal{M}(E)$. Out of this knowledge we construct the σ-algebra $\mathcal{N}(E) = \{N \cap U \mid U \in \mathcal{M}(E)\}$, so that $(N(E), \mathcal{N}(E))$ also becomes a measurable space.

Lemma 2.7.8. *For any $\mu \in N(E)$, there exist measurable functions $\xi_i : N(E) \to E$ $(i = 1, \ldots, \mu(E))$, such that*

$$\mu = \sum_{i=1}^{\mu(E)} \delta_{\xi_i(\mu)}, \tag{2.17}$$

where δ_x denotes the Dirac measure for $x \in E$.

Definition 2.7.9. We define $N_s(E)$ to be the set of all counting Borel measures μ on E that are *simple*, which means that $\mu(\{x\}) \leq 1$ for each $x \in E$.

Definition 2.7.10. Let $\mu \in M(E)$. The *support of* μ is the smallest closed set $A \subseteq E$ such that $\mu(E \setminus A) = 0$ and we write $\operatorname{supp} \mu = A$.

Remark 2.7.11. (1) It can be shown that $N_s(E) \in \mathcal{N}(E)$, so $N_s(E)$ is a measurable subset of $N(E)$. Similarly to remark 2.7.7, we define $\mathcal{N}_s(E)$ to be the trace σ-algebra and $(N_s(E), \mathcal{N}_s(E))$ becomes a measurable space.
(2) Let $\mu \in N(E)$. We then obtain that

$$\operatorname{supp} \mu = \{x \in E \mid \mu(\{x\}) \geq 1\} \subseteq E$$

is a locally finite subset. Going one step further, for simple counting measures we get the non-trivial isomorphism

$$N_s(E) \to \Sigma_{loc}, \quad \mu \mapsto \operatorname{supp} \mu$$

into the set Σ_{loc} of all locally finite subsets of E. In particular, notice that $\mu \in N(E)$ is simple if and only if μ can be written in the form of (2.17) where $\xi_i(\mu) \neq \xi_j(\mu)$ whenever $i \neq j$. Writing $x_i = \xi_i(\mu) \in E$, This yields,

$$\operatorname{supp} \mu = \{x \in E \mid \mu(\{x\}) = 1\} = \{x_i\}_{i=1}^{\mu(E)}$$

which may be a finite or infinite sequence of distinct points in E. Surely, an analogous result is achievable for $\mu \in N(E)$. However, beware that, in this case, the sequence does not necessarily consist of distinct points and we lose the isomorphism.

Definition 2.7.12. A *random measure X on E* is a measurable function $X : \Omega \to M(E)$ from a probability space (Ω, Σ, P) into the measurable space $(M(E), \mathcal{M}(E))$. In this case, we equip $(M(E), \mathcal{M}(E))$ with the pushforward measure of P and X and we call $X(P)$ the *distribution of X*.
If a random measure X on E satisfies $\{X \in N(E)\}$ P-almost surely, then we say that X is a *point process in E*. If such a point process X meets the condition that $\{X \in N_s(E)\}$ holds P-almost surely, we call X a *simple point process*.

Remark 2.7.13. By remark 2.7.11/(2), we make the following observation:
Let X be a point process in E with $\omega \in \{X \in N_s(E)\}$, then $X(\omega)$ can be considered as a simple counting measure or as a finite or infinite sequence in E depending on whether $X(\omega)(E) < \infty$ or $X(\omega)(E) = \infty$. If $X(\omega)$ is simple, the above ways of viewing are equivalent. Now the reason of calling X a *(simple) point process* has a solid justification.

Definition 2.7.14. Let X be a random measure on E. Then the set function

$$\Lambda : \mathcal{B}(E) \to [0, \infty], \quad \Lambda(A) = \mathbb{E}[X(.)(A)]$$

is a measure on E and it is called the *intensity measure of X*.

Definition 2.7.15. Let X be a point process in E with intensity measure Λ. We say that X has *Poisson counting variables* if for every $A \in \mathcal{B}(E)$ such that $\Lambda(A) < \infty$ the random variable $X(.)(A) : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ has a *Poisson distribution*, so

$$P(\{X(.)(A) = k\}) = \mathrm{e}^{-\Lambda(A)} \frac{\Lambda(A)^k}{k!} \quad (k \in \mathbb{N}_0)$$

Definition 2.7.16. Let X be a point process in E. For each choice of pairwise disjoint Borel sets $A_1, \ldots, A_m \in \mathcal{B}(E)$ $(m \in \mathbb{N})$, let the random variables $X(.)(A_1), \ldots, X(.)(A_m)$ be independent. Then we say that X has *independent increments*.

Definition 2.7.17. Let X be a simple point process in E. We say that X is a *Poisson point process* if it has both Poisson counting variables and independent increments.

Corollary 2.7.18. *Let X be a simple point process in E such that X has Poisson counting variables. Then X is a Poisson point process.*

From now on we focus on the special case where $E = \mathbb{R}^d$.

Definition 2.7.19. Let X be a random measure on $\mathbb{R}^d$. Then X is called *stationary* if the distribution of X and its translate $X + x$ coincide for all possible $x \in \mathbb{R}^d$.

Remark 2.7.20. For a stationary random measure X on $\mathbb{R}^d$ with locally finite intensity measure Λ we obtain that Λ has a particularly nice form. Indeed, first observe that, within this setting, Λ is a measure on $\mathbb{R}^d$ which is translation-invariant. The only locally finite and translation-invariant measure on $\mathbb{R}^d$, however, is a multiple of the *Lebesgue measure* $\|.\|$, which yields

$$\Lambda = \lambda \|.\|$$

for some constant $\lambda \in [0, \infty)$ that is refered to as *intensity of* X.

Within the scope of this thesis we exclusively need stationary Poisson point processes in $\mathbb{R}^2$. Since we can do experiments only on a finite data set, we are going to focus our attention on an observation window W inside $\mathbb{R}^2$ and we always choose $W = [0,1]^2$ to be the *unit square*. The question that is unanswered yet is how to simulate a stationary Poisson point process X that is restricted to this region.
Consider the following construction:
First, take a random variable $\nu : \Omega \to \mathbb{N}_0$ equipped with a Poisson distribution with parameter $\Lambda = \lambda \|W\| = \lambda$. We have

$$P(\{\omega \in \Omega \mid \nu(\omega) = k\}) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}$$

and for the outcome $\nu(\omega) = k$, we generate k-many independent random variables $\hat{\xi}_1, \dots \hat{\xi}_k : \Omega \to W$ which are uniformly distributed in W. The point process $\hat{X}$ in W we construct in this way has the property $\hat{X}(\omega)(W) = \nu(\omega)$ and is of the form

$$\hat{X}(\omega)(\hat{A}) = \sum_{i=1}^{\nu(\omega)} \delta_{\hat{\xi}_i(\omega)}(\hat{A})$$

Notice that for $\hat{A} = A \cap W$ with $A \in \mathcal{B}(\mathbb{R}^2)$ we have the equality

$$\begin{aligned} P(\{\omega \in \Omega \mid \hat{X}(\omega)(\hat{A}) = k\}) = \mathrm{e}^{-\lambda\|\hat{A}\|} \frac{(\lambda \left\|\hat{A}\right\|)^k}{k!} &= \mathrm{e}^{-\lambda\|A\cap W\|} \frac{(\lambda \|A \cap W\|)^k}{k!} \\ &= P(\{\omega \in \Omega \mid X(\omega)(A \cap W) = k\}), \end{aligned}$$

so $\hat{X}$ has the same distribution as X restricted to W.

2.7.1. On the Poisson-Delaunay Complex

Taking a realization $X(\omega)$ of a stationary Poisson point process X in $\mathbb{R}^2$ with intensity $\lambda = n$, we can construct its Delaunay complex[7], restrict our region of interest to the unit square $[0,1]^2$ and ask ourselves about the expected number of critical and non-critical k-simplices[8] we have there. However, since we started with an infinitely large and random

[7] Initially, we defined the Delaunay complex for finite point sets only, but it is also possible to build it with an infinite, but locally finite one. For $X(\omega)$, this is the case due to the isomorphism stated in remark 2.7.11/(2).

[8] See section 2.2.1 for the definition of these types of simplices.

Delaunay complex, in this case sometimes called the *Poisson-Delaunay complex*, some simplices get cut off at the boundary of the unit square. In order to deal with this issue, we consider all k-simplices that have their *center* inside $[0,1]^2$, by which we mean the center of their smallest empty sphere. For an interval $[\beta, \alpha]$ this is the common center of all of its elements.

Before stating the result on this question, that was given and proven as Theorem 1 in [19], we need to remind ourselves what the (incomplete) Gamma function is.

Definition 2.7.21. The *Gamma function* Γ is given by the integral

$$\Gamma(a) = \int_0^\infty s^{a-1} e^{-s}\, ds$$

and the *incomplete Gamma function* γ is written as

$$\Gamma(a,x) = \int_0^x s^{a-1} e^{-s}\, ds.$$

Clearly, we have the relation $\gamma(a,x) \to \Gamma(a)$ as $x \to \infty$.

Theorem 2.7.22. *Let X be a stationary Poisson point process with intensity $n > 0$ in $\mathbb{R}^2$ and let $X(\omega)$ be a realization viewed as a locally finite set of points scattered in $\mathbb{R}^2$. For each $R > 0$, the expected number of intervals $[\beta, \alpha]$ of the Poisson-Delaunay complex with center in the unit square $[0,1]^2$ and radius at most R is given by*

$$\frac{\gamma(a, n\pi R^2)}{\Gamma(a)} C_{b,a} n, \tag{2.18}$$

in which $b = \dim(\beta)$, $a = \dim(\alpha)$ and in which $C_{b,a}$ are specific coefficients that take the values

$$C_{0,1} = C_{0,2} = 0, \quad C_{0,0} = C_{1,2} = C_{2,2} = 1, \quad C_{1,1} = 2. \tag{2.19}$$

A direct consequence of theorem 2.7.22 is summarized in the following corollary.

Corollary 2.7.23. *In the same setting as in theorem 2.7.22, view the Poisson-Delaunay complex of intervals with center in the unit square $[0,1]^2$.*

(i) In expectation there are $2n$ critical edges and n non-critical ones.

(ii) In expectation there are n critical triangles and n non-critical ones.

(iii) In expectation out of all critical edges it follows that n of them are death-giving to connected components and n are birth-giving to loops. Moreover, all critical triangles are death-giving to loops.

Proof. We look at the Poisson-Delaunay complex itself, thus we imagine R to go to infinity. Consequently, the incomplete Gamma function cancels and becomes the Gamma function at a and we are left with $C_{b,a}n$. Hence, the number of critical edges is $C_{1,1}n = 2n$ and the number of non-critical ones is $C_{1,2}n = n$, which proves (i). For the non-critical triangles we again have $C_{1,2}n = n$ and $C_{2,2}n = n$ gives the number of critical triangles, which shows (ii).
We are left to prove (iii). For this we construct a bijective correspondence between critical edges that give birth to loops and critical triangles, since every of the latter fills a loop created of the former at some radius. We conclude there are n such birth-giving critical edges, which results in n death-giving critical edges.
Finally, all critical triangles contribute as death-giving ones because we do not have any tetrahedra in the Poisson-Delaunay complex in the plane, so there cannot be any triangles that give birth to tetrahedra. □

2.7.2. The Euclidean Minimum Spanning Tree Problem

Imagine n isolated electricity networks that you want to connect together. You hire a company that lay cables, but of course you want it to be cost-efficient. Thus, you construct a Euclidean minimum spanning tree with the networks forming its vertex set embedded in $\mathbb{R}^2$ and compute its length. In this way, the amount of needed cables is kept to a minimum.
Now we further suppose that these electricity networks are uniformly distributed random variables within the unit square $[0,1]^2$. For each realization of these random variables we most likely obtain a different vertex set V and so, also $|\mathrm{EMST}(V)|$ might vary. For a larger number n of electricity networks, however, the distances to nearest neighbors necessarily decrease and we can ask ourselves whether there exists an asymptotic limit the length of the Euclidean minimum spanning tree tends to.
In 1988 this question was successfully answered by Steele in [28]. He not only showed a specific asymptotic behavior for the Euclidean minimum spanning tree in 2 dimensions but more general in d dimensions. He even considered weight functions like in (2.6) with integer powers $a \in (0,d)$. For the purpose of this thesis, it suffices to consider the case where $d = 2$ and $a = 1$, which we state next.

Theorem 2.7.24. *Let $G = (V, E; \mathcal{W})$ be a complete, weighted graph with $\mathcal{W}$ like in equation (2.6) and where the vertex set V consists of the realization of n independent and uniformly distributed random variables $\hat{\xi}_1, \dots, \hat{\xi}_n$ in the unit square $[0,1]^2 \subseteq \mathbb{R}^2$. Then with probability 1 we obtain that*

$$\lim_{n\to\infty} \left(\frac{|\mathrm{EMST}(V)|}{\sqrt{n}} \right) = c \tag{2.20}$$

where $c > 0$ is a constant. Furthermore, we even obtain the same behavior

$$\lim_{n\to\infty} \left(\frac{\mathbb{E}[|\mathrm{EMST}(V)|]}{\sqrt{n}} \right) = c \tag{2.21}$$

for the expected length of the Euclidean minimum spanning tree.

Up until now, nobody has derived an explicit value of c, although there was some partial success in developing a formula for computing this constant by Avram and Bertsimas [1]. Unfortunately, this formula contains a series expansion whose summands are increasingly harder to calculate analytically. Nevertheless, Avram and Bertsimas achieved the current best lower bound by approximating the series up to the fifth summand and we have

$$0.6008 \leq c \leq \frac{\sqrt{2}}{2} \approx 0.7072, \tag{2.22}$$

where the best upper bound is due to a theoretical argument of Gilbert [21].

3. Computational Setup

In this chapter we explain the computational pipeline that was needed for the experimental investigation of both 0- and 1-norm of the six-pack on a random 2-colored point cloud in the unit square with varying number of points (n) and varying probability of the colors (t). By *random* we mean a stationary Poisson point process $X : \Omega \to N_s(\mathbb{R}^2)$ restricted to the observation window $W = [0, 1]^2$, which we can model as described in section 2.7.

The intensity we use is the number of points inside the unit square, so $\lambda = n$. Since we do many experiments and are only interested in the average cases of the computed data, we further simplify the situation as we fix the number of points to be n, instead of introducing a random variable $\nu : \Omega \to \mathbb{N}_0$ that in expectation gives us the outcome n.

Dealing with randomness always leads to large amounts of data because each experiment has its own unique point cloud structure that is obtained by a uniform distribution and that we like to store somehow, in order to always be able to connect computational outcomes with a visualization of the points' positions in W. We use *pseudo-randomly generated seeds* for this matter. Each experiment is assigned such a seed, which is a uniformly chosen integer between 0 and 2^{30}, and it serves as a key for reconstructing the point cloud to the corresponding experiment. We cannot assure that no two experiments have the same seed and, therefore, have the same point clouds. However, we argue that this possibility does not occur within our experiment series as its probability is 2^{-30}, hence effectively 0. Algorithm 1 shows the construction of a list of seeds *seed_list*, which we feed into the main algorithm 3 as one of many input parameters.

Algorithm 1 `create_seedlist`

Require: *number_exps*
Ensure: Creation of *seed_list*
  seed_list $\leftarrow$ create list of (*number_exps*)-many random seeds lying between 0 and 2^{30}

The reason why we first construct all seeds and only then run the individual experiments, depending on these seeds, is because we also want to compare different choices of filtration settings with each other. Recall from section 2.6 that we can fix a filtration by choosing a simplicial complex K and a subcomplex $L \subseteq K$. In this thesis, K always refers to the chromatic Delaunay complex $K = \mathrm{Del}(\chi)$. The mapping $\chi : C \to \{0, 1\}$ refers to the random chromatic point set that is constructed by equipping each random point in C (that we get out of our simulated and simplified Poisson point process) with the label 0 or 1, which should always be understood as the colors *blue* or *orange* in this context. For the subcomplex L, three obvious choices appear in this framework: (i) $L = \mathrm{Del}(C_0)$, (ii) $L = \mathrm{Del}(C_1)$ or (iii) $L = \mathrm{Del}(C_0) \sqcup \mathrm{Del}(C_1)$.

The first two cases are symmetric to each other, so we choose (i) to work with, which is the subcomplex of $\mathrm{Del}(\chi)$ consisting of all blue simplices. The third case is interesting on its own again. There, L contains all simplices that are of exactly one color. In other words, a simplex lies in L if it is *either* blue *or* orange. This is why we sometimes call choice (iii) the *mono-chromatic* case, even though the vertex set of L is far from being one-colored!
Coming back to our list of seeds, we construct random 2-colored point clouds in the unit square with the fixed parameters n and t and still be flexible with our choice on L. Looking at Algorithm 3, we pick *filtration* to be one of the cases (i) - (iii) ahead of any computation, but since we generated *seed_list* beforehand, we may apply Algorithm 3 multiple times – one for each choice of *filtration*.

The main purpose of Algorithm 3 is to build a dictionary with different keys, each of which becoming the column of a *.csv* data frame structure in the last step within Algorithm 4. Notice that we always set $\mathit{number\_exps} = 100$, $\mathit{number\_colors} = 2$ and $\mathit{spatial\_dim} = 2$. Moreover, $\mathit{norms} = [0, 1]$ and $\mathit{diagrams} =$ [“kernel”, “relative”, “cokernel”, “domain”, “image”, “codomain”] since we want to have 0- and 1-norms for all six persistence diagrams. For better readability, we further set $\mathit{number\_pts} = n$ and $\mathit{prob\_colors} = [t, 1 - t]$.

Algorithm 2 `add_value_to_dict`

Require: *dict*, *key*, *val*, *current_step*, *total_steps*
Ensure: Add the value *val* to position *current_step* in list with (*total_steps*)-many elements that lies inside the dictionary *dict* within the key *key*

```
if key is not in dict then
    dict[key] ← create list with (total_steps-1)-many 0's and the value val at position
                current_step
else
    dict[key][current_step] ← val
end if
```

At first, we create the key “seed” and use the method `add_value_to_dict` (see Algorithm 2) to do so. In lines 7 and 8 in Algorithm 3 we construct for each experiment pair (*exp*, *seed*) a random 2-colored chromatic point cloud in $[0, 1]^2$. The next lines of code are responsible for computing the filtration functions of the six-pack and all their persistence pairs. For this, we use the free PyPI package `chromatic_tda` (version 1.1.6). For further background on this package and a great manual for first use, see [12]. With the help of the last three for-loops we get to the bottom of calculating the 0- and 1-norm for each k^{th} persistence diagram of the six-pack where k can go up to 2 for $\mathit{spatial\_dim} = 2$.

In the end, we obtain the dictionary *exps_dict*, which we pour into the shape of a data frame structure in Algorithm 4.

Algorithm 3 `create_exps_dict`

Require: *number_exps*, *number_pts*, *number_colors*, *prob_colors*, *filtration*, *spatial_dim*, *diagrams*, *norms*, *seed_list*

Ensure: Creation of *exps_dict*

```
1: exps_dict ← {}
2: experiments ← list of integers with range number_exps
3:
4: for (exp, seed) in zip(experiments, seed_list) do
5:     activate random seed seed
6:     add_value_to_dict(exps_dict, “seed”, seed, exp, number_exps)
7:     points ← list of (number_pts)-many randomly chosen coordinates with respect to
           uniform distribution inside [0, 1]^spatial_dim
8:     labels ← list of (number_pts)-many elements, each being an integer between 0 and
           number_colors with weight prob_colors
9:
10:    /* From line 11 to 16 we work with methods from the package chromatic_tda */
11:    alpha_complex ← compute alpha complex with input points, labels
12:    simplicial_complex ← get filtration simplicial complexes out of alpha_complex with
           input filtration
13:    compute persistence of simplicial_complex
14:
15:    for dgm in diagrams do
16:        persistence_dict ← dictionary with possible dimensions k as keys and all k^th
               persistence pairs with finite values as corresponding values
17:        k_list ← list of keys of persistence_dict
18:
19:        for k in k_list do
20:            persistence_array ← array of persistences of k^th persistence pairs
21:
22:            for p in norms do
23:                p_norm ← array with p^th norm for each element in persistence_array
24:                add_value_to_dict(exps_dict, “dgm_k-dim_p-norm”, p_norm, exp,
                       number_exps)
25:            end for
26:        end for
27:    end for
28: end for
29: exps_dict ← dictionary with all the data
```

Algorithm 4 `create_csv`

Require: *number_exps*, *number_pts*, *number_colors*, *prob_colors*, *filtration*, *spatial_dim*, *exps_dict*, *filepath*
Ensure: Saving the data in *exps_dict* as .csv file
dataframe ← create data frame structure with content of *exps_dict*

/∗ The method `filename_from_input` is used for constructing a string with all the important information inside to distinguish the different experiment series and their data ∗/
filename ← `filename_from_input`(*number_exps*, *number_pts*, *number_colors*, *prob_colors*, *filtration*, *spatial_dim*)
Save *dataframe* as *filename*.csv with file path *filepath*

In Table 3.1 we illustrate an example of such a data frame of the configuration with 100 experiments, two equally likely colors, 1000 points in the unit square and in the mono-chromatic case. Apart from the first two, all column captions are of the form "*dgm_k*-dim_*p*-norm" like we have it in Algorithm 3, line 24.

exp	seed	kernel_0-dim_0-norm	kernel_0-dim_1-norm	...	codomain_1-dim_1-norm
1	515938188	491.0	4.4562...	...	2.1021...
2	362593555	510.0	4.4936...	...	2.0135...
3	802025811	541.0	4.5496...	...	1.9995...
⋮	⋮	⋮	⋮	⋮	⋮
98	633635510	498.0	4.2411...	...	1.9617...
99	642053118	520.0	4.4501...	...	2.0066...
100	489887328	500.0	4.5637...	...	1.9787...

Table 3.1.: An excerpt of the file *exp100_n1000_col2_prob50-50_mono_dim2.csv*

Remark 3.1. Since we work with many different parameters, let us fix the notation of a triple (n, t, L) which defines the setting we work in, where n indicates the numbers of points, t gives the probability that we color a point blue[1] and L represents the type of filtration, just like before.

[1]Abstractly speaking, this corresponds to color 0 for $\omega = \{0, 1\}$.

4. Results

Recall from chapter 3 that for each of the 100 experiments, we are given a random 2-colored point set $\chi : C \to \omega$ inside the unit square $[0,1]^2$, where we use a uniform distribution to simulate the realizations of a Poisson point process within our observation window. For a specific combination (n,t,L) we generate the persistence pairs of all k^{th} persistence diagrams of the six-pack, we apply the 0- and 1-norm, respectively, and calculate their means.

Remark 4.1. For obtaining reasonable values with the 1-norm, we only take finite persistence pairs into consideration and drop the usage of any cutoff. For better comparability, we further use no infinite persistence pairs for the computation of the 0-norm as well. This different way of viewing, however, does not affect the validity of the three equalities we get in theorem 2.6.4.

To give a concrete example of the outcomes, let us turn to Figure 4.1. It shows two graphics, both for the 0^{th} persistence diagram of the domain L. We plot the number of points against the average 0-norm on the *left* and the average 1-norm on the *right* for three different values of t and for the two interesting cases we have for L, which are $L = \mathrm{Del}(C_0)$ (abbrev. "0-filt") and $L = \mathrm{Del}(C_0) \sqcup \mathrm{Del}(C_1)$ (abbrev. "mono-filt"). For better visualization, the settings with the former choice of L are always depicted by *empty markers*, whereas the ones of the latter are given by *solid markers*.

We extract two major observations out of these plots now: firstly, for each choice of L we have the same behavior of the means for increasing numbers of points and, secondly, this behavior appears to be that of a linear function for the 0-norm and that of a square root function for the 1-norm. In fact, these observations stay true for the 1^{st} persistence diagrams of the six-pack as well, and all these plots can be looked up in appendix A.1. Coming back to Figure 4.1, we further highlight that we get similar curves for different probability distributions and in the following, we are going to state and prove explicit equations for the expected 0- and 1-norms for some k^{th} persistence diagrams in dependence of n and t. Moreover, the theoretical results will be supported by our experimental data.

Since we work in the plane, the interesting cases appear for $k = 0, 1, 2$, where $k = 2$ only affects the relative filtration $f_{L,K}$.[1] However, we are going to focus on $k = 0$ (corresponding to connected components) and $k = 1$ (corresponding to loops) only.

[1] See also the last paragraph in section 2.6.

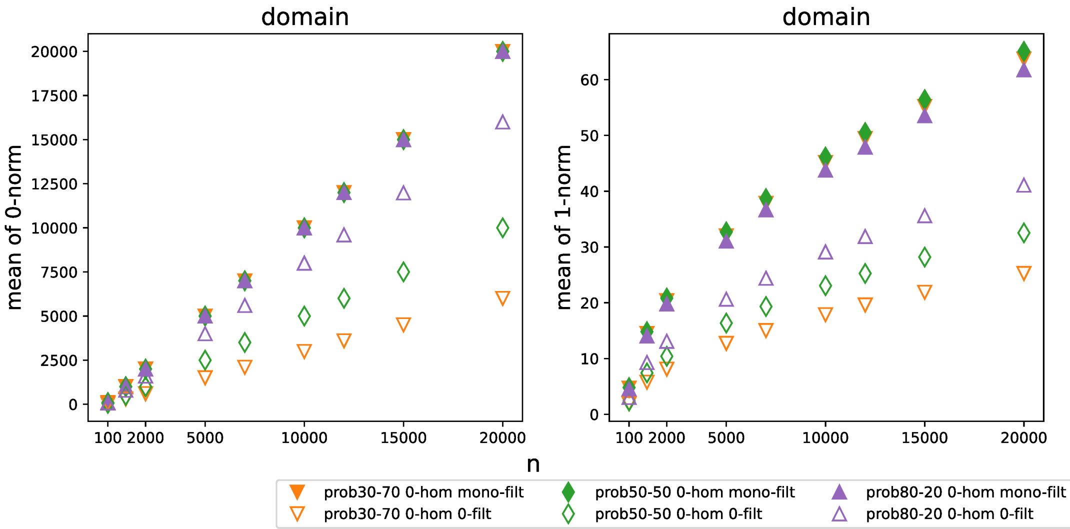

Figure 4.1.: Different developments for means of the 0-norm (*left*) and 1-norm (*right*) in 0^{th} homology for the domain in the cases of "0-filt" and "mono-filt" and for three probability distributions.

At first, we state and prove a proposition similar to theorem 2.7.22, in which we derive a formula for the expected length of intervals, that is the expected sum of the intervals' radii. This will become important later on when showing expected 1-norms in 1^{st} persistence homology. Finally, let us state here that the proofs of proposition 4.2 as well as the ones of proposition 4.6 and 4.10 (for the special case $t = 1 - t = \frac{1}{2}$) are going to be included in an upcoming paper that is the joint work of Draganov, Edelsbrunner, Saghafian and the author of this thesis. [13]

Proposition 4.2. *Let X be a stationary Poisson point process with intensity $n > 0$ in $\mathbb{R}^2$ and let $X(\omega)$ be a realization viewed as a locally finite set of points scattered in $\mathbb{R}^2$. For each $R > 0$, the expected length of intervals $[\beta, \alpha]$ of the Poisson-Delaunay complex with center in the unit square $[0,1]^2$ and radius at most R is given by*

$$\frac{\gamma(a + \frac{1}{2}, n\pi R^2)}{\Gamma(a)} \frac{C_{b,a}}{\sqrt{\pi}} \sqrt{n}, \tag{4.1}$$

where $b = \dim(\beta) \geq 0$, $a = \dim(\alpha) \geq 1$ and where $C_{b,a}$ denote the same constants as given in theorem 2.7.22. Explicitly,

$$C_{0,1} = C_{0,2} = 0, \quad C_{0,0} = C_{1,2} = C_{2,2} = 1, \quad C_{1,1} = 2. \tag{4.2}$$

Proof. Take (2.18) from theorem 2.7.22, set $x = n\pi R^2$ and multiply the integrand of the incomplete Gamma function with

$$r : [0, x] \to \mathbb{R}_0^+, \quad r(s) = \sqrt{\frac{s}{n\pi}}.$$

Going from $0 = r(0)$ to $R = r(x)$, r goes through all possible radii in between and so,

$$\frac{C_{b,a} n}{\Gamma(a)} \int_0^x r(s) s^{a-1} e^{-s}\, ds$$

yields the expected sum of radii of all considered intervals $[\beta, \alpha]$. Rewriting the above brings us to the desired formula

$$\frac{1}{\Gamma(a)} \frac{C_{b,a}}{\sqrt{\pi}} \int_0^x s^{a+\frac{1}{2}-1} e^{-s}\, ds = \frac{\gamma(a+\frac{1}{2}, n\pi R^2)}{\Gamma(a)} \frac{C_{b,a}}{\sqrt{\pi}} \sqrt{n}.$$

□

0^{th} homology, 0-norm

Proposition 4.3. *Let $\chi : C \to \{0,1\}$ be a random 2-colored chromatic point set inside the unit square $[0,1]^2$ in the setting (n,t,L) with $L = \mathrm{Del}(C_0)$ and $K = \mathrm{Del}(\chi)$. Moreover, let f_K denote the corresponding filtration function of K with f_L respective $f_{K,L}$ being the restrictions of f_K to L respective $K \setminus L$. Then*

$$\mathbb{E}[\|\mathrm{Dgm}_0(f_K)\|_0] = n, \tag{4.3}$$
$$\mathbb{E}[\|\mathrm{Dgm}_0(f_L)\|_0] = \mathbb{E}[\|\mathrm{Dgm}_0(\mathrm{im}\, f_L \to f_K)\|_0] = tn, \tag{4.4}$$
$$\mathbb{E}[\|\mathrm{Dgm}_0(\ker f_L \to f_K)\|_0] = t(1-t)n, \tag{4.5}$$
$$\mathbb{E}[\|\mathrm{Dgm}_0(f_{L,K})\|_0] = \mathbb{E}[\|\mathrm{Dgm}_0(\mathrm{coker}\, f_L \to f_K)\|_0] = (1-t)n. \tag{4.6}$$

Proof. Notice that $\|\mathrm{Dgm}_0(f_K)\|_0$ gives us the number of connected components in the codomain filtration, which can be considered as without any memory of the colors on the points because of the lifting property (see 2.5.1). In expectation, this number of connected components is n as we have n points in the filtration. This yields (4.3).
In the filtration of the domain observe that we expect tn blue points in there, which means tn blue connected components. The second equality in (4.4) follows because out of all expected tn blue points in the domain, each one remains a connected component in the image.
In the relative filtration, on the other hand, there are $(1-t)n$ orange points in expectation. This is also true for the cokernel, where we view the codomain without the image. Thus, we have shown (4.6). Equation (4.5) is a bit more work. This is the only case, where vertices are not yet born in the first step of the filtration but only in the moment they die in the image by getting connected with another vertex. Since the domain L consists of all blue points, we are just interested in those.
Concretely, take an arbitrary point of our chromatic point set. The probability that this point is colored blue equals t. Now, if this blue point is first connected with another blue point, it dies in the image as well as in the domain, so nothing[2] happens in the kernel.

[2] Actually, the blue point then has same birth and death time in the kernel diagram. But we do not take such cases into consideration.

The other option is that our blue point is first connected with an orange point. In the domain, the blue point still lives on but in the image it dies, so it appears in the kernel. This birth in the kernel is realized by a mixed edge. The probability that a death-giving edge with the first boundary vertex being blue is a mixed one equals, therefore, $t(1-t)$ and since there are n-many death-giving edges in expectation (see corollary 2.7.23), the equality in (4.5) follows. □

Proposition 4.4. *Let $\chi : C \to \{0,1\}$ be a random 2-colored chromatic point set inside the unit square $[0,1]^2$ in the setting (n,t,L) with $L = \mathrm{Del}(C_0) \sqcup \mathrm{Del}(C_1)$ and $K = \mathrm{Del}(\chi)$. Moreover, let f_K denote the corresponding filtration function of K with f_L being the restrictions of f_K to L. Then*

$$\mathbb{E}[\|\mathrm{Dgm}_0(f_K)\|_0] = n, \tag{4.7}$$
$$\mathbb{E}[\|\mathrm{Dgm}_0(f_L)\|_0] = \mathbb{E}[\|\mathrm{Dgm}_0(\mathrm{im}\, f_L \to f_K)\|_0] = n, \tag{4.8}$$
$$\mathbb{E}[\|\mathrm{Dgm}_0(\ker f_L \to f_K)\|_0] = 2t(1-t)n. \tag{4.9}$$

Proof. Because the codomain K has stayed the same as in proposition 4.3, equation (4.7) immediately follows from (4.3). Also, (4.8) comes from an analogous argumentation. Here, L consists not only of the set of all blue points but of the disjoint union of blue and orange points. Therefore, we have $tn + (1-t)n = n$ many connected components in expected in the domain filtration as well as in the image.
For the last equality, we already know that we have $t(1-t)n$ connected components in the kernel stemming from blue points by the proof of proposition 4.3. Now, we go through the same logic by starting with an orange point. The probability that we pick an orange one to begin with is $(1-t)$ and the probability that this is a connected component in the kernel is given by $t(1-t)$ due to a death-giving mixed edge that has an orange vertex as its first boundary vertex. In the end, the expected number of connected components in the kernel is equal to $t(1-t)n + t(1-t)n = 2t(1-t)n$, which is also exactly the expected number of mixed death-giving edges in total.
For the relative filtration or the cokernel we do not get any results as there are no connected components in either of them.[3] □

0^{th} homology, 1-norm

Proposition 4.5. *Let $\chi : C \to \{0,1\}$ be a random 2-colored chromatic point set inside the unit square $[0,1]^2$ in the setting (n,t,L) with $L = \mathrm{Del}(C_0)$ and $K = \mathrm{Del}(\chi)$. Moreover, let f_K denote the corresponding filtration function of K with f_L being the restrictions of f_K to L. Then for large enough n,*

$$\mathbb{E}[\|\mathrm{Dgm}_0(f_K)\|_1] = \frac{c}{2}\sqrt{n}, \tag{4.10}$$
$$\mathbb{E}[\|\mathrm{Dgm}_0(f_L)\|_1] = \frac{c}{2}\sqrt{t}\sqrt{n}, \tag{4.11}$$

[3] At least no connected components with finite persistence.

where $c > 0$ is the constant from theorem 2.7.24.

Proof. By proposition 2.4.18, we already know that the 1-norm of connected components in the codomain diagram is half the length of the Euclidean minimum spanning tree of the vertices within the filtration.[4] In expectation and for large n this yields

$$\mathbb{E}[\|\mathrm{Dgm}_0(f_K)\|_1] = \frac{1}{2}\mathbb{E}[|\mathrm{EMST}(C)|] \overset{2.7.24}{=} \frac{1}{2}c\sqrt{n}.$$

The same procedure also leads us to the proof of the second equality with the only difference being that we have C_0 as the underlying point set of L instead of C. Again, for large enough n we obtain

$$\mathbb{E}[\|\mathrm{Dgm}_0(f_L)\|_1] = \frac{1}{2}\mathbb{E}[|\mathrm{EMST}(C_0)|] \overset{2.7.24}{=} \frac{1}{2}c\sqrt{tn}$$

in the expected case. $\square$

Proposition 4.6. *Let $\chi : C \to \{0, 1\}$ be a random 2-colored chromatic point set inside the unit square $[0, 1]^2$ in the setting (n, t, L) with $L = \mathrm{Del}(C_0) \sqcup \mathrm{Del}(C_1)$ and $K = \mathrm{Del}(\chi)$. Moreover, let f_K denote the corresponding filtration function of K with f_L being the restrictions of f_K to L. Then for large enough n,*

$$\mathbb{E}[\|\mathrm{Dgm}_0(f_K)\|_1] = \mathbb{E}[\|\mathrm{Dgm}_0(\mathrm{im}\, f_L \to f_K)\|_1] = \frac{c}{2}\sqrt{n}, \tag{4.12}$$

$$\mathbb{E}[\|\mathrm{Dgm}_0(f_L)\|_1] = \frac{c}{2}(\sqrt{t} + \sqrt{1-t})\sqrt{n}, \tag{4.13}$$

$$\mathbb{E}[\|\mathrm{Dgm}_0(\ker f_L \to f_K)\|_1] = \frac{c}{2}(\sqrt{t} + \sqrt{1-t} - 1)\sqrt{n}, \tag{4.14}$$

where $c > 0$ is the constant from theorem 2.7.24.

Proof. The expected 1-norm of connected components in the codomain filtration follows immediately from (4.10). Notice now that the set of connected components in the image is $C_0 \sqcup C_1 = C$ and they are all born at $r = 0$ and die as soon as they get merged with another vertex regardless their color. We conclude that the 0^{th} persistence pairs of the image equal the ones of the codomain, which gives us (4.12).
For (4.13), recall that the underlying point set of L is the disjoint union of C_0 and C_1, so we view these sets separately and obtain

$$\mathbb{E}[\|\mathrm{Dgm}_0(f_L)\|_1] = \frac{1}{2}\big(\mathbb{E}[|\mathrm{EMST}(C_0)|] + \mathbb{E}[|\mathrm{EMST}(C_1)|]\big) = \frac{c}{2}(\sqrt{t} + \sqrt{1-t})\sqrt{n}.$$

The last equation (4.14) about the kernel can be easily deduced from (2.11) in theorem 2.6.4. Like in the proof of proposition 4.4, neither are there connected components present in the relative filtration nor in the cokernel. Thus, we do not get any results for these two cases. $\square$

[4] The cutoff δ disappears in this case as we are only interested in the finite persistence pairs by remark 4.1.

1^{st} homology, 0-norm

Proposition 4.7. *Let $\chi : C \to \{0,1\}$ be a random 2-colored chromatic point set inside the unit square $[0,1]^2$ in the setting (n,t,L) with $L = \mathrm{Del}(C_0)$ and $K = \mathrm{Del}(\chi)$. Moreover, let f_K denote the corresponding filtration function of K with f_L respective $f_{K,L}$ being the restrictions of f_K to L respective $K \setminus L$. Then*

$$\mathbb{E}[\|\mathrm{Dgm}_1(f_K)\|_0] = n, \tag{4.15}$$
$$\mathbb{E}[\|\mathrm{Dgm}_1(f_L)\|_0] = tn, \tag{4.16}$$
$$\mathbb{E}[\|\mathrm{Dgm}_1(f_{L,K})\|_0] = (2t+1)(1-t)n. \tag{4.17}$$

Proof. First of all, there exists a bijective correspondence between the number of loops within a filtration of the Delaunay complex and the number of critical triangles. In the expected case, by corollary 2.7.23 the number of critical triangles is equal to n in the codomain filtration and tn in the domain filtration, which provides the proof of (4.15) and (4.16) simultaneously. The last equation (4.17) is a bit more work. Observe that the expected number of critical triangles in the relative diagram also includes triangles, which are non-critical in the codomain.[5] Moreover, there can be critical triangles in the codomain that are no death-giving triangles in the relative but birth-giving to 2-dimensional holes instead.[6] Thus, we tackle this problem by counting all critical birth-giving edges in the relative filtration. To begin with, it is easily checked that all critical edges in the relative have to be critical in the codomain. By corollary 2.7.23, there are $2n$ of them in expectation. Since $L = \mathrm{Del}(C_0)$, all blue-colored edges are mod out in the relative, so the only remaining critical edges have to be purely orange or mixed. This results in $(1-t^2) \cdot 2n = (2-2t^2)n$ critical edges in the relative on average. Subtracting from this number all expected critical death-giving edges in the relative, which are $(1-t)n$ by proposition 4.3[7], gives the claim. ☐

Proposition 4.8. *Let $\chi : C \to \{0,1\}$ be a random 2-colored chromatic point set inside the unit square $[0,1]^2$ in the setting (n,t,L) with $L = \mathrm{Del}(C_0) \sqcup \mathrm{Del}(C_1)$ and $K = \mathrm{Del}(\chi)$. Moreover, let f_K denote the corresponding filtration function of K with f_L respective $f_{K,L}$ being the restrictions of f_K to L respective $K \setminus L$. Then*

$$\mathbb{E}[\|\mathrm{Dgm}_1(f_K)\|_0] = \mathbb{E}[\|\mathrm{Dgm}_1(f_L)\|_0] = n, \tag{4.18}$$
$$\mathbb{E}[\|\mathrm{Dgm}_1(f_{L,K})\|_0] = 4t(1-t)n. \tag{4.19}$$

Proof. The expected number of loops in the codomain filtration follows immediately from (4.15). Moreover, since the number of critical triangles in the domain filtration equals $tn + (1-t)n = n$, this shows the second equality in (4.18). The relative filtration

[5] As an example, consider a mixed-colored obtuse triangle where the hypotenuse has two blue endpoints.

[6] For instance, take a critical orange-colored triangle in the codomain that is surrounded by many blue points.

[7] To be precise, there are $(1-t)n-1$ such, but we want we want to exclude the one infinite persistence pair we have in the relative in the end of our argumentation, so we add $+1$.

is again more interesting. According to theorem 2.3.39, $H_1(K_r, L_r) \cong H_1(|K_r| / |L_r|)$ in each filtration step $r \geq 0$ and we can think of $|K_r| / |L_r|$ as the underlying space of the lifted chromatic Delaunay complex, where we shrink all blue simplices to one point, all orange ones to another and identify these two remaining points. In the end, only mixed-colored simplices are not mod out and what first were edges in the codomain filtration become loops in the relative filtration. In fact, those are all loops there are. Therefore, the expected number of loops in the relative case equals the number of critical mixed-colored edges, which is $2t(1-t) \cdot 2n = 4t(1-t)n$. □

1^{st} homology, 1-norm

Proposition 4.9. *Let $\chi : C \to \{0,1\}$ be a random 2-colored chromatic point set inside the unit square $[0,1]^2$ in the setting (n,t,L) with $L = \mathrm{Del}(C_0)$ and $K = \mathrm{Del}(\chi)$. Moreover, let f_K denote the corresponding filtration function of K with f_L being the restrictions of f_K to L. Then for large enough n*

$$\mathbb{E}[\|\mathrm{Dgm}_1(f_K)\|_1] = \left(\frac{c}{2} - \frac{1}{4}\right)\sqrt{n}, \tag{4.20}$$

$$\mathbb{E}[\|\mathrm{Dgm}_1(f_L)\|_1] = \left(\frac{c}{2} - \frac{1}{4}\right)\sqrt{t}\sqrt{n}. \tag{4.21}$$

Proof. To begin with we want to fully concentrate on proving (4.20). To compute the expected length of loops in the codomain filtration, we have to sum up all persistences of individual 1-holes. That is

$$\sum_{(r_i,r_j) \text{ loop}} (r_j - r_i) = \sum_{(r_i,r_j) \text{ loop}} r_j - \left(\sum_{(r_i,r_j) \text{ loop}} r_i\right).$$

The first sum on the right hand side is only contributed by death values of loops whereas the second one adds up all birth values of loops. By corollary 2.7.23, the appearance of a critical triangle always marks the death of an existing 1-hole.
With the help of proposition 4.2, the first sum on the right hand side in expectation is, thus, $\frac{\Gamma(\frac{5}{2})}{\Gamma(2)} \frac{C_{2,2}}{\sqrt{\pi}} \sqrt{n} = \frac{3}{4}\sqrt{n}$.
The other sum on the right hand side, on the contrary, consists of the radii of all critical birth-giving edges. In proposition 4.2 we cannot separate expected birth-giving from death-giving persistences and so the trick is to subtract the expected lengths of all critical edges and to add back the ones of all death-giving ones. In the end, we obtain the desired equality

$$\mathbb{E}[\|\mathrm{Dgm}_1(f_K)\|_1] = \frac{3}{4}\sqrt{n} - \frac{\Gamma(\frac{3}{2})}{\Gamma(2)} \frac{C_{1,1}}{\sqrt{\pi}} \sqrt{n} + \mathbb{E}[\|\mathrm{Dgm}_0(f_K)\|_1] = \left(\frac{c}{2} - \frac{1}{4}\right)\sqrt{n}.$$

Going through the same thinking with the domain filtration, this yields

$$\mathbb{E}[\|\mathrm{Dgm}_1(f_L)\|_1] = \frac{3}{4}\sqrt{tn} - \sqrt{tn} + \mathbb{E}[\|\mathrm{Dgm}_0(f_L)\|_1] = \left(\frac{c}{2} - \frac{1}{4}\right)\sqrt{tn}.$$

□

Proposition 4.10. *Let $\chi : C \to \{0,1\}$ be a random 2-colored chromatic point set inside the unit square $[0,1]^2$ in the setting (n,t,L) with $L = \mathrm{Del}(C_0) \sqcup \mathrm{Del}(C_1)$ and $K = \mathrm{Del}(\chi)$. Moreover, let f_K denote the corresponding filtration function of K with f_L being the restrictions of f_K to L. Then for large enough n*

$$\mathbb{E}[\|\mathrm{Dgm}_1(f_K)\|_1] = \left(\frac{c}{2} - \frac{1}{4}\right)\sqrt{n}, \tag{4.22}$$

$$\mathbb{E}[\|\mathrm{Dgm}_1(f_L)\|_1] = \left(\frac{c}{2} - \frac{1}{4}\right)(\sqrt{t} + \sqrt{1-t})\sqrt{n}. \tag{4.23}$$

Proof. Equation (4.22) immediately follows from (4.20). To show (4.23), we again notice that the underlying point set of L is the disjoint union of C_0 and C_1, which yields

$$\mathbb{E}[\|\mathrm{Dgm}_1(f_L)\|_1] = \left(\frac{c}{2} - \frac{1}{4}\right)\sqrt{tn} + \left(\frac{c}{2} - \frac{1}{4}\right)\sqrt{(1-t)n}.$$

□

In appendix A.2, all outcomes of the computational experiments are given and compared to the theoretical results if available. In some instances, where we could not come up with a proof for the corresponding expectation, we state conjectures of possible explicit formulae. In Tables 4.1 and 4.2 we compactly list all current theoretically proven results on expected 0- and 1-norms (*blue*) opposed to conjectures due to the experiments we have done (*yellow*). White boxes with a question mark indicate the cases where we neither have a theoretical idea nor a concrete computational hint to what the explicit form could look like.

The first table is about the 0^{th} homology, so connected components, whereas the second one is about the 1^{st} homology, meaning loops. Each big row stands for one diagram and is divided into two smaller rows – one for the 0-norm and the other one for the 1-norm. The two columns are the two kinds of interesting filtrations there are for two colors.

Observe that nearly all different cases are worked out in Table 4.1. Also, due to the symmetry in t for $L = \mathrm{Del}(C_0) \sqcup \mathrm{Del}(C_1)$, we obtain rather simple functions. This changes for $L = \mathrm{Del}(C_0)$, where the conjectured explicit formula seems to be way more delicate in the 1-norm setting. The good news is, though, that once we are able to show either one of these conjectures in Table 4.1, we can deduce the rest of them with the help of the powerful relations we have in (2.11), (2.12) and (2.13).
In Table 4.2 roughly half of the boxes are still given with a question mark, most of which are regarding the expected length of loops in some persistence diagrams. The computational experiments, however, shed light on some very simple and, at the same time, also well-fitting functions that might be rigorously provable. Especially intriguing seems to be the expected number of loops in the kernel diagram for both filtrations.

0^{th} homology		mono-filt	0-filt
kernel	0	$2t(1-t)n$	$t(1-t)n$
	1	$\frac{c}{2}\big(\sqrt{t}+\sqrt{1-t}-1\big)\sqrt{n}$	$\big((\frac{c}{2}-\frac{1}{4})t^2-\frac{c}{2}t+\frac{1}{4}\sqrt{t}\big)\sqrt{n}$
relative	0		$(1-t)n$
	1	—	$\big((\frac{c}{2}-\frac{1}{4})t^2-\frac{c}{2}t-(\frac{c}{2}-\frac{1}{4})\sqrt{t}+\frac{c}{2}\big)\sqrt{n}$
cokernel	0		$(1-t)n$
	1		$\big((\frac{c}{2}-\frac{1}{4})t^2-\frac{c}{2}t-(\frac{c}{2}-\frac{1}{4})\sqrt{t}+\frac{c}{2}\big)\sqrt{n}$
domain	0	n	tn
	1	$\frac{c}{2}\big(\sqrt{t}+\sqrt{1-t}\big)\sqrt{n}$	$\frac{c}{2}\sqrt{t}\sqrt{n}$
image	0	n	tn
	1	$\frac{c}{2}\sqrt{n}$	$\big(-(\frac{c}{2}-\frac{1}{4})t^2+\frac{c}{2}t+(\frac{c}{2}-\frac{1}{4})\sqrt{t}\big)\sqrt{n}$
codomain	0	n	
	1	$\frac{c}{2}\sqrt{n}$	

Table 4.1.: Explicit proven formulae (*blue*) or conjectures (*yellow*) for expectations in 0^{th} persistent homology. Each row for the diagrams is subdivided into 0-norm (*upper part*) and 1-norm (*lower part*).

1^{st} homology		mono-filt	0-filt
kernel	0	$3t(1-t)n$	$t(1-t)(1+t)n$
	1	?	?
relative	0	$4t(1-t)n$	$(2t+1)(1-t)n$
	1	?	?
cokernel	0	$\mu t(1-t)n$, $\mu \approx 2.8$	?
	1	?	?
domain	0	n	tn
	1	$(\frac{c}{2}-\frac{1}{4})\big(\sqrt{t}+\sqrt{1-t}\big)\sqrt{n}$	$(\frac{c}{2}-\frac{1}{4})\sqrt{t}\sqrt{n}$
image	0	$1-\kappa t(1-t)n$, $\kappa \approx 2.4$	?
	1	?	?
codomain	0	n	
	1	$(\frac{c}{2}-\frac{1}{4})\sqrt{n}$	

Table 4.2.: Explicit proven formulae (*blue*), conjectures (*yellow*) or completely unknown cases (*white with question mark*) for expectations in 1^{st} persistent homology. Each row for the diagrams is subdivided into 0-norm (*upper part*) and 1-norm (*lower part*).

5. Outlook

It is noticable from the previous two tables that this thesis was the starting point for many more open questions that we have on the six-pack. First of all, it would be interesting to verify or falsify the mentioned conjectures. Secondly, it seems as if the boundary effects that influence the number of critical triangles might be the reason why it is difficult to come up with concrete theoretical conjectures out of the experiments for the 1^{st} persistent homology. Going from the unit square to the torus could resolve this issue and lead us to better guesses on the expectations for loops. We close this thesis with even further questions:

$\boxed{1}$ Can we also find formulae for the 2-norm?

$\boxed{2}$ Instead of using two colors, consider now three. Then we are suddenly confronted with essentially four different filtrations. Moreover, we have two varying parameters on the probability. Can we still derive general explicit formulae for the expectations in 0- and 1-norm?

$\boxed{3}$ Instead of having a chromatic point cloud in the plane, we could go one dimension higher and ask the same questions within the unit cube. Will the 3-norm then be interesting to look at, too? And what about three colors in three dimensions?

$\boxed{4}$ Is there any way to generalize Theorem 1 from [19] and our modified version for the expected length (see proposition 4.2) to the chromatic setting?

$\boxed{5}$ In our results, the unknown asymptotic constant c crossed our way many times. Is it possible to improve lower or upper bounds on c due to the relations of the persistence diagrams and proven formulae we have?

A. Appendix

A.1. Behavior of Means of Norms

On the next four pages we present figures in the style of a six-pack, where we plot the number of points against the mean of a specific norm – either 0- or 1-norm. Depending on the 0^{th} or 1^{st} homology as well as on the choice of filtration for L we obtain different results. For each fixed setting of these parameters we, moreover, choose three probability distributions for coloring the points in two colors, namely $[0.3, 0.7]$, $[0.5, 0.5]$ and $[0.8, 0.2]$. Being of the form $[t, 1-t]$, the variable t denotes the probability of picking a 0-colored point out of all points in the chromatic point set $\chi : C \to \{0, 1\}$.
In appendix A.2 we further investigate the relation between changing values of means with varying probability distributions and increasing number of points.

<u>HARD FACTS:</u>
Number of experiments: 100
Number of points: parameter (n) on x-axis with values
100, 1000, 2000, 5000, 7000, 10000, 12000, 15000, 20000
Number of colors: 2
Probability of colors: $[0.3, 0.7]$, $[0.5, 0.5]$, $[0.8, 0.2]$
Filtrations:
$K = \mathrm{Del}(\chi)$ with $L = \mathrm{Del}(C_0)$ ("0-filt") or $L = \mathrm{Del}(C_0) \sqcup \mathrm{Del}(C_1)$ ("mono-filt")
Spatial dimension: 2
Distribution of points: uniform
Location of points: unit square
Diagrams: six-pack fashion
Homology: 0^{th}, 1^{st}
Norms: 0, 1

0^{th} homology, 0-norm

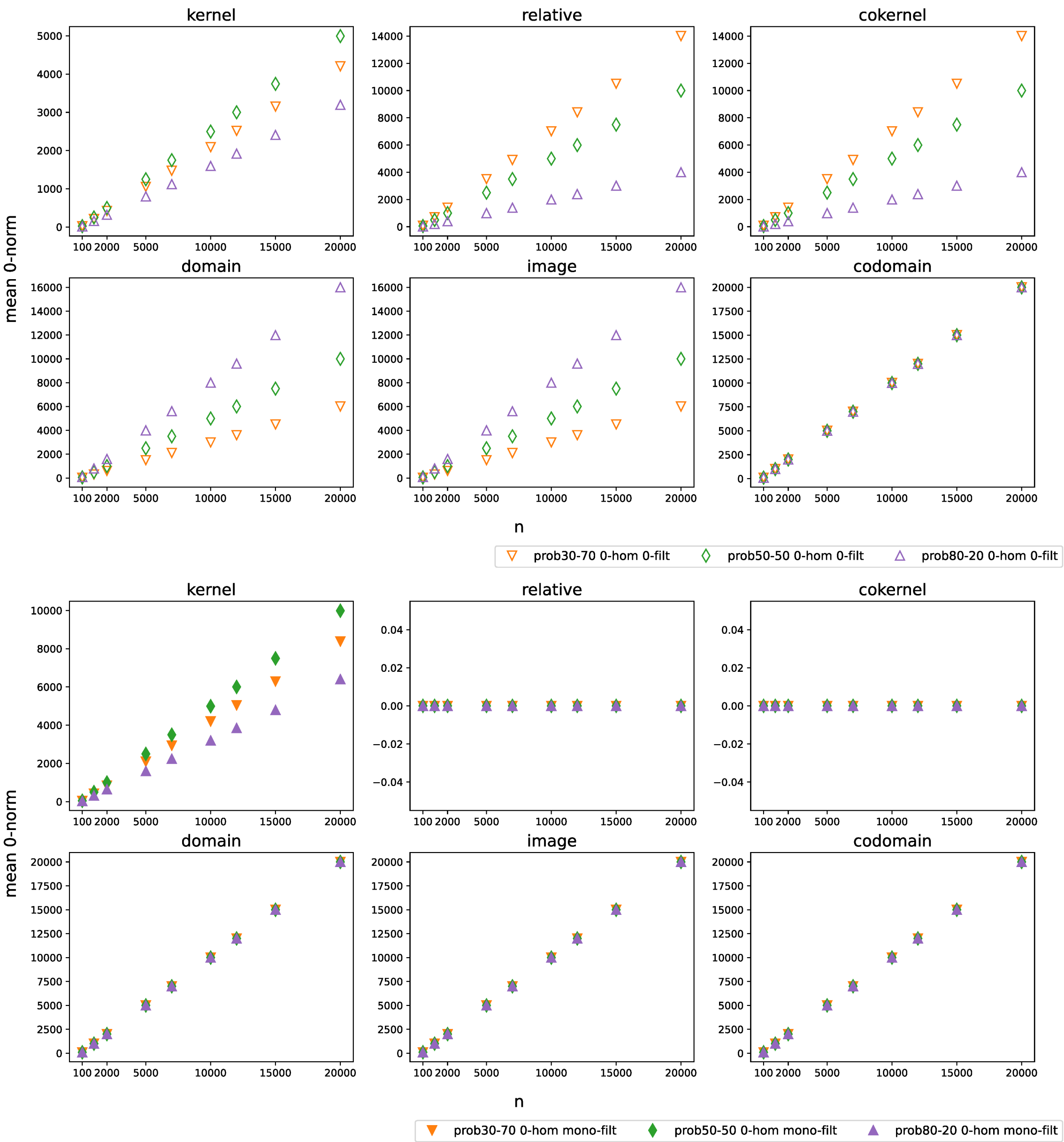

Figure A.1.: Different developments for means of 0-norm in 0^{th} persistent homology in a six-pack fashion for increasing number of points and various probability distributions for coloring in the case $L = \mathrm{Del}(C_0)$ (*top*) and $L = \mathrm{Del}(C_0) \sqcup C_1$ (*bottom*).

0^{th} homology, 1-norm

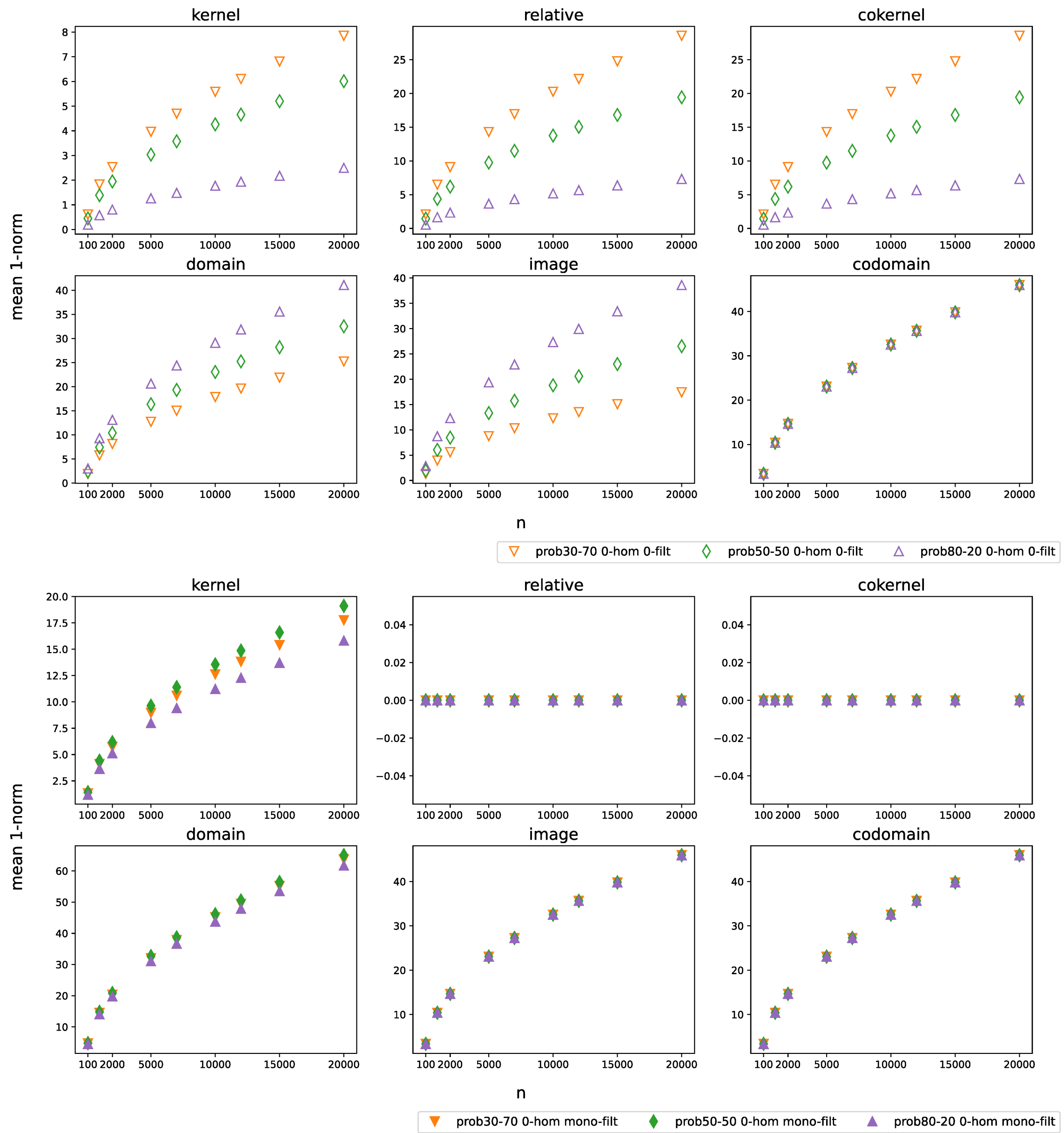

Figure A.2.: Different developments for means of 1-norm in 0^{th} persistent homology in a six-pack fashion for increasing number of points and various probability distributions for coloring in the case $L = \mathrm{Del}(C_0)$ (*top*) and $L = \mathrm{Del}(C_0) \sqcup C_1$ (*bottom*).

1^{st} homology, 0-norm

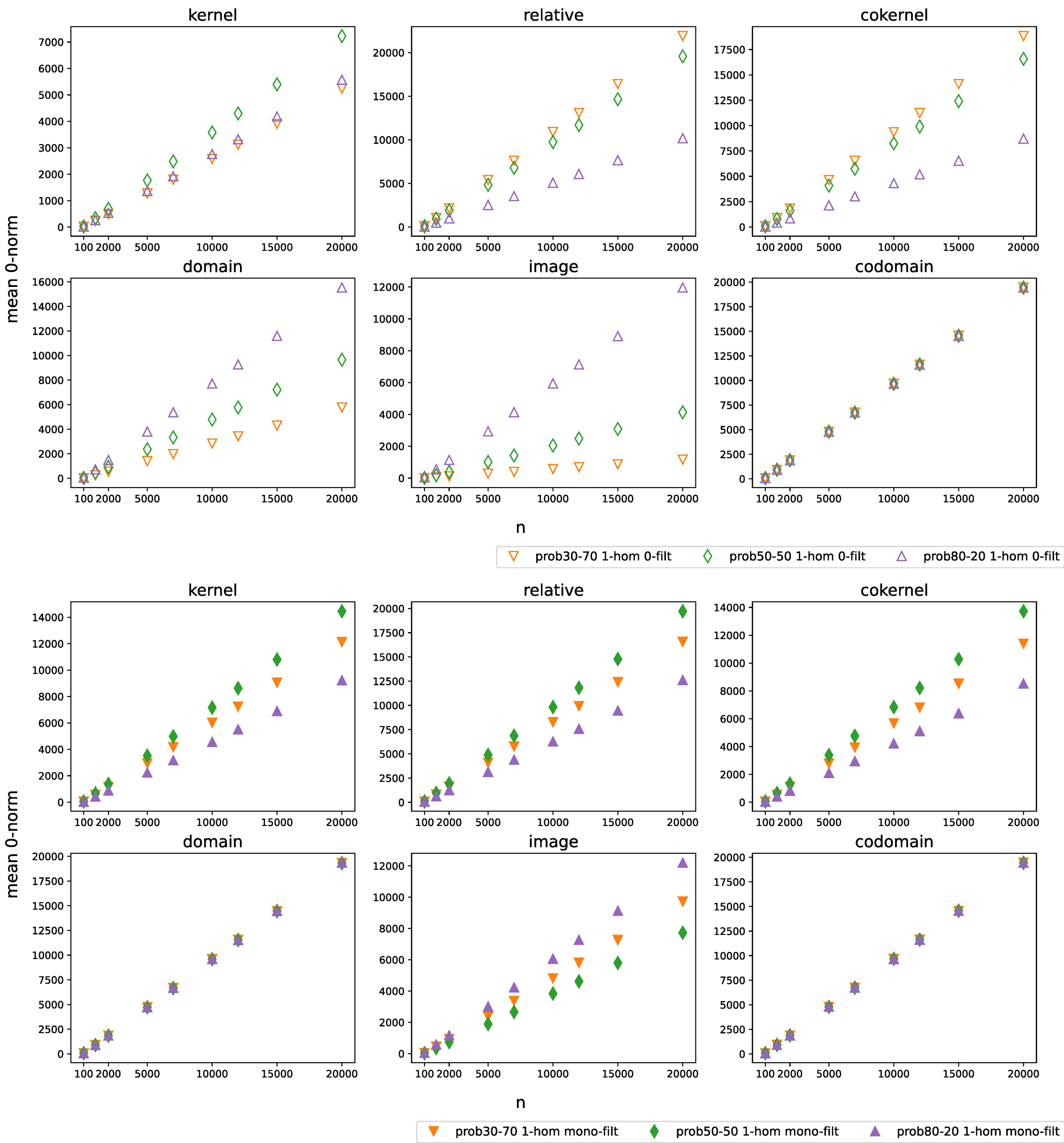

Figure A.3.: Different developments for means of 0-norm in 1^{st} persistent homology in a six-pack fashion for increasing number of points and various probability distributions for coloring in the case $L = \mathrm{Del}(C_0)$ (*top*) and $L = \mathrm{Del}(C_0) \sqcup C_1$ (*bottom*).

1^{st} homology, 1-norm

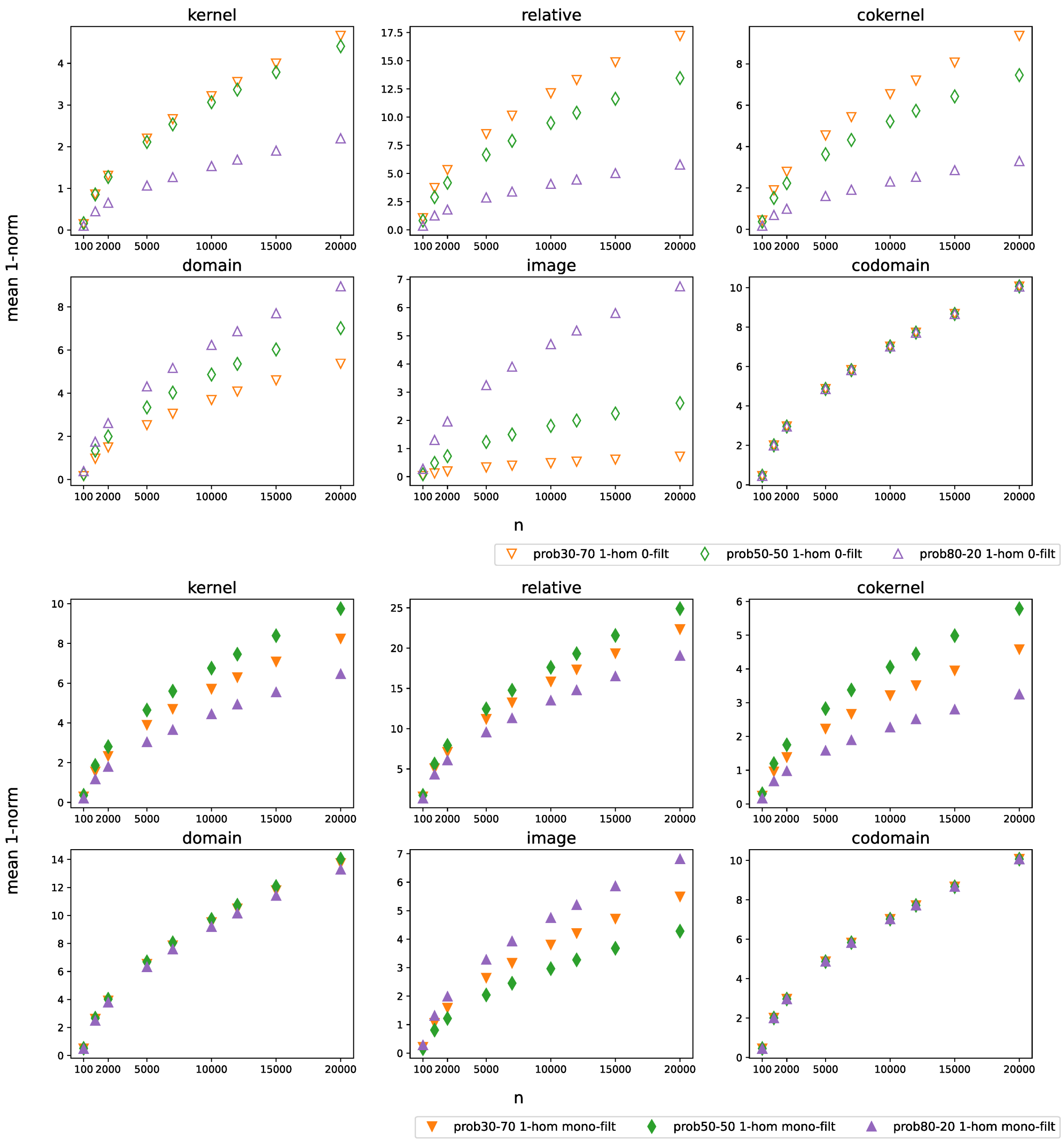

Figure A.4.: Different developments for means of 1-norm in 1^{st} persistent homology in a six-pack fashion for increasing number of points and various probability distributions for coloring in the case $L = \mathrm{Del}(C_0)$ (*top*) and $L = \mathrm{Del}(C_0) \sqcup C_1$ (*bottom*).

A.2. Best Fitting Curves for Means of Norms

Notice that we have a linear behavior with respect to n when taking the 0-norm and a square root behavior with respect to n when taking the 1-norm (see appendix A.1). Thus, we see that for n points the means of the ...

... 0-norm form a curve of the form

$$n \mapsto c_1 \cdot n + c_0,$$

... 1-norm form a curve of the form

$$n \mapsto c_1 \cdot \sqrt{n} + c_0,$$

where c_1 and c_0 might depend on the probability t that a point is 0-colored. For 1^{st} homology the parameter c_0 may even depend on n due to boundary effects. Concretely, points that lie near the boundary of the unit square are more likely to form obtuse triangles than acute ones. This results in an imbalance that is possibly the reason why, for instance, $\mathbb{E}[\|\mathrm{Dgm}_1(f_K)\|_0]$ empirically only reaches $0.9747n + c_0$ instead of n as the theory suggests.

In the following pages we illustrate the best fitting curves that we obtain when plotting the experimentally computed $c_1^{(\mathrm{exp})}$ against the probability t. Note that for better visualization, all values $c_1^{(\mathrm{exp})}$ are rounded up to four decimals. The goodness-of-fit is given by the *root mean square error* or *rmse* for short.

<u>HARD FACTS:</u>
Number of experiments: 100
Number of points: parameter (n) on x-axis with values
100, 300, 500, 700, 1000, 1200, 1500, 1700, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 12000, 15000, 17000 20000
Number of colors: 2
Probability of colors: parameter ($[t, 1-t]$)
$[0.01, 0.99]$, $[0.05, 0.95]$, $[0.07, 0.93]$, $[0.1, 0.9]$, $[0.15, 0.85]$, $[0.2, 0.8]$, $[0.25, 0.75]$, $[0.3, 0.7]$, $[0.35, 0.65]$, $[0.4, 0.6]$, $[0.45, 0.55]$, $[0.5, 0.5]$, $[0.55, 0.45]$, $[0.6, 0.4]$, $[0.65, 0.35]$, $[0.7, 0.3]$, $[0.75, 0.25]$, $[0.8, 0.2]$, $[0.85, 0.15]$, $[0.9, 0.1]$, $[0.93, 0.07]$, $[0.95, 0.05]$, $[0.99, 0.01]$
Filtrations:
$K = \mathrm{Del}(\chi)$ with $L = \mathrm{Del}(C_0)$ ("0-filt") or $L = \mathrm{Del}(C_0) \sqcup \mathrm{Del}(C_1)$ ("mono-filt")
Spatial dimension: 2
Diagrams: kernel, relative, cokernel, domain, image, codomain
Distribution of points: uniform
Location of points: unit square
Homology: 0^{th}, 1^{st}

FIXED PARAMETERS:
Diagram: codomain
Homology: 0^{th}
Norm: 0, 1
Filtration: Diagram does not depend on the chosen filtration

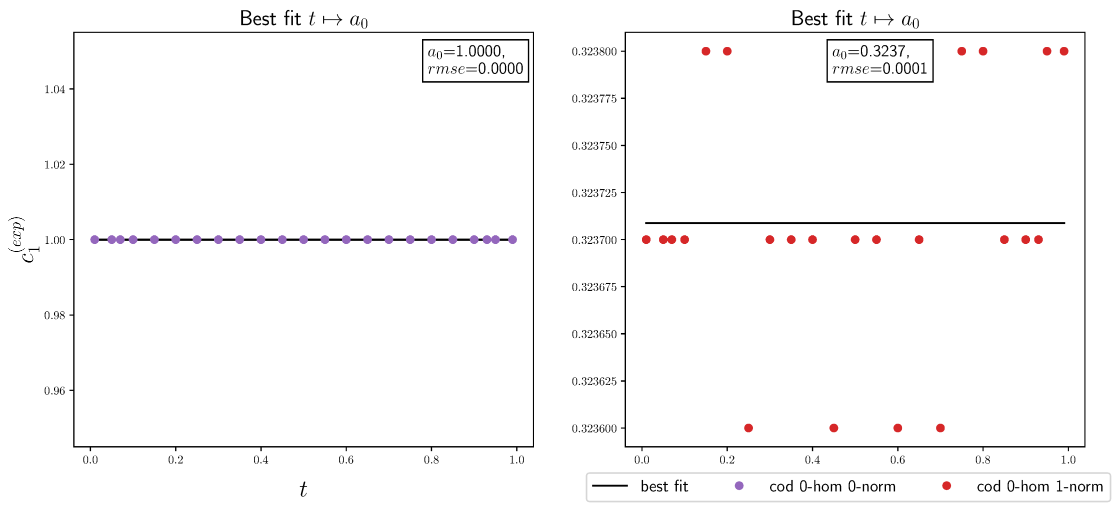

Figure A.5.: Best fits for $c_1^{(\text{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectations

$$\boxed{\mathbb{E}[\|\mathrm{Dgm}_0(f_K)\|_0] = n, \quad \mathbb{E}[\|\mathrm{Dgm}_0(f_K)\|_1] = \frac{c}{2}\sqrt{n}} \tag{A.1}$$

we get from propositions 4.3 and 4.5, respectively. In particular, our experimentally obtained constant $c^{(\text{exp})} = 2a_0 \approx 0.6474$ fits with the bounds in (2.22).

FIXED PARAMETERS:
Diagram: domain
Homology: 0^{th}
Norm: 0, 1
Filtration: 0-filt

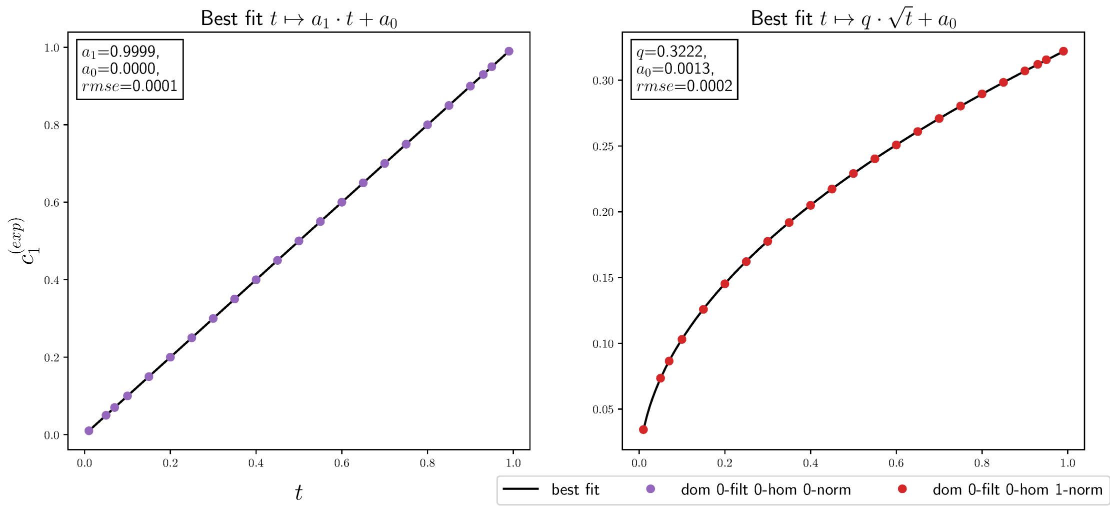

Figure A.6.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectations

$$\boxed{\mathbb{E}[\|\mathrm{Dgm}_0(f_L)\|_0] = tn, \quad \mathbb{E}[\|\mathrm{Dgm}_0(f_L)\|_1] = \frac{c}{2}\sqrt{t}\sqrt{n}} \tag{A.2}$$

we get from propositions 4.3 and 4.5, respectively. In particular, our experimentally obtained constant $c^{(\mathrm{exp})} = 2a_0 \approx 0.6444$ fits with the bounds in (2.22).

<u>FIXED PARAMETERS:</u>
Diagram: domain
Homology: 0^{th}
Norm: 0, 1
Filtration: mono-filt

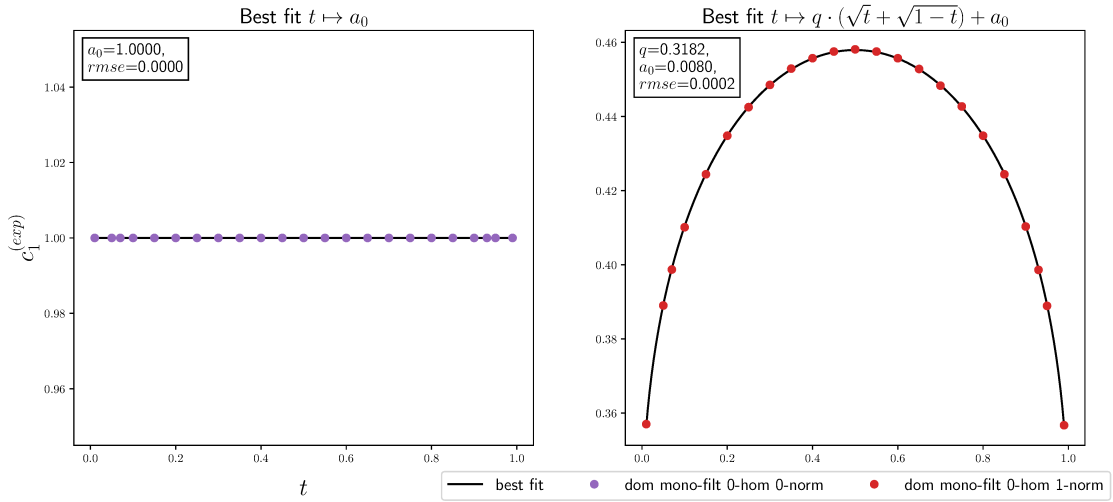

Figure A.7.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectations

$$\boxed{\mathbb{E}[\|\mathrm{Dgm}_0(f_L)\|_0] = n, \quad \mathbb{E}[\|\mathrm{Dgm}_0(f_L)\|_1] = \frac{c}{2}(\sqrt{t} + \sqrt{1-t})\sqrt{n}} \tag{A.3}$$

we get from propositions 4.4 and 4.6, respectively. In particular, our experimentally obtained constant $c^{(\mathrm{exp})} = 2q \approx 0.6364$ fits with the bounds in (2.22).

FIXED PARAMETERS:
Diagram: image
Homology: 0^{th}
Norm: 0, 1
Filtration: 0-filt

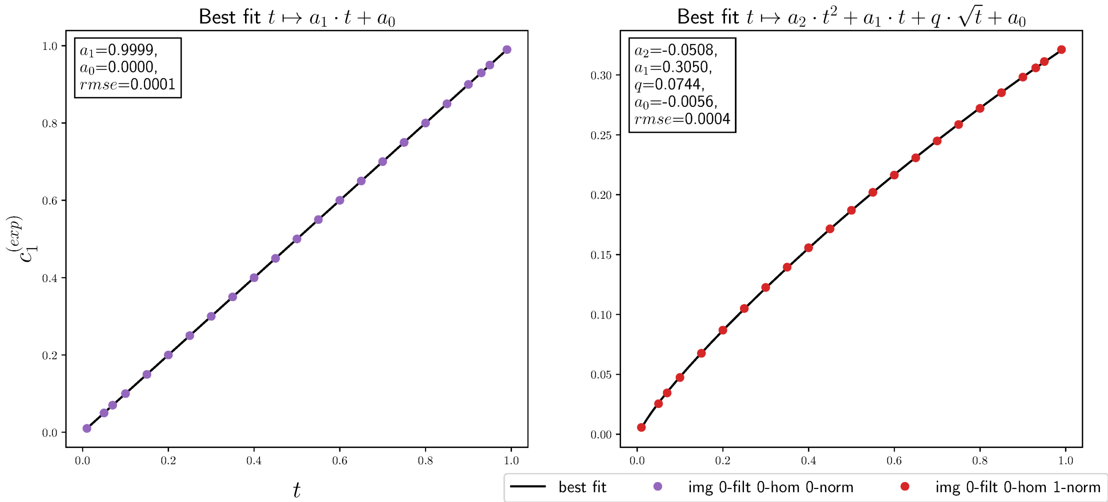

Figure A.8.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectation

$$\boxed{\mathbb{E}[\|\mathrm{Dgm}_0(\mathrm{im}\, f_L \to f_K)\|_0] = tn} \tag{A.4}$$

we get from proposition 4.3. Moreover, we observe a potential explicit form for the 1-norm and conjecture that

$$\mathbb{E}[\|\mathrm{Dgm}_0(\mathrm{im}\, f_L \to f_K)\|_1] = \left(-\left(\frac{c}{2} - \frac{1}{4}\right)t^2 + \frac{c}{2}t + \left(\frac{c}{2} - \frac{1}{4}\right)\sqrt{t}\right)\sqrt{n}. \tag{A.5}$$

<u>FIXED PARAMETERS:</u>
Diagram: image
Homology: 0^{th}
Norm: 0, 1
Filtration: mono-filt

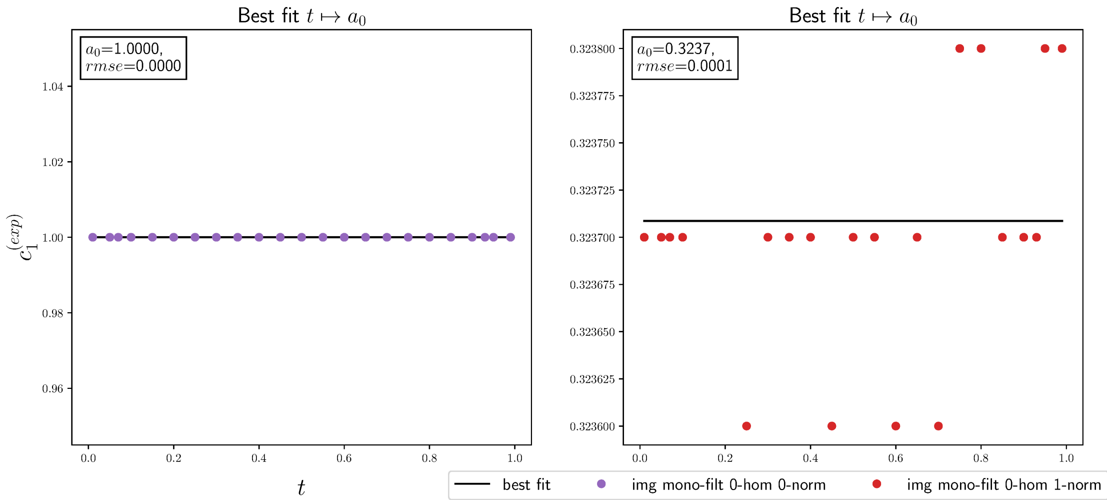

Figure A.9.: Best fits for $c_1^{(\exp)}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectations

$$\boxed{\mathbb{E}[\|\mathrm{Dgm}_0(\operatorname{im} f_L \to f_K)\|_0] = n, \quad \mathbb{E}[\|\mathrm{Dgm}_0(\operatorname{im} f_L \to f_K)\|_1] = \frac{c}{2}\sqrt{n}} \tag{A.6}$$

we get from propositions 4.4 and 4.6, respectively. In particular, our experimentally obtained constant $c^{(\exp)} = 2a_0 \approx 0.6474$ fits with the bounds in (2.22).

FIXED PARAMETERS:
Diagram: kernel
Homology: 0^{th}
Norm: 0, 1
Filtration: 0-filt

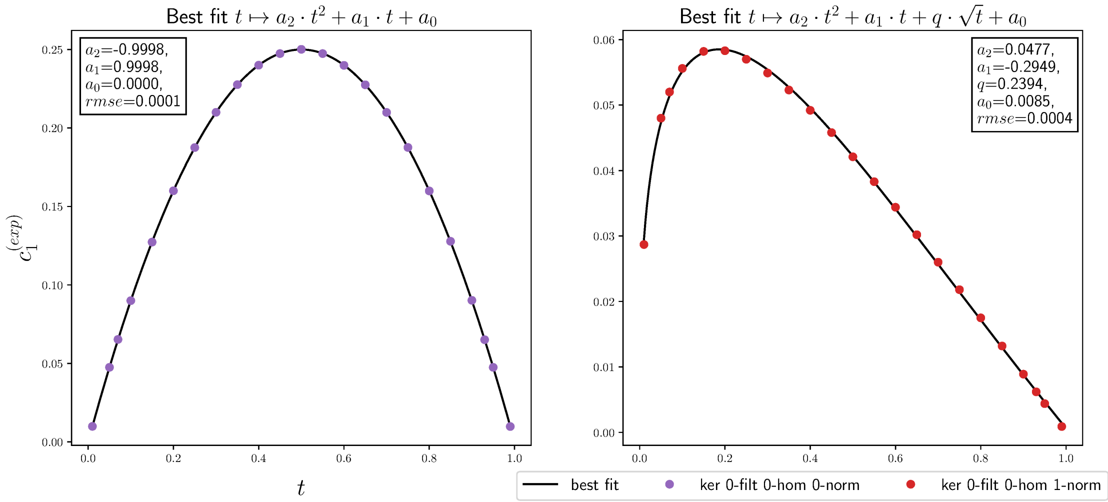

Figure A.10.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectation

$$\boxed{\mathbb{E}[\|\mathrm{Dgm}_0(\ker f_L \to f_K)\|_0] = t(1-t)n} \tag{A.7}$$

we get from proposition 4.3. Moreover, we observe a potential explicit form for the 1-norm and conjecture that

$$\mathbb{E}[\|\mathrm{Dgm}_0(\ker f_L \to f_K)\|_1] = \left(\left(\frac{c}{2} - \frac{1}{4}\right)t^2 - \frac{c}{2}t + \frac{1}{4}\sqrt{t}\right)\sqrt{n}. \tag{A.8}$$

FIXED PARAMETERS:
Diagram: kernel
Homology: 0^{th}
Norm: 0, 1
Filtration: mono-filt

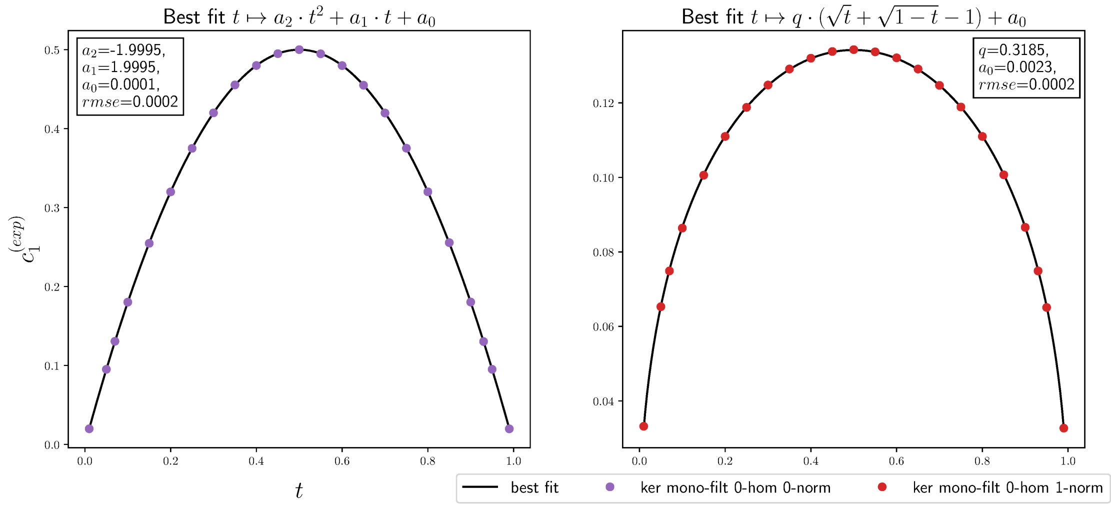

Figure A.11.: Best fits for $c_1^{(\exp)}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectations

$$\boxed{\begin{aligned} \mathbb{E}[\|\mathrm{Dgm}_0(\ker f_L \to f_K)\|_0] &= 2t(1-t)n, && \text{(A.9)} \\ \mathbb{E}[\|\mathrm{Dgm}_0(\ker f_L \to f_K)\|_1] &= \frac{c}{2}(\sqrt{t} + \sqrt{1-t} - 1)\sqrt{n} && \text{(A.10)} \end{aligned}}$$

we get from propositions 4.4 and 4.6, respectively. In particular, our experimentally obtained constant $c^{(\exp)} = 2q \approx 0.6370$ fits with the bounds in (2.22).

FIXED PARAMETERS:
Diagram: cokernel
Homology: 0^{th}
Norm: 0, 1
Filtration: 0-filt

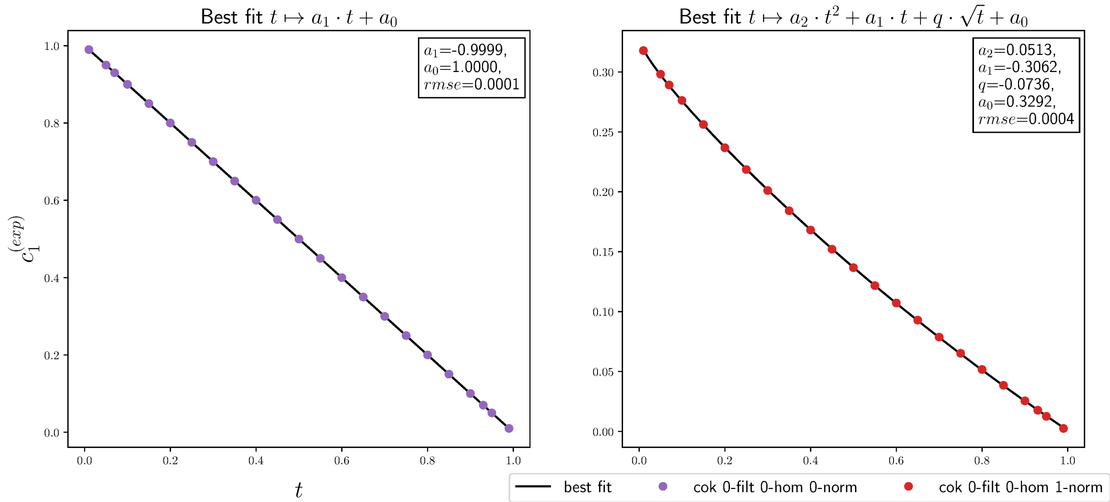

Figure A.12.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectation

$$\boxed{\mathbb{E}[\|\mathrm{Dgm}_0(\mathrm{coker} f_L \to f_K)\|_0] = (1-t)n} \tag{A.11}$$

we get from proposition 4.3. Moreover, we observe a potential explicit form for the 1-norm and conjecture that

$$\mathbb{E}[\|\mathrm{Dgm}_0(\mathrm{coker} f_L \to f_K)\|_1] = \left(\left(\frac{c}{2} - \frac{1}{4}\right)t^2 - \frac{c}{2}t - \left(\frac{c}{2} - \frac{1}{4}\right)\sqrt{t} + \frac{c}{2}\right)\sqrt{n}. \tag{A.12}$$

FIXED PARAMETERS:
Diagram: relative
Homology: 0^{th}
Norm: 0, 1
Filtration: 0-filt

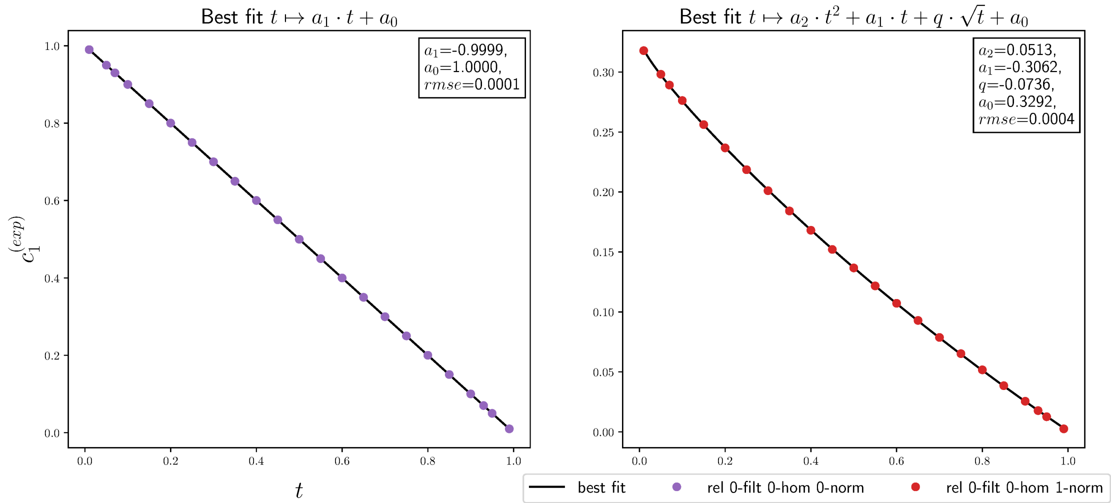

Figure A.13.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectation

$$\boxed{\mathbb{E}[\|\mathrm{Dgm}_0(f_{L,K})\|_0] = (1-t)n} \tag{A.13}$$

we get from proposition 4.3. Moreover, we observe a potential explicit form for the 1-norm and conjecture that

$$\mathbb{E}[\|\mathrm{Dgm}_0(f_{L,K})\|_1] = \left(\left(\frac{c}{2} - \frac{1}{4}\right)t^2 - \frac{c}{2}t - \left(\frac{c}{2} - \frac{1}{4}\right)\sqrt{t} + \frac{c}{2}\right)\sqrt{n}. \tag{A.14}$$

<u>FIXED PARAMETERS:</u>
Diagram: codomain
Homology: 1^{st}
Norm: 0, 1
Filtration: Diagram does not depend on the chosen filtration

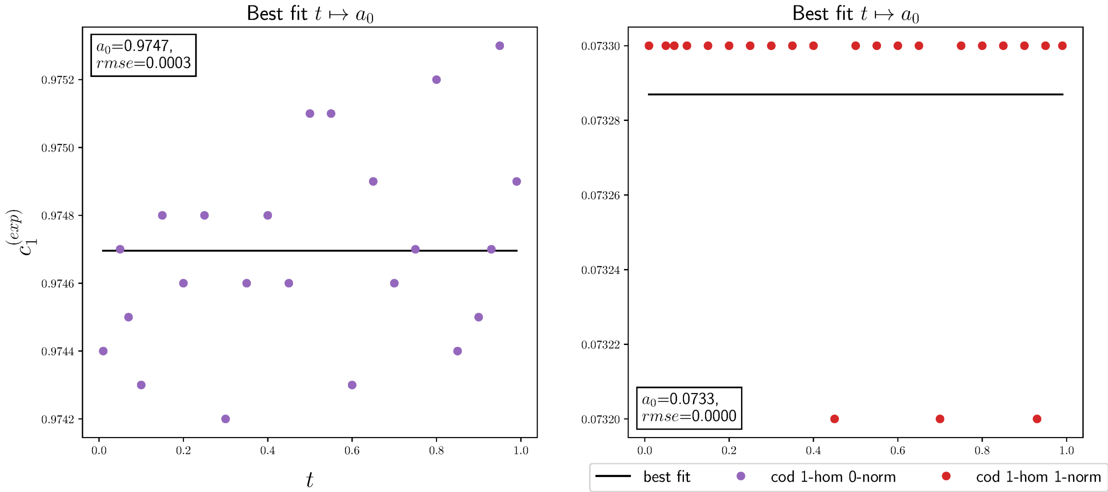

Figure A.14.: Best fits for $c_1^{(\exp)}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectation

$$\boxed{\mathbb{E}[\|\mathrm{Dgm}_1(f_K)\|_0] = n, \quad \mathbb{E}[\|\mathrm{Dgm}_1(f_K)\|_1] = \left(\frac{c}{2} - \frac{1}{4}\right)\sqrt{n}} \tag{A.15}$$

we get from propositions 4.7 and 4.9, respectively. In particular, our experimentally obtained constant $c^{(\exp)} = 2\left(a_0 + \frac{1}{4}\right) \approx 0.6466$ fits with the bounds in (2.22).

<u>FIXED PARAMETERS:</u>
Diagram: domain
Homology: 1^{st}
Norm: 0, 1
Filtration: 0-filt

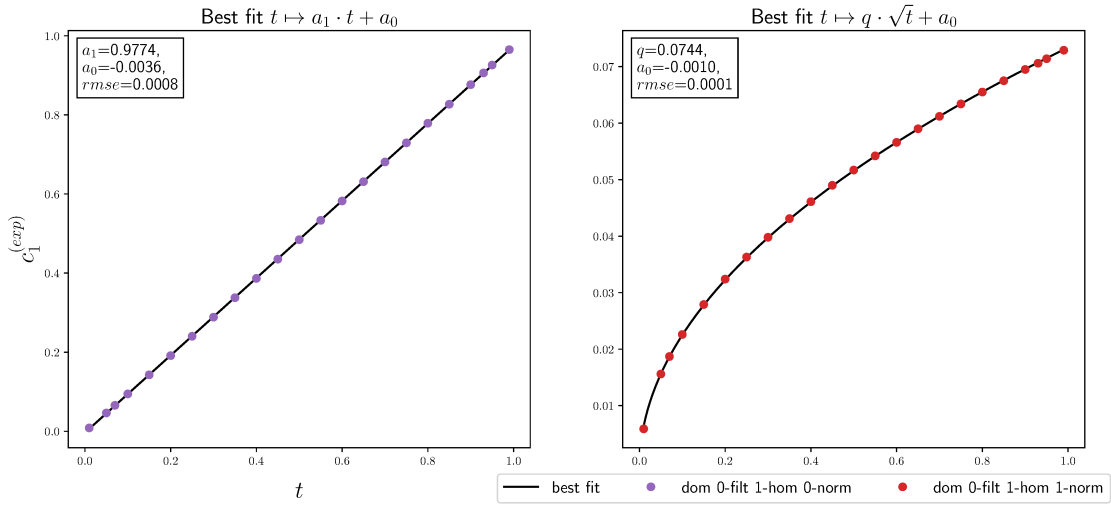

Figure A.15.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectation

$$\boxed{\mathbb{E}[\|\mathrm{Dgm}_1(f_L)\|_0] = tn, \quad \mathbb{E}[\|\mathrm{Dgm}_1(f_L)\|_1] = \left(\frac{c}{2} - \frac{1}{4}\right)\sqrt{t}\sqrt{n}} \tag{A.16}$$

we get from propositions 4.7 and 4.9, respectively. In particular, our experimentally obtained constant $c^{(\mathrm{exp})} = 2\big(q + \frac{1}{4}\big) \approx 0.6488$ fits with the bounds in (2.22).

FIXED PARAMETERS:
Diagram: domain
Homology: 1^{st}
Norm: 0, 1
Filtration: mono-filt

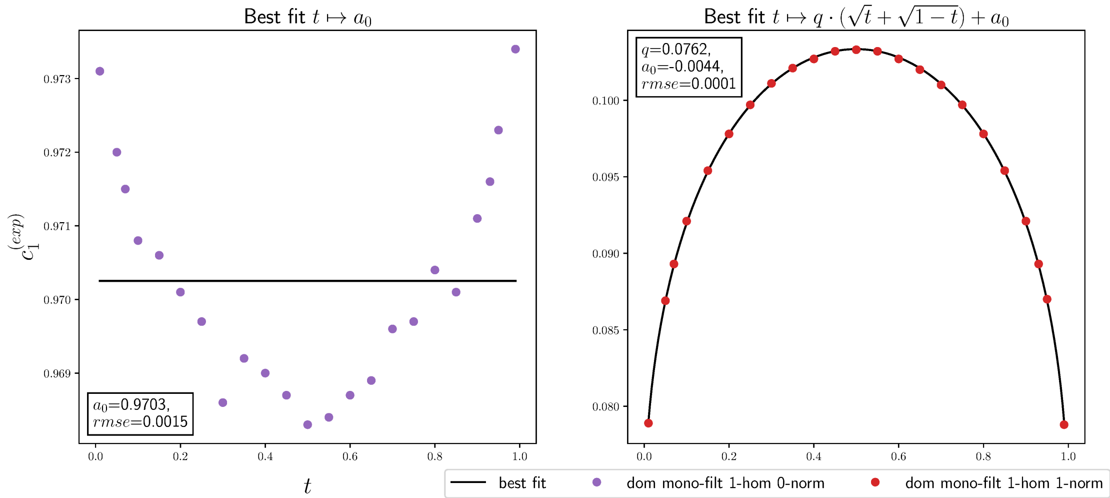

Figure A.16.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectation

$$\boxed{\begin{aligned} \mathbb{E}[\|\mathrm{Dgm}_1(f_L)\|_0] &= n, \qquad &\text{(A.17)} \\ \mathbb{E}[\|\mathrm{Dgm}_1(f_L)\|_1] &= \left(\frac{c}{2} - \frac{1}{4}\right)(\sqrt{t} + \sqrt{1-t})\sqrt{n} \qquad &\text{(A.18)} \end{aligned}}$$

we get from propositions 4.8 and 4.10, respectively. In particular, our experimentally obtained constant $c^{(\mathrm{exp})} = 2\big(q + \frac{1}{4}\big) \approx 0.6524$ fits with the bounds in (2.22).

<u>FIXED PARAMETERS:</u>
Diagram: image
Homology: 1^{st}
Norm: 0, 1
Filtration: 0-filt

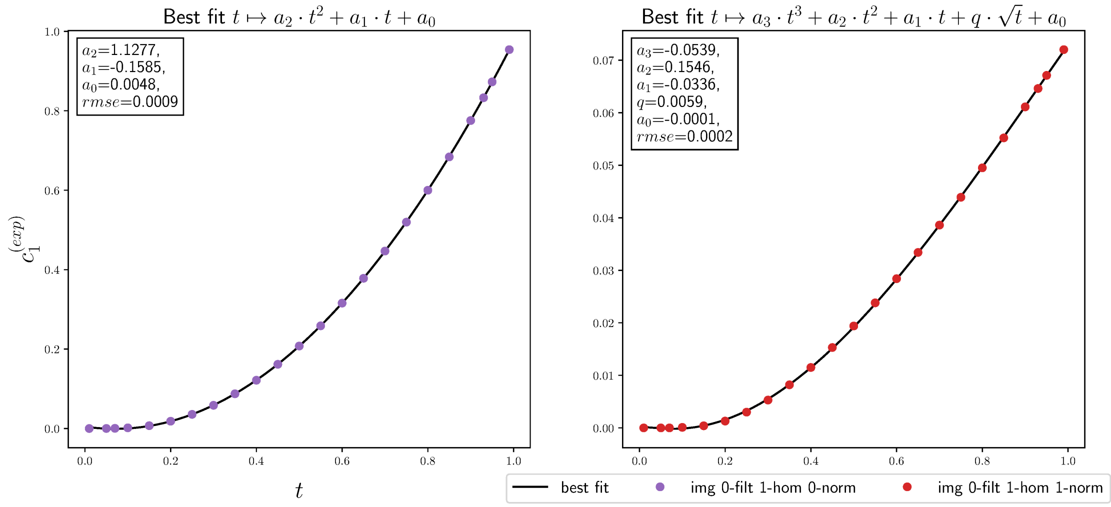

Figure A.17.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

Currently, these experiments do not give a concrete hint on the explicit form of the expectation for neither the 0- nor the 1-norm.

<u>FIXED PARAMETERS:</u>
Diagram: image
Homology: 1^{st}
Norm: 0, 1
Filtration: mono-filt

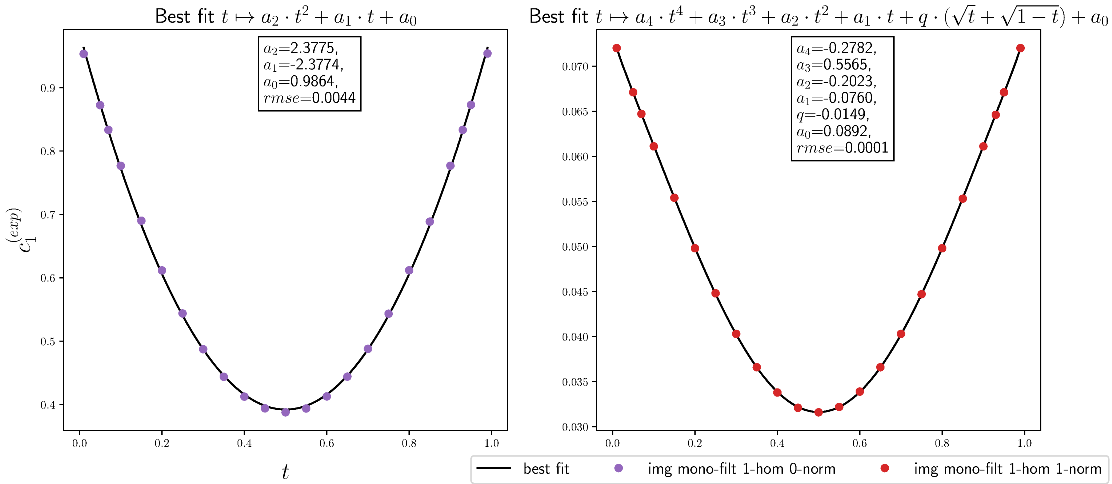

Figure A.18.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

Currently, these experiments do not give a concrete hint on the explicit form of the expectation for the 1-norm. However, at least for the 0-norm we conjecture it to be

$$\mathbb{E}[\|\mathrm{Dgm}_1(\mathrm{im}\, f_L \to f_K)\|_0] = \big(1 - \kappa t(1-t)\big)n \text{ with } \kappa \approx 2.4. \tag{A.19}$$

<u>FIXED PARAMETERS:</u>
Diagram: kernel
Homology: 1^{st}
Norm: 0, 1
Filtration: 0-filt

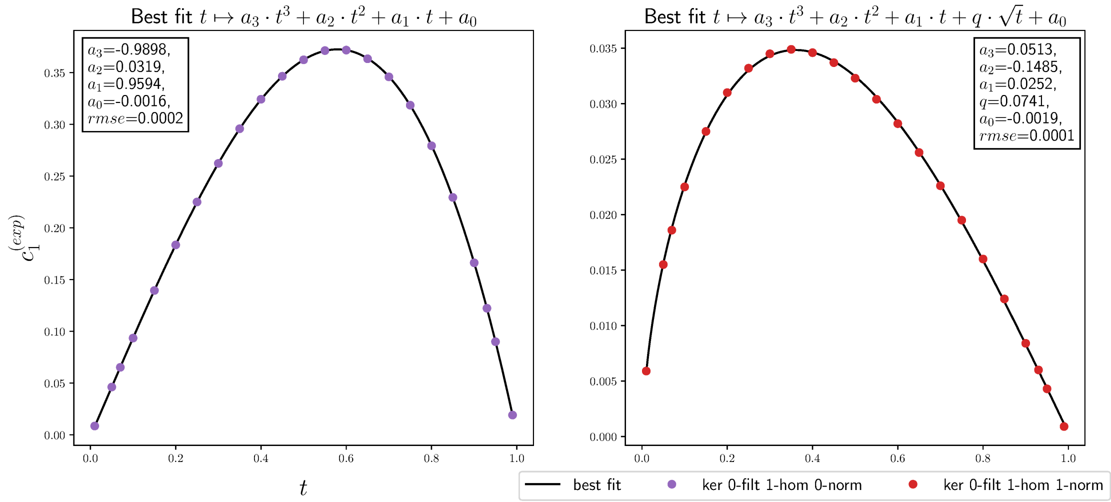

Figure A.19.: Best fits for $c_1^{(\text{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

Currently, these experiments do not give a concrete hint on the explicit form of the expectation for the 1-norm. However, at least for the 0-norm we conjecture it to be

$$\mathbb{E}[\|\mathrm{Dgm}_1(\ker f_L \to f_K)\|_0] = t(1-t)(1+t)n. \tag{A.20}$$

<u>FIXED PARAMETERS:</u>
Diagram: kernel
Homology: 1^{st}
Norm: 0, 1
Filtration: mono-filt

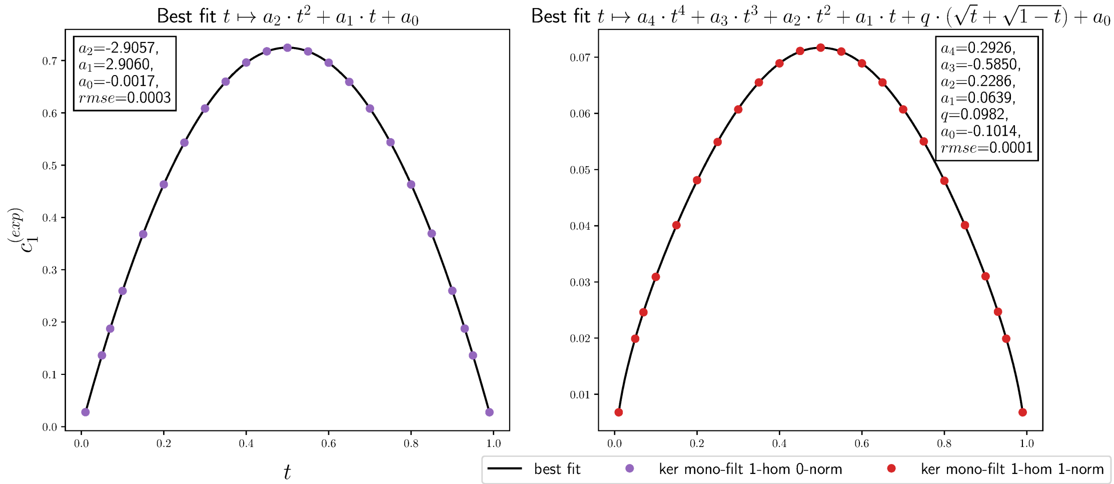

Figure A.20.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

Currently, these experiments do not give a concrete hint on the explicit form of the expectation for the 1-norm. However, at least for the 0-norm we conjecture it to be

$$\mathbb{E}[\|\mathrm{Dgm}_1(\ker f_L \to f_K)\|_0] = 3t(1-t)n. \tag{A.21}$$

<u>FIXED PARAMETERS:</u>
Diagram: cokernel
Homology: 1^{st}
Norm: 0, 1
Filtration: 0-filt

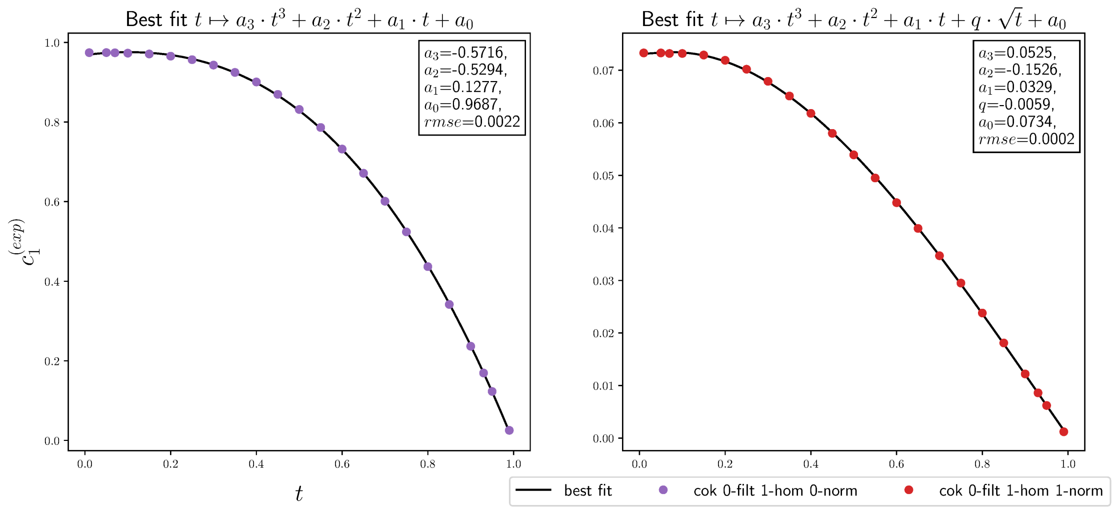

Figure A.21.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

Currently, these experiments do not give a concrete hint on the explicit form of the expectation for neither the 0- nor the 1-norm.

FIXED PARAMETERS:
Diagram: cokernel
Homology: 1^{st}
Norm: 0, 1
Filtration: mono-filt

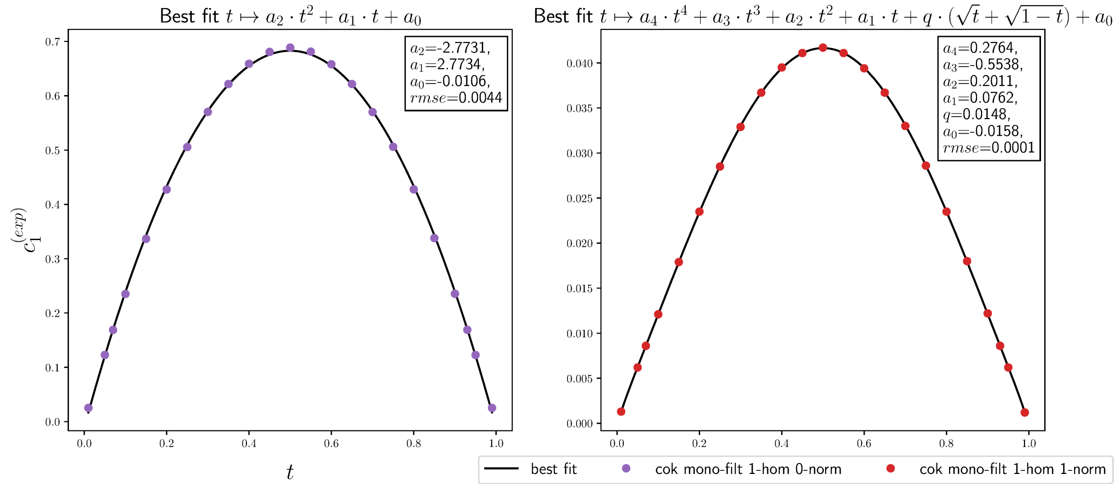

Figure A.22.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

Currently, these experiments do not give a concrete hint on the explicit form of the expectation for the 1-norm. However, at least for the 0-norm we conjecture it to be

$$\mathbb{E}[\|\mathrm{Dgm}_1(\mathrm{coker} f_L \to f_K)\|_0] = \mu t(1-t)n \text{ with } \mu \approx 2.8. \tag{A.22}$$

<u>FIXED PARAMETERS:</u>
Diagram: relative
Homology: 1^{st}
Norm: 0, 1
Filtration: 0-filt

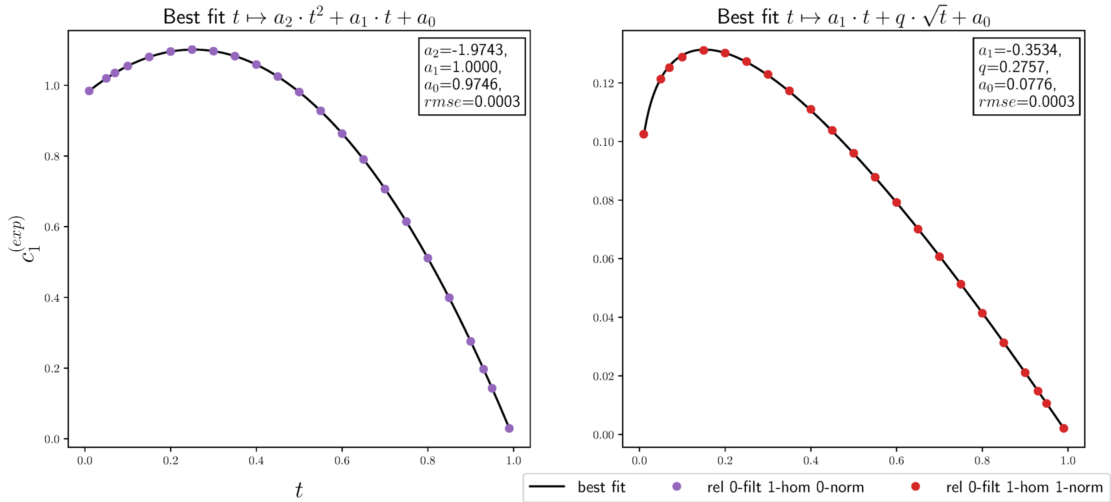

Figure A.23.: Best fits for $c_1^{(\text{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectation

$$\boxed{\mathbb{E}[\|\mathrm{Dgm}_1(f_{L,K})\|_0] = (2t+1)(1-t)n} \tag{A.23}$$

we get from proposition 4.7. For the 1-norm we currently do not have a concrete hint on the explicit form of the expectation.

<u>FIXED PARAMETERS:</u>
Diagram: relative
Homology: 1^{st}
Norm: 0, 1
Filtration: mono-filt

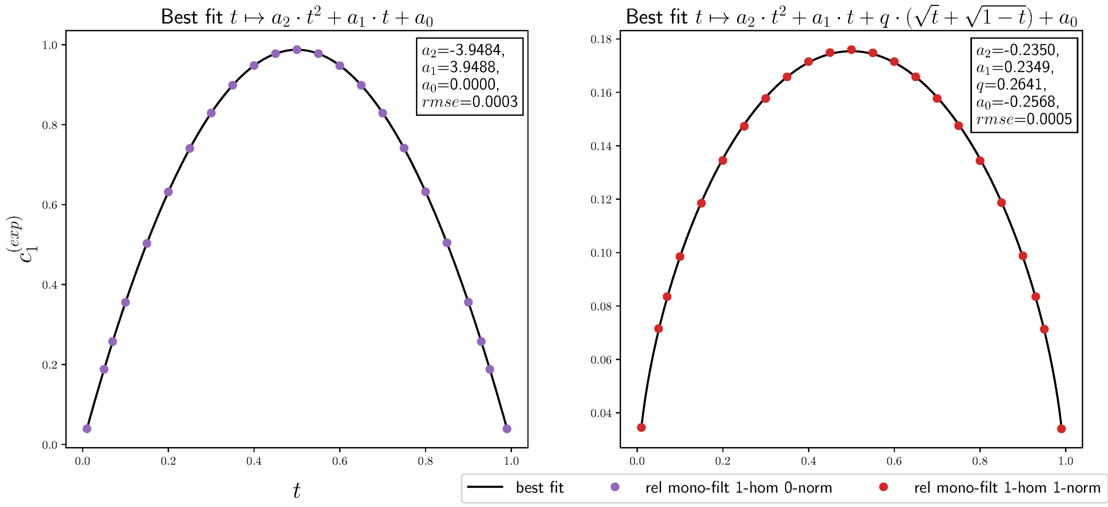

Figure A.24.: Best fits for $c_1^{(\mathrm{exp})}$ when taking the 0-norm (*left*) and the 1-norm (*right*).

These experiments verify the theoretical expectation

$$\boxed{\mathbb{E}[\|\mathrm{Dgm}_1(f_{L,K})\|_0] = 4t(1-t)n} \tag{A.24}$$

we get from proposition 4.8. For the 1-norm we currently do not have a concrete hint on the explicit form of the expectation.

List of Figures

All figures were designed by the author and, apart from the ones in the appendix, with the help of the Software program *Wolfram Mathematica.* [23]
Some of them were created with source code from the specific Wolfram Demonstration Project *Alpha Complex and Union of Union Disks* contributed by Ondrej Draganov. [11]

List of Tables

Bibliography

[1] Florin Avram and Dimitris Bertsimas. "The Minimum Spanning Tree Constant in Geometrical Probability and Under the Independent Model: A Unified Approach". In: *Ann. Appl. Probab.* 2.4 (1992), pp. 113–130.

[2] Jean-Baptiste Bardin, Gard Spreemann, and Kathryn Hess. "Topological exploration of artificial neuronal network dynamics". In: *Network Neuroscience* 3.3 (Jan. 2019), pp. 725–743.

[3] Ulrich Bauer and Herbert Edelsbrunner. "The Morse theory of Čech and Delaunay complexes". In: *Transactions of the American Mathematical Society* 369.5 (Dec. 2016), pp. 3741–3762.

[4] Ulrich Bauer et al. "A unified view on the functorial nerve theorem and its variations". In: *Expositiones Mathematicae* 41.4 (2023), p. 125503.

[5] Ranita Biswas et al. *On the Size of Chromatic Delaunay Mosaics*. 2022. arXiv: `2212.03121 [math.CO]`.

[6] J. A. Bondy and U. S. R. Murty. *Graph Theory with Applications*. New York: Elsevier, 1976.

[7] Helen M. Byrne et al. *Topological Methods for Characterising Spatial Networks: A Case Study in Tumour Vasculature*. 2019. arXiv: `1907.08711 [q-bio.QM]`.

[8] Sebastiano Cultrera di Montesano et al. *Chromatic Alpha Complexes*. 2024. arXiv: `2212.03128 [math.AT]`.

[9] Sebastiano Cultrera di Montesano et al. *The Euclidean MST-ratio for Bi-colored Lattices*. 2024. arXiv: `2403.10204 [cs.CG]`.

[10] Cecil Jose A. Delfinado and Herbert Edelsbrunner. "An incremental algorithm for Betti numbers of simplicial complexes". In: *Proceedings of the Ninth Annual Symposium on Computational Geometry*. SCG '93. San Diego, California, USA: Association for Computing Machinery, 1993, pp. 232–239.

[11] Ondrej Draganov. *Alpha Complex and Union of Growing Disks*. Wolfram Demonstrations Project. Open content licensed under CC BY-NC-SA. 2022. URL: `https://demonstrations.wolfram.com/AlphaComplexAndUnionOfGrowingDisks/`.

[12] Ondrej Draganov and Mohammad Mahini. *chromatic-tda*. URL: `https://github.com/OnDraganov/chromatic-tda`.

[13] Ondřej Draganov et al. *Expected Length of the Euclidean Minimum Spanning Tree and 1-norms of Chromatic Persistence Diagrams in the Plane*. Unpublished paper as of November 2024.

[14] H. Edelsbrunner, D. Letscher, and A. Zomorodian. "Topological persistence and simplification". In: *Proceedings 41st Annual Symposium on Foundations of Computer Science*. 2000, pp. 454–463.

[15] Herbert Edelsbrunner. *A Short Course in Computational Geometry and Topology.* Springer International Publishing, 2014. ISBN: 978-3-319-05956-3.

[16] Herbert Edelsbrunner and John Harer. *Computational Topology - An Introduction.* American Mathematical Society, 2010. ISBN: 978-0-8218-4925-5.

[17] Herbert Edelsbrunner and Teresa Heiss. *Merge Trees of Periodic Filtrations.* 2024. arXiv: `2408.16575 [math.AT]`.

[18] Herbert Edelsbrunner and Ernst P. Mücke. "Three-dimensional alpha shapes". In: *ACM Trans. Graph.* 13.1 (Jan. 1994), pp. 43–72.

[19] Herbert Edelsbrunner, Anton Nikitenko, and Matthias Reitzner. "Expected sizes of Poisson–Delaunay mosaics and their discrete Morse functions". In: *Advances in Applied Probability* 49.3 (2017), pp. 745–767.

[20] Patrizio Frosini. "Measuring shapes by size functions". In: *Intelligent Robots and Computer Vision X: Algorithms and Techniques.* Ed. by David P. Casasent. Vol. 1607. International Society for Optics and Photonics. SPIE, 1992, pp. 122–133.

[21] E. N. Gilbert. "Random Minimal Trees". In: *Journal of the Society for Industrial and Applied Mathematics* 13.2 (1965), pp. 376–387.

[22] Allen Hatcher. *Algebraic Topology.* Cambridge: Cambridge University Press, 2002. ISBN: 0-521-79540-0.

[23] Wolfram Research Inc. *Mathematica, Version 13.2.0.0.* Champaign, IL, 2024. URL: `https://www.wolfram.com/mathematica`.

[24] James R. Munkres. *Elements of Algebraic Topology.* Addison Wesley Publishing Company, 1984. ISBN: 0-201-04586-9.

[25] Vanessa Robins. "Toward computing homology from finite approximations". In: *Topology Proceedings* 24 (1999), pp. 503–532.

[26] M. Saadatfar et al. "Pore configuration landscape of granular crystallization". In: *Nature Communications* 8.15082 (2017).

[27] Rolf Schneider and Wolfgang Weil. *Stochastic and Integral Geometry.* Probability and Its Applications. Springer Berlin Heidelberg, 2008. ISBN: 978-3-540-78859-1.

[28] J. Michael Steele. "Growth Rates of Euclidean Minimal Spanning Trees with Power Weighted Edges". In: *The Annals of Probability* 16.4 (1988), pp. 1767–1787.

[29] R. Stöcker and H. Zieschang. *Algebraische Topologie: Eine Einführung.* Mathematische Leitfäden. Vieweg+Teubner Verlag, 2013. ISBN: 978351912265.

[30] Afra Zomorodian. "Fast construction of the Vietoris-Rips complex". In: *Computers & Graphics* 34 (June 2010), pp. 263–271.

[31] Afra Zomorodian and Gunnar Carlsson. "Computing persistent homology". In: *Proceedings of the Twentieth Annual Symposium on Computational Geometry.* SCG '04. Brooklyn, New York, USA: Association for Computing Machinery, 2004, pp. 347–356.