



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Detecting Cycles in Time Series with Persistent Homology

Master's Thesis

Katharina Pêtre

kpetre@student.ethz.ch

Department of Mathematics D-MATH

Advisor:

Dr. Sara Kališnik Hintz

March 10, 2025

Abstract

Time series are lists of measurements taken from dynamical systems, i.e. systems that change over time according to certain rules. Detecting cyclic motion in time series is insightful, because it tells us that there are recurring patterns in the underlying system. Finding such patterns can help us predict the future evolution of the system.

The aim of this master's thesis is to introduce persistent homology, an important tool in topological data analysis, and in later chapters, to understand and elaborate on a new approach suggested by Bauer et al. in [Bau+23b], which aims to detect cycles in time series data using persistent homology, and more importantly, to classify cycles that behave similarly. Cycling signatures capture the different types of cycles that occur within segments of a given time series. The core work done for this thesis is the study of how to move from the rather abstract definition of cycling signatures to more combinatorial objects, do the computation there and come back without losing significance and accuracy.

Acknowledgements

First and foremost, I would like to express my sincere appreciation and gratitude to Dr. Sara Kališnik Hintz for her insight and support throughout the entire process of writing this thesis.

I am grateful to my family and friends for helping me keep my cool during this intense semester, and especially to my friend Renato Farruggio, who is a computer scientist and helped me get started with some of the more computational work for this thesis.

Finally, I would like to thank ETH Zurich for giving me the opportunity to attend a number of excellent mathematics courses during my master's studies and for providing the framework that made this thesis possible.

Abbreviations and Acronyms

The following list contains abbreviations and acronyms used in this thesis.

\mathbb{R}^+	The set of positive real numbers.
\mathbb{R}_0^+	The set of non-negative real numbers.
\mathbb{N}	The set of positive integers.
\mathbb{N}_0	The set of non-negative integers.
$\#X$	The size of the set X .
$\mathcal{P}(X)$	The power set of X , i.e. the set of all subsets.
\overline{X}	The closure of the set X , i.e. the smallest closed set containing X .
X°	The interior of the set X , i.e. the largest open set contained in X .
∂X	The boundary of the set X , i.e. $\partial X = \overline{X} \setminus X^\circ$.
\subseteq	An inclusion which can be an equality.
\subset	A strict inclusion.
\cong	An isomorphism (mostly of vector spaces).
\simeq	A homeomorphism of topological spaces.
\sim	An equivalence relation.
$[x], \bar{x}$ or $[x]_\sim$	The equivalence class of an element x with respect to an equivalence relation \sim .
$[v_0, \dots, v_k]$	The geometric k -simplex spanned by the vertices v_0, \dots, v_k .
$\text{VR}(X, r)$	The Vietoris-Rips complex with vertex set X and radius r .
$\check{C}(X, r)$	The Čech complex with vertex set X and radius r .
$ \mathcal{K} $	The geometric realisation of the simplicial/cubical complex \mathcal{K} .
$B(p, r)$ or $B_2(p, r)$	The open ball of radius r around p in \mathbb{R}^d with respect to Euclidean metric.
$U(X, r)$	The thickening of the set X by r .
$B_\infty(p, r)$	The open ball of radius r around p with respect to the ∞ -metric.
S^d or S_2^d	The sphere of dimension d with respect to Euclidean metric (embedded in \mathbb{R}^{d+1}).
S_∞^d	The sphere of dimension d with respect to the ∞ -metric.

Contents

Abstract	i
Acknowledgements	ii
Abbreviations and Acronyms	iii
1 Introduction	1
2 Simplicial Complexes, Cubical Complexes and Homology	7
2.1 Simplicial Complexes	7
2.2 Barycentric Subdivision	9
2.3 Simplicial Homology	14
2.4 Cubical Complexes and Homology	21
3 Metric Spaces and Persistent Homology	25
3.1 Metric Spaces to Simplicial Complexes	25
3.2 Čech Complexes and Thickenings	26
3.3 Persistent Homology and its Representation	29
3.4 Distances between Persistence Diagrams	44
4 Differential Geometry, Dynamical Systems and Time Series	48
4.1 Smooth Manifolds and Maps	48
4.2 Tangent Spaces	51
4.3 Dynamical Systems	54
4.4 Time Series	57
5 Cycling Signatures	60
5.1 Finding a Good Metric	60
5.2 Cycling Segments and Classification	61
5.3 Internal and External Thickenings	65
5.4 Construction of Comparison Spaces	68
5.5 Choice of the Parameters r_0 and C	69
5.6 Acyclic Covers and Carriers	70

5.7 Summarised Computation Pipeline	78
Bibliography	81
Declaration of originality	83

Introduction

Data often come in the form of time series, i.e. a sequence of measurements, usually recorded at equidistant time stamps. More precisely, they are sampled from dynamical systems, which means systems that evolve over time according to a set of rules. For example, time series can be lists of weather data, stock prices, brain monitoring measurements and so on. With time series, the goal is to use the information it contains to construct a model that approximates the underlying dynamical system. In this thesis, we assume that the time series come from dynamical systems (X, Φ) , where $X \subseteq \mathbb{R}^d$ is a subset of Euclidean space and $\Phi : \mathbb{R} \times X \rightarrow X$ is a smooth map describing how the system changes over time. This means that for any point x in X , we can trace the evolution $\Phi(t, x)$ of x with the flow Φ over time t . The word "flow" indicates that if we track x starting from $t = 0$ and observe its position after some time T , the result is the same as if we first measured its position at an intermediate time step s (where $s \in [0, T]$) and then measured its movement from s to T .

Assuming our time series comes from a dynamical system, we are interested in finding out whether the system contains any cyclic motion and if yes, what type of cyclic motion. For example, the plot in Figure 1.1 shows the evolution of temperatures in Basel (75 km from Zurich) over the past ten years. Clearly, there is some periodicity present. The temperature always falls in the winter months and rises again in the summer. Identification of such trends helps us to predict the future evolution of the dynamical system. In the temperature example, we can now say with high confidence that it will be cold next winter as well.

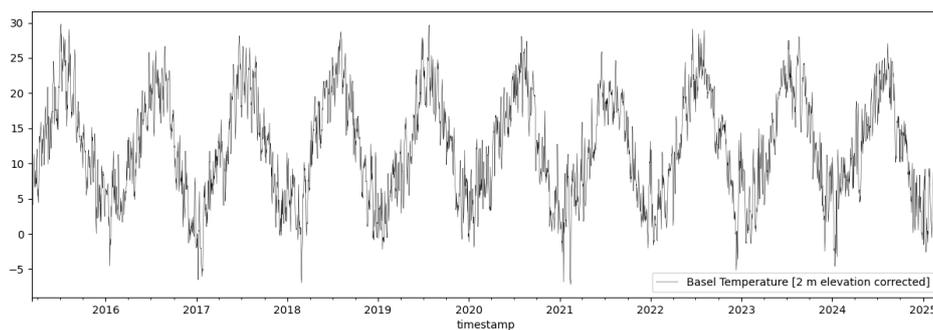


Figure 1.1: Temperature measurements (daily average) in Basel (75 km from Zurich) from March 2015 to March 2025. The data can be freely accessed with the following link: <https://www.meteoblue.com/de/wetter/archive/export>.

A mathematical example of a time series coming from a dynamical system is illustrated in Figure 1.2. In mathematics, dynamical systems often appear in the

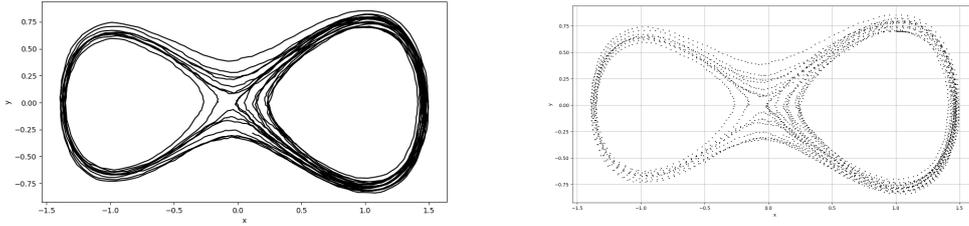


Figure 1.2: Double well dynamical system and time series.

form of solutions to differential equations. The dynamical system from Figure 1.2 comes as the solution of the stochastic differential equation

$$dx = f(x)dt + \sigma dW,$$

where W is a d -dimensional stochastic process called Wiener process and σ is a parameter which controls the importance of the randomness introduced by W . This type of differential equation models systems from a range of different areas. Example include dynamics in chemical reactions [JLR20] and neural dynamics [Mel19]. As can be seen in Figure 1.2, the time series includes two obvious voids (hence the name). The results studied in this thesis will allow us to identify the cycles that go around the voids, and will also make it possible to distinguish between the cycles that go around the left void, those that go around the right void, and the longer segments that go around both. To detect and classify those cycles, we will use objects from the field of algebraic topology called **homology groups**.

Homology groups capture information about loops and higher-dimensional voids in topological spaces. The numbers that count these loops and voids are called **Betti numbers**. For example, the first Betti number of a circle is one, because a circle consists of one 1-dimensional loop. The problem with time series is that they are only finite collections of points. This means that we cannot use a time series as is to compute Betti numbers, since a collection of points has no interesting topological properties.

One way to approach the above-mentioned problem is to construct **thickenings** (Definition 3.5). Let X be a collection of points in \mathbb{R}^d . For example, this could be a sample from a dynamical system. Then

$$U(X, r) = \bigcup_{p \in X} B(p, r/2)$$

is the thickening of X by the radius $r/2$, where $B(p, r/2)$ is the open ball of radius $r/2$ around the point p . As can be observed in Figure 1.3, the radii of these thickenings strongly influence which topological features (voids) can be detected. Persistent homology will allow us to collect the topological features that occur for a range of different radii. To compute homology of the thickenings, we show (Theorem 3.6) that the closure of a thickening $U(X, r)$ is topologically equivalent to the more combinatorial Čech complex $\check{C}(X, r)$.

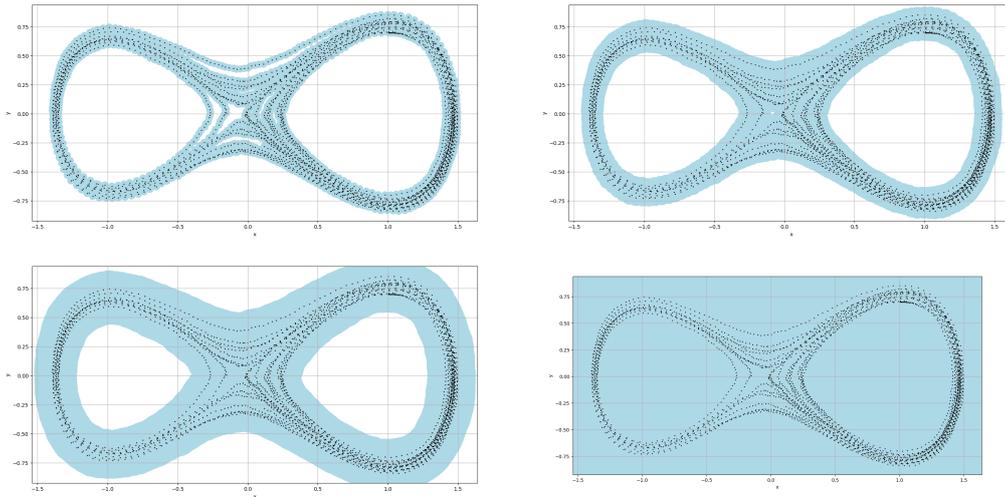


Figure 1.3: Double well time series with thickenings for increasing radii.

A Čech complex is a **simplicial complex**, i.e. a collection of simplices - convex hulls of points - glued together along common boundaries. So, 0-simplices are points, 1-simplices are edges, 2-simplices are triangles, 3-simplices are tetrahedra and for larger dimensions k , k -simplices are their k -dimensional counterparts. The simplex $[p_0, \dots, p_k]$ spanned by the points $p_0, \dots, p_k \in X$ is a simplex of the Čech complex $\check{C}(X, r)$ (Definition 3.4) if and only if

$$\bigcap_{i=0}^k B(p_i, r/2) \neq \emptyset.$$

A picture of a Čech complex and the corresponding thickening can be seen in Figure 1.4.

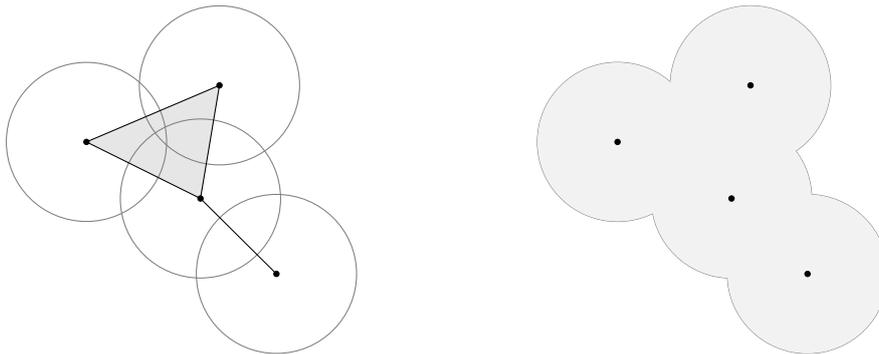


Figure 1.4: A Čech complex and the corresponding thickening of points.

Computing persistent homology means taking a collection of nested simplicial complexes, and recording the radii at which a topological feature is born and at which it dies. Looking at Figure 1.3 once again, there are many small voids that appear in the first thickening, but already disappear in the second. On the other hand, the

right and left loops are born at around the same time, but only disappear at a much larger radius. The observation that they persist over the largest range of radii hints at the fact that they are the most significant topological features in the data.

Another question that arises when looking at time series data, is how to include the direction of the flow of the system into the computation. As we can see in Figure 1.5, the middle points merge in the thickening even though the flow goes in opposite directions. Bauer et al. address this issue in [Bau+23b] by using a different distance than the standard Euclidean metric for the balls $B(p, r/2)$ in the definition of the thickenings. We add a directional component to each point by considering the tangent vectors with respect to the flow Φ from the dynamical system. The map that formalises this is

$$\rho: X \setminus X_{fix} \longrightarrow UTX, x \longmapsto \left(x, \frac{v(x)}{\|v(x)\|_2} \right),$$

where X_{fix} collects all fixed points of the dynamical system (X, Φ) , i.e. points that do not move at all over time, UTX is the set containing all tangent vectors of length 1 for all points in X , and

$$v(x) = \frac{d}{dt}\Phi(t, x)\Big|_{t=0}$$

is the tangent vector induced by the flow Φ at x . In other words, we do not just look at the points $x \in X$ from the data per se, but we also take into account the direction towards the following points in the time series, which roughly models the direction of the flow at x . For the new thickenings $U(\rho(\gamma), r)$ of segments γ of the time series, this means that two points going in very different directions are now further away with respect to the new distance, even if they look like they are close in space. We will look at this construction in more detail in Section 5.1.

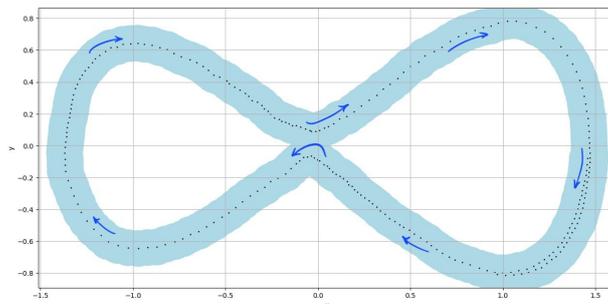


Figure 1.5: Short subsequence of the time series from Figure 1.3, thickened and with arrows to indicate direction of the flow.

One more thing to think about is how to distinguish between different types of cycles. As we can infer from Figure 1.3, there are three different types of cycles at first glance. The left one, the right one and one going around both. This distinction is made by looking at the thickening as a subset of a larger space Y called a **comparison space**. In practice, this comparison space is a collection of cubes Y_∞ (squares in

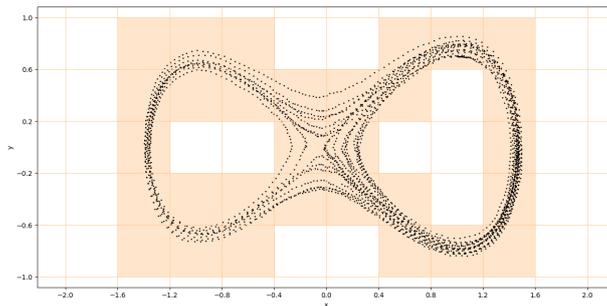


Figure 1.6: Sketch of a comparison space (orange squares) for the time series from Figure 1.3. The representation is not strictly accurate, as the directional component of the time series is not taken into consideration.

dimension two, lines in dimension one and points in dimension zero). A visualisation of a comparison space for the double well time series can be found in Figure 1.6.

The comparison space already contains the two holes from the double well system. If we look at the thickenings as a subspace of this comparison space, the different cycles have different generators (from the comparison space) in homology. This is the core idea of the cycling signatures introduced by Bauer et al. in [Bau+23b]. The **cycling signature** of a time series segment is the image of the map in homology induced by this inclusion. This translation of maps is justified by the fact that homology is a functor, as we show in Theorem 2.29.

For the cycling signatures to be computed efficiently, we perform a number of tweaks to enhance the performance. The detour we take is summarised in the commutative diagram

$$\begin{array}{ccc}
 \{H_1(U(\rho(\gamma), r))\}_{r \geq 0} & \xrightarrow{H_1(i_\gamma)} & H_1(Y) \\
 \cong \uparrow & & \uparrow \cong \\
 \{H_1(\check{C}(\rho(\gamma), r))\}_{r \geq 0} & \xrightarrow{H_1(\phi)} & H_1(Y_\infty),
 \end{array} \tag{1.1}$$

where $\rho(\gamma)$ is a short piece of a time series with added tangent vectors, and the Čech complex $\check{C}(\rho(\gamma), r)$ is constructed with the adapted metric defined on $\rho(\gamma)$. The parameter $r \geq 0$ indicates that we are working with persistent homology groups, so we capture the homology groups for all r within a predetermined range and observe when and how they change. The left arrow is a result of the topological equivalence between thickenings and Čech complexes, and the right arrow essentially relates round balls and cubes. The top map is the translation (using functoriality of homology, Theorem 2.29) of the above-mentioned inclusion of the time series into a comparison space Y , and the map on the bottom is the computational approach with the algorithm provided in [Bau+23b].

In the algorithm, we first compute the homology of the cubical comparison space

Y_∞ and decide on a collection of segments of the time series for which we check if they contain any cycles. For these segments, we construct simplicial complexes similar to Čech complexes for increasing radii and compute persistent homology. Lastly, we find a basis of the image of $H_1(\phi)$ from the above diagram by performing a subdivision of the simplicial complex and using a clever construction which is roughly sketched in Figure 1.7 and will be explained in more detail in Section 5.7.

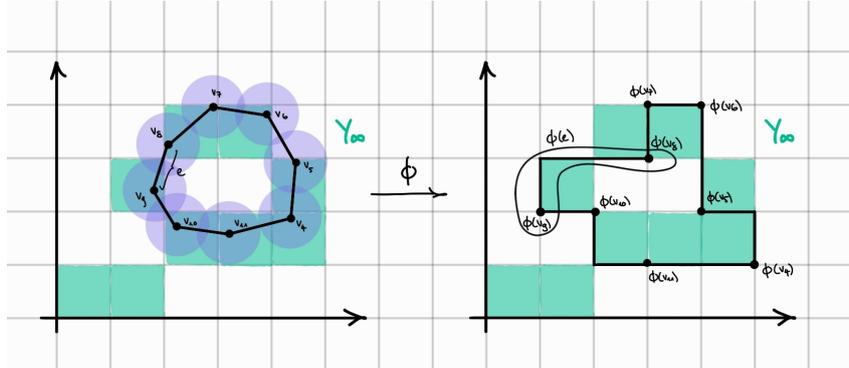


Figure 1.7: Sketch of an example for the map ϕ used for the commutative diagram 1.1.

In the sketch from Figure 1.7, the cycling signature we obtain for this specific radius is the space $H_1(Y_\infty)$ itself. Going back to the double well system, the algorithm computes three distinct cycling signatures of rank 1. They are represented with three different colours in Figure 1.8.

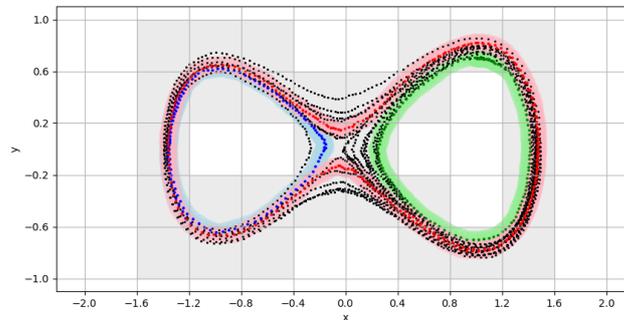


Figure 1.8: Double well time series inside a cubical comparison space and with different colours indicating the three different rank 1 cycling signatures.

For a segment going once around the left void (blue in the figure), we obtain a trivial cycling signature as long as the radius for the construction of the Čech complex is too small to connect all the dots, and then get a rank 1 signature generated by the generator of $H_1(Y_\infty)$ corresponding to the left void. The same argument holds for segments going around the right void once, which are highlighted in green in the figure. The cycling signature of longer segments going around both voids (coloured in red in Figure 1.8) have their cycling signature generated by the sum of the generators of $H_1(Y_\infty)$.

Simplicial Complexes, Cubical Complexes and Homology

2.1 Simplicial Complexes

Simplicial complexes are constructions consisting of points, lines, triangle and higher-dimensional polytopes glued together along common boundaries. They can be used to describe many topological spaces in a combinatorial way, which makes it easier to compute certain topological invariants (like homology, which we introduce in Section 2.3). We follow chapter 2 of [Car14], as well as [EH10] and [Hat01].

If we draw three points on a plane, we want to avoid the case in which they all lie on the same line. This will be important in the definition of simplices, as it ensures that the set spanned by $k + 1$ points has dimension k , and not a lower dimension. For our three points, this means connecting them and filling in the resulting triangle gives us a surface, so a 2-dimensional object.

Definition 2.1 (General Position). *Let $S \subset \mathbb{R}^d$ be a finite set. Then S is in **general position**, if it is not contained in a hyperplane $H \subseteq \mathbb{R}^d$ of dimension $\dim H \leq \#S - 1$.*

Going forward, we assume our finite sets in \mathbb{R}^d to be in general position. The geometric simplices introduced in the next definition are the building blocks which we will use for the construction of simplicial complexes.

Definition 2.2 (Geometric Simplex).

(a) *Let $S = \{x_0, \dots, x_k\}$ be a finite subset of \mathbb{R}^d . The **convex hull** of the set S is given by*

$$\sigma(S) = \left\{ \sum_{i=0}^k \lambda_i x_i : x_i \in S, 0 \leq \lambda_i \leq 1 \text{ for all } i \text{ and } \sum_{i=0}^k \lambda_i = 1 \right\}.$$

(b) *Let $S \subset \mathbb{R}^d$ be a finite set in general position. Then the **geometric simplex spanned by S** is the convex hull of S in \mathbb{R}^d . Write $\sigma = \sigma(S)$ or $[x_0, \dots, x_k]$ for $S = \{x_0, \dots, x_k\}$. The points x_i , $i = 0, \dots, k$ are called **vertices** and the simplices generated by subsets of S are called **faces** of σ . A simplex spanned by a set containing $k + 1$ points is called a **k -simplex**, as it has dimension k .*

(c) *The **standard k -simplex** is the simplex*

$$\Delta^k = \left\{ (t_0, \dots, t_k) \in \mathbb{R}^{k+1} : \sum_{i=0}^k t_i = 1 \right\}.$$

In other words, it is the k -simplex spanned by all points of the form

$$(0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^{k+1}.$$

(d) Let σ be a k -simplex. Then its **boundary** is the set containing all $(k - 1)$ -simplices which are faces of σ .

In Figure 2.1 we illustrate what 0- 1- and 2-simplices looks like.

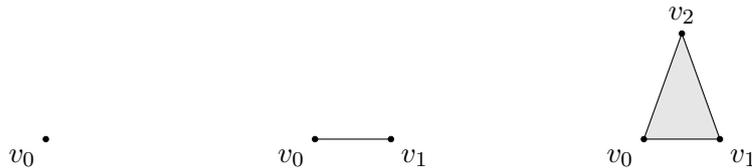


Figure 2.1: A geometric 0-simplex or vertex v_0 , a 1-simplex or edge $[v_0, v_1]$ and a 2-simplex or triangle $[v_0, v_1, v_2]$.

When we glue geometric simplices together along their boundaries, we get a so-called geometric simplicial complex. The precise rules for this construction are given by the following definition.

Definition 2.3 (Geometric Simplicial Complex). A *geometric simplicial complex* is a finite collection \mathcal{K} of geometric simplices σ such that if $\sigma \in \mathcal{K}$, then all faces of σ are also contained in \mathcal{K} . Moreover, for two simplices σ, τ in \mathcal{K} , their intersection must also be a simplex, which is a face both of σ and τ . We use the following notation.

- $V(\mathcal{K})$ is the set of vertices of \mathcal{K} (i.e. 0-dimensional simplices).
- $\Sigma_k(\mathcal{K})$ is the collection of k -simplices of \mathcal{K} .
- The k -**skeleton** of \mathcal{K} is the subcomplex

$$\mathcal{K}^{(k)} = \bigcup_{i \leq k} \Sigma_i(\mathcal{K})$$

containing all simplices of \mathcal{K} that have dimension up to k .

The topology on a geometric simplicial complex is the quotient topology obtained by identifying the boundary of each k -simplex with the corresponding $(k - 1)$ -simplices. We look at the simplices as subspaces of \mathbb{R}^d .

Example 2.4. Consider the constructs in Figure 2.2. The leftmost construction is a geometric simplicial complex \mathcal{K} with vertices $V(\mathcal{K}) = \{v_1, \dots, v_5\}$, edges $\Sigma_1(\mathcal{K}) = \{[v_1, v_2], [v_1, v_3], [v_2, v_3], [v_3, v_4]\}$ and one triangle $\Sigma_2 = \{[v_1, v_2, v_3]\}$. The second construct is not a geometric simplicial complex, as the edge $[v_2, v_3]$ of the triangle $[v_1, v_2, v_3]$ is not included in the set of simplices. The third construct is no geometric simplicial complex either, as the edge set is $\{[v_1, v_2], [v_1, v_3], [v_2, v_6], [v_6, v_3], [v_3, v_4]\}$, so again, the face $[v_2, v_3]$ of the triangle $[v_1, v_2, v_3]$ is missing. Moreover, the vertex v_6 is in the intersection $[v_2, v_6] \cap [v_1, v_2, v_3]$ but is not a face of $[v_1, v_2, v_3]$.

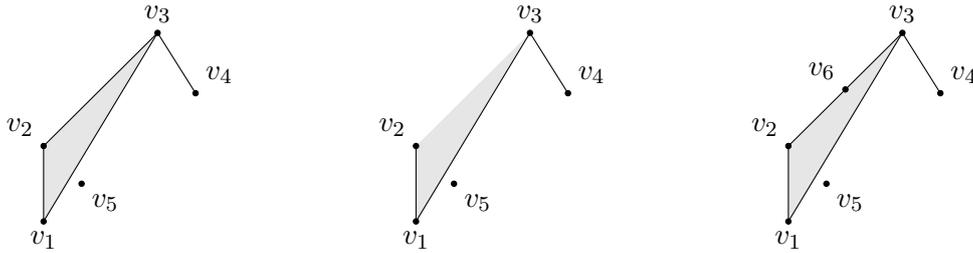


Figure 2.2: On the left, we see an example of a geometric simplicial complex. The two other constructs are non-examples of geometric simplicial complexes.

Geometric simplicial complexes are bound by their position in \mathbb{R}^d . We want to identify all geometric simplicial complexes which are homeomorphic, or more intuitively, for which there is a bijection between the set of all simplices which preserves faces.

Definition 2.5 (Abstract Simplicial Complex). An *abstract simplicial complex* is a pair $\mathcal{K} = (V(\mathcal{K}), \Sigma(\mathcal{K}))$ consisting of a finite set of **vertices** $V(\mathcal{K})$ and a collection of **simplices** $\Sigma(\mathcal{K}) \subseteq \mathcal{P}(V(\mathcal{K})) \setminus \{\emptyset\}$, such that if $\sigma \in \Sigma(\mathcal{K})$ and $\tau \subseteq \sigma$, $\tau \in \Sigma(\mathcal{K}) \setminus \{\emptyset\}$, then $\tau \in \Sigma(\mathcal{K})$.

A *geometric realisation* $|\mathcal{K}|$ of an abstract simplicial complex \mathcal{K} is a geometric simplicial complex which arises from assigning the vertices $V(\mathcal{K})$ points in Euclidean space \mathbb{R}^d and mapping every abstract simplex to the geometric simplex spanned by the points corresponding to its vertices.

From now onwards, when we write "simplicial complex" we mean "abstract simplicial complex". When we think of maps of simplicial complexes, it makes sense to expect that such a function should map a simplex to another simplex. The following definition formalises this idea.

Definition 2.6 (Simplicial Map). Let \mathcal{K} and \mathcal{L} be simplicial complexes, and let $f: \mathcal{K} \rightarrow \mathcal{L}$ be a map. f is called *simplicial*, if for any simplex $\sigma \in \mathcal{K}$, $f(\sigma)$ is a simplex of \mathcal{L} .

2.2 Barycentric Subdivision

The most commonly studied subdivisions of simplicial complexes are barycentric subdivisions. In particular, they allow us to define coordinates for all points in a geometric simplicial complex using only the vertices. Moreover, they play an important role in the proofs of some fundamental results in the field algebraic topology, like the Excision theorem and Mayer-Vietoris sequences (see [Hat01] if interested). In this section, we use theory from [Bau+23a], [Wof18] and [Hat01].

Definition 2.7 (Barycentric Subdivision and Star). Let \mathcal{K} be a simplicial complex and let $|\mathcal{K}|$ be its geometric realisation.

- (a) A *simplicial subdivision* or *refinement* of \mathcal{K} is another simplicial complex \mathcal{L} , such that $|\mathcal{K}| = |\mathcal{L}|$, and every simplex of \mathcal{L} is contained in a simplex of \mathcal{K} .

(b) Let $\sigma = [v_0, \dots, v_k]$ be a simplex of \mathcal{K} . The **barycenter** of σ is

$$b(\sigma) = \frac{1}{k+1} \sum_{i=0}^k v_i.$$

(c) The **barycentric subdivision** of \mathcal{K} is the simplicial subdivision $\text{sd}\mathcal{K}$ with vertices

$$V(\text{sd}\mathcal{K}) = \{b(\sigma) : \sigma \in \Sigma(\mathcal{K})\}$$

and simplices

$$\Sigma(\text{sd}\mathcal{K}) = \{[b(\sigma_0), \dots, b(\sigma_l)] : \sigma_0 \subset \dots \subset \sigma_l \text{ in } \mathcal{K}\}.$$

(d) The **(closed) barycentric star** of a vertex v of \mathcal{K} is the subcomplex

$$\text{bst}(v, \mathcal{K}) = |\{\tau \in \text{sd}\mathcal{K} : \tau \cup \{v\} \in \text{sd}\mathcal{K}\}| \subseteq |\text{sd}\mathcal{K}|.$$

In Figure 2.3, we take the simplices from Figure 2.1 and add their barycentre to illustrate Definition 2.7.

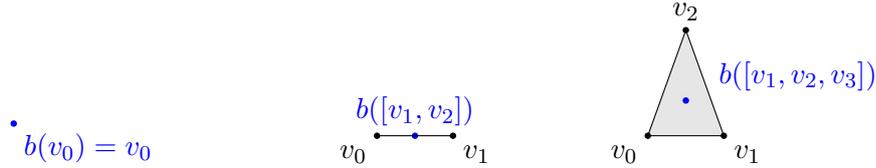


Figure 2.3: The same simplices as in Figure 2.1 with their respective barycentres added as blue dots.

Remark 2.8. *Intuitively, the barycentric star of a vertex v in \mathcal{K} is the union of all simplices of $|\text{sd}\mathcal{K}|$ which contain the vertex v , plus all of their faces (this is the reason why it is closed). Moreover, the collection of all barycentric stars gives a closed cover of $|\text{sd}\mathcal{K}|$.*



Figure 2.4: A simplicial complex and its barycentric subdivision. The simplices highlighted in blue are a visualisation of the closed barycentric star of v .

An example of a barycentric subdivision and star can be found in Figure 2.4. The following lemma tells us that the barycentric stars $\text{bst}(v, \mathcal{K})$ form a so-called good cover of $|\mathcal{K}|$.

Lemma 2.9. *Let σ be a simplex of a simplicial complex \mathcal{K} . Then the intersection*

$$\bigcap_{v \in \sigma} \text{bst}(v, \mathcal{K})$$

is contractible.

Proof. Let $\sigma = [v_0, \dots, v_k]$ be a simplex of \mathcal{K} . We consider the subcomplex \mathcal{L} of $\text{sd}\mathcal{K}$ which contains all simplices $[b(\sigma_0), \dots, b(\sigma_l)]$ for which

$$\sigma \subseteq \sigma_0 \subset \dots \subset \sigma_l.$$

Let $\tau = [b(\sigma_0), \dots, b(\sigma_l)]$ be such a simplex. Then τ is contained in the simplex $[b(\sigma), b(\sigma_0), \dots, b(\sigma_l)]$ of $\text{sd}\mathcal{K}$, thus $|\tau|$ is contained in $\text{bst}(v_i, \mathcal{K})$ for all v_i . In particular, this implies

$$|\mathcal{L}| \subseteq \bigcap_{i=0}^k \text{bst}(v_i, \mathcal{K}).$$

On the other hand, let $\tau = [b(\sigma_0), \dots, b(\sigma_l)]$ be a simplex of $\text{sd}(\mathcal{K})$ such that

$$|\tau| \subseteq \bigcap_{i=0}^k \text{bst}(v_i, \mathcal{K}).$$

Then $|\tau| \subseteq \text{bst}(v_i, \mathcal{K})$ for all $i \in \{0, \dots, k\}$. This implies $v_i \in \sigma_0$ for all i . Therefore, $\sigma = [v_0, \dots, v_k] \subseteq \sigma_0$, which in turn implies that $\tau \in \mathcal{L}$, so we get

$$\bigcap_{i=0}^k \text{bst}(v_i, \mathcal{K}) \subseteq |\mathcal{L}| \quad \text{and so,} \quad \bigcap_{i=0}^k \text{bst}(v_i, \mathcal{K}) = |\mathcal{L}|.$$

By definition of \mathcal{L} and the previous argument, any geometric simplex $|\tau| \subseteq \bigcap_{v \in \sigma} \text{bst}(v, \mathcal{K})$ contains $|\sigma|$, so it also contains $|b(\sigma)|$ by convexity of simplices. Thus, $\bigcap_{v \in \sigma} \text{bst}(v, \mathcal{K})$ is star-shaped with respect to $|b(\sigma)|$, which implies that it is contractible. \square

Definition 2.10 (Barycentric Coordinates). *Let \mathcal{K} be a simplicial complex and let $\sigma = [v_0, \dots, v_k] \in \Sigma_k(\mathcal{K})$ be a k -simplex.*

(a) *Consider the linear homeomorphism*

$$\beta: \Delta^k \longrightarrow |\sigma|, \quad (t_0, \dots, t_k) \longmapsto \sum_{i=0}^k t_i |v_i|.$$

*The tuple (t_0, \dots, t_k) specifies the **barycentric coordinates** of $x = \sum_{i=0}^k t_i |v_i|$ with respect to σ .*

(b) *Let v_i be a vertex of \mathcal{K} . We define the map*

$$b_{v_i}: |\mathcal{K}| \longrightarrow [0, 1], \quad x \longmapsto \begin{cases} 0 & |v_i| \text{ is not a vertex of the smallest} \\ & \text{geometric simplex containing } x \\ t_i & \text{else,} \end{cases}$$

where t_i is the barycentric coordinate of x corresponding to the vertex v_i (see (b)).

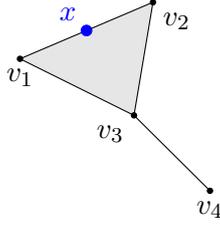


Figure 2.5: A simplicial complex with a highlighted point $v \in [v_1, v_2]$ contained in it.

The above enables us to write any point contained in \mathcal{K} as a linear combination of vertices of \mathcal{K} . More precisely, let x be a point in $|\mathcal{K}|$ and let $|\sigma| = |[v_0, \dots, v_k]|$ be the smallest geometric simplex of $|\mathcal{K}|$ containing x . Then we can write

$$x = \sum_{v \in V(\mathcal{K})} b_v(x)|v| = \sum_{i=0}^k t_i|v_i|.$$

Example 2.11. Consider the simplicial complex \mathcal{K} in Figure 2.5. The smallest geometric simplex of $|\mathcal{K}|$ containing the blue point x is $[[v_1, v_2]]$. So, we can write

$$x = b_{v_1}(x)|v_1| + b_{v_2}(x)|v_2|,$$

and $b_{v_i}(x) = 0$ for all $i \neq 1, 2$.

Lemma 2.12. Let v be a vertex of a simplicial complex \mathcal{K} . Then

$$\text{bst}(v, \mathcal{K}) = \{x \in |\mathcal{K}| : b_v(x) \geq b_w(x) \text{ for all } w \in V(\mathcal{K})\}.$$

Proof. Let x be a point in the geometric realisation $|\mathcal{K}|$ of \mathcal{K} . The first step in the proof is to find a way to switch from the barycentric coordinates of x with respect to the geometric complex $|\mathcal{K}|$ to those with respect to the barycentric subdivision $|\text{sd}\mathcal{K}|$. Let $w_i, i = 0, \dots, m$ be the vertices of the smallest simplex of \mathcal{K} containing x . We assume without loss of generality that they are ordered in decreasing order with respect to the barycentric coordinates $b_{w_i}(x)$, so

$$b_{w_0}(x) \geq b_{w_1}(x) \geq \dots \geq b_{w_m}(x) > 0.$$

Let $\sigma_i = [w_0, \dots, w_i]$ for all $i \leq m$. By construction, x is contained in the geometric simplex $[[b(\sigma_0), \dots, b(\sigma_m)]]$ of $|\text{sd}\mathcal{K}|$.

If, on the other hand, x is a point of $|\text{sd}\mathcal{K}|$ with barycentric coordinates

$$x = \sum_{j=0}^m b_{b(\sigma_j)}(x)b(\sigma_j)$$

for some simplices

$$\sigma_0 \subset \sigma_1 \subset \dots \subset \sigma_m = [w_0, \dots, w_m]$$

of \mathcal{K} , then clearly, x is contained in the simplex σ_m . So, we can write x in barycentric coordinates with respect to $[w_0, \dots, w_m]$. A sketch to encourage intuition can be found in Figure 2.6.

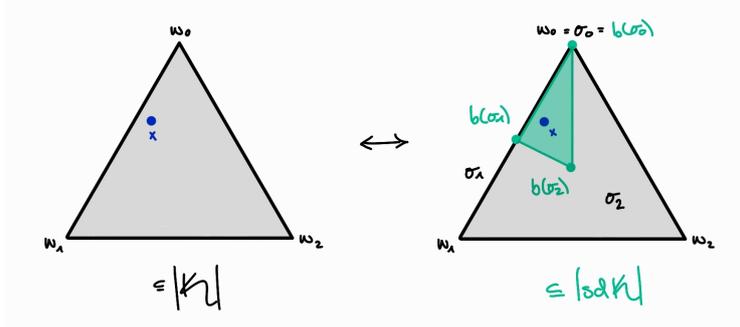


Figure 2.6: Sketch of how the change of barycentric coordinates from the proof of Lemma 2.12 might look.

The precise formulas for how to switch between the coordinates can be found in [Bau+23b], Lemmas A.2 and A.3.

We fix a vertex $v \in V(\mathcal{K})$ and let x be a point in the geometric realisation $|\mathcal{K}|$ such that

$$b_v(x) \geq b_w(x) \quad \text{for all } w \in V(\mathcal{K}).$$

We show that x is contained in a simplex of $|\text{sd}\mathcal{K}|$ which contains the vertex v . Let $|\sigma| = |[w_0, \dots, w_m]|$ be the smallest simplex of $|\mathcal{K}|$ which contains x and assume like before that the vertices w_0, \dots, w_m are ordered in decreasing order with respect to the barycentric coordinates $b_{w_i}(x)$. This implies $v = w_0$, since the barycentric coordinate $b_v(x)$ is maximal by assumption. By the precedent argument, this means that x is contained in a simplex $[v = b(\sigma_0), b(\sigma_1), \dots, b(\sigma_m)]$. Thus, $x \in \text{bst}(v, \mathcal{K})$.

On the other hand, let x be contained in $\text{bst}(v, \mathcal{K})$. Then there exists some simplex τ of $\text{sd}\mathcal{K}$ with v as a vertex, for which $x \in |\tau|$. We write $\tau = [b(\sigma_0), \dots, b(\sigma_m)]$ for $v = \sigma_0, \subset \dots, \subset \sigma_m$ a chain of simplices of \mathcal{K} . By the coordinate switch argument, this means, that the barycentric coordinate of x in \mathcal{K} with respect to v is maximal. \square

We end the section with a useful lemma about simplicial complexes and subdivisions in general.

Lemma 2.13. *Let \mathcal{K} and \mathcal{L} be simplicial complexes and let*

$$f: |\mathcal{K}| \longrightarrow |\mathcal{L}|$$

be a map which is linear on all simplices of \mathcal{K} . There exists a simplicial subdivision \mathcal{K}' of \mathcal{K} and \mathcal{L}' of \mathcal{L} , such that

$$f: |\mathcal{K}'| \longrightarrow |\mathcal{L}'|$$

is simplicial, i.e. maps every simplex of \mathcal{K}' to a simplex of \mathcal{L}' .

For a proof, we refer to [Col82], Lemma 2.13.

2.3 Simplicial Homology

In this section, we will describe collections of adjacent simplices using vector spaces. These collections are usually called chains and can be written as formal sums of simplices. Using those, we will be able to identify loops and higher-dimensional voids in simplicial complexes. We present theory from [Car14],[EH10] and [Hat01].

To define said chains, we make use of free vector spaces.

Definition 2.14 (Free Vector Space). *Let K be a field and S a finite set. The free K -vector space on S is the vector space $V_K(S)$ which has basis S and contains all formal sums*

$$\sum_{s \in S} \lambda_s s,$$

where $\lambda_s \in K$ for all s . Its dimension is $\dim V_K(S) = \#S$.

The next lemma gives a characterisation for free quotient vector spaces.

Lemma 2.15. *Let X be a finite set and R be an equivalence relation on X . Define $V_K(R)$ to be the subspace of $V_K(X)$ spanned by $\{x - y : x \sim_R y\}$. Then*

$$V_K(X)/V_K(R) \cong V_K(X/R).$$

Proof. The map above is given by

$$V_K(X)/V_K(R) \longrightarrow V_K(X/R), \quad [x] \longmapsto \bar{x}.$$

We check that it is bijective on the basis elements. First, let $\bar{x} = \bar{y}$ in $V_K(X/R)$. This directly implies $x \sim_R y$ by definition of the free vector spaces. Thus, by definition of $V_K(R)$, $[x] = [y]$ in $V_K(X)/V_K(R)$. We have shown injectivity.

For surjectivity, let \bar{x} be an element of $V_K(X/R)$. Since any element $x \in X$ is contained in exactly one equivalence class $[x]$ in X/R , it follows by construction of $V_K(R)$, that there is exactly one element $[x]$ in the preimage of \bar{x} . \square

From now on, we will always assume $K = \mathbb{F}_2$ to be the the field containing two elements 0, 1 and such that $1 + 1 = 0$. We define the free K -vector spaces that we use in the context of simplicial complexes.

Definition 2.16. *Let \mathcal{K} be a simplicial complex. We define the free K -vector spaces*

$$C_k(\mathcal{K}) = V_K(\Sigma_k(\mathcal{K})) \quad \text{for all } k \in \mathbb{N}_0.$$

Moreover, we consider the maps

$$\partial_k : C_k(\mathcal{K}) \longrightarrow C_{k-1}(\mathcal{K}), \quad \sigma = [v_0, \dots, v_k] \longmapsto \sum_{i=0}^k [v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_k].$$

We call $C_k(\mathcal{K})$ the **vector space of k -chains of \mathcal{K}** , its elements **k -chains** and ∂_k the **k -th boundary operator**, as it maps any k -simplex to the $(k-1)$ -chain which is the sum of the simplices in its boundary.

Remark 2.17. We are already using the fact that $K = \mathbb{F}_2$, so $1 = -1$. Otherwise, we would have to add signs in the definition of the boundary operator to take the orientation of simplices into consideration. This will not be discussed in this thesis, but we refer to [Hat01], Section 2.1 for further reading.

Example 2.18. Consider the simplicial complex \mathcal{K} from Figure 2.2 on the left. We get the following free vector spaces.

$$\begin{aligned} C_0(\mathcal{K}) &= \text{span}_K(v_1, v_2, v_3, v_4, v_5) \\ C_1(\mathcal{K}) &= \text{span}_K([v_1, v_2], [v_1, v_3], [v_2, v_3], [v_3, v_4]) \\ C_2(\mathcal{K}) &= \text{span}_K([v_1, v_2, v_3]) \\ C_k(\mathcal{K}) &= \text{span}_K(\emptyset) = \{0\} \text{ for } k \geq 3 \end{aligned}$$

The boundary operators ∂_k can be written as matrices with columns corresponding to the k -simplices of the simplicial complex and rows corresponding to the $(k-1)$ -simplices. The (i, j) -th entry of the matrix is one, if the lower-dimensional $(k-1)$ -simplex corresponding to the i -th row is a face of the higher-dimensional k -simplex corresponding to the j -th column, and zero if this is not the case. So the 1-entries of the matrix indicate faces which occur in the formal sum that is the boundary of the simplex. For example, the boundary operator ∂_1 corresponding to the simplicial complex in Figure 2.2 can be represented by the matrix

$$\partial_1 = \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{array} \begin{array}{cccc} [v_1, v_2] & [v_1, v_3] & [v_2, v_3] & [v_3, v_4] \\ \left[\begin{array}{cccc} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array} \right] \end{array}.$$

We check that this matrix satisfies Definition 2.16 for the edge $[v_1, v_2]$.

$$\partial_1([v_1, v_2]) = \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{array} \begin{array}{cccc} [v_1, v_2] & [v_1, v_3] & [v_2, v_3] & [v_3, v_4] \\ \left[\begin{array}{cccc} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array} \right] \cdot \begin{array}{c} \left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \end{array} \right] \begin{array}{c} [v_1, v_2] \\ [v_1, v_3] \\ [v_2, v_3] \\ [v_3, v_4] \end{array} \end{array} = \begin{array}{c} \left[\begin{array}{c} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{array} \right] \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{array} = v_1 + v_2,$$

which are exactly the vertices incident to $[v_1, v_2]$ as can be seen in Figure 2.2.

Using the spaces $C_k(\mathcal{K})$ and the maps ∂_k , we can build a chain of vector spaces connected by maps, which looks like

$$\dots \xrightarrow{\partial_{k+1}} C_k(\mathcal{K}) \xrightarrow{\partial_k} C_{k-1}(\mathcal{K}) \xrightarrow{\partial_{k-1}} \dots \xrightarrow{\partial_1} C_0(\mathcal{K}) \xrightarrow{\partial_0} 0.$$

Such chains are called chain complexes.

Definition 2.19 (Chain Complex). A *chain complex* is a sequence

$$C_\bullet = \{C_k\}_{k \in \mathbb{N}_0}$$

of vector spaces with linear maps $\partial_k: C_k \rightarrow C_{k-1}$ for all k , where $\partial_0: C_0 \rightarrow 0$ is the last map, and such that $\partial_k \circ \partial_{k+1} = 0$ for all $k \in \mathbb{N}_0$.

Remark 2.20. The condition $\partial_k \circ \partial_{k+1} = 0$ is equivalent to $\text{im}(\partial_{k+1}) \subseteq \ker(\partial_k)$.

Lemma 2.21. Let \mathcal{K} be a simplicial complex and let

$$\dots \xrightarrow{\partial_{k+1}} C_k(\mathcal{K}) \xrightarrow{\partial_k} C_{k-1}(\mathcal{K}) \xrightarrow{\partial_{k-1}} \dots \xrightarrow{\partial_1} C_0(\mathcal{K}) \xrightarrow{\partial_0} 0$$

be the resulting chain of vector spaces. Then for all $k \in \mathbb{N}_0$,

$$\partial_k \circ \partial_{k-1} = 0,$$

so $C_\bullet(\mathcal{K})$ is indeed a chain complex.

Proof. Let $\sigma \in C_k(\mathcal{K})$ be a k -simplex of \mathcal{K} . We write $\sigma = [v_0, \dots, v_k]$, where v_0, \dots, v_k are the vertices spanning σ . We compute

$$\begin{aligned} (\partial_{k-1} \circ \partial_k)(\sigma) &= \partial_{k-1} \left(\sum_{i=0}^k [v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_k] \right) \\ &= \sum_{i=0}^k \partial_{k-1} ([v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_k]) \\ &= \sum_{i=0}^k \left(\sum_{j < i} [v_0, \dots, v_{j-1}, v_{j+1}, \dots, v_{i-1}, v_{i+1}, \dots, v_k] \right. \\ &\quad \left. + \sum_{j > i} [v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_{j-1}, v_{j+1}, \dots, v_k] \right) \\ &= 0, \end{aligned}$$

by switching the indices j and i in the second sub-sum. Showing that the linear transformation $\partial_{k-1} \circ \partial_k$ is zero on the basis elements of $C_k(\mathcal{K})$ (which are the k -simplices of \mathcal{K}) is enough to show that it is zero on all elements by linearity, so we are done. \square

Speaking in the language of chains, a cycle is a chain whose boundary is empty. If we think about a cycle made of edges, e.g. the cycle $[v_1, v_2] + [v_2, v_3] + [v_3, v_4] + [v_4, v_1]$ in Figure 2.7, we see that when applying the boundary operator ∂_1 , every vertex appears exactly twice in the resulting sum, which means the sum cancels out as $K = \mathbb{F}_2$.

In order to find voids in a simplicial complex \mathcal{K} , we consider k -chains c for which $\partial_k(c) = 0$, which is exactly the subspace $\ker(\partial_k)$ of $C_k(\mathcal{K})$ and check that they are not the boundary of any $(k+1)$ -simplex, i.e. they are not in the image $\text{im}(\partial_{k+1})$, since in that case the cycle is filled in with a higher-dimensional simplex. Formally, this idea gives rise to homology groups, which are vector spaces in our setting.

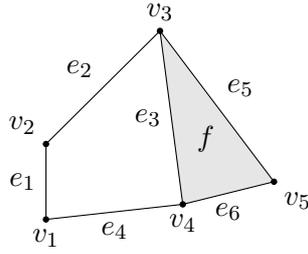


Figure 2.7: A simplicial complex consisting of a 1-dimensional cycle plus a triangle.

Definition 2.22 (Homology). Let

$$\dots \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \xrightarrow{\partial_{k-1}} \dots \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

be a chain complex. The ***k*-th homology group** of the chain complex is the quotient vector space $H_k = \ker \partial_k / \text{im } \partial_{k+1}$. The dimension $\beta_k = \dim(H_k)$ is called the ***k*-th Betti number**.

Remark 2.23. This notion makes sense, as $\ker \partial_k$ and $\text{im } \partial_{k+1}$ are both subspaces of C_k , and by the definition of a chain complex, $\text{im } \partial_{k+1} \subseteq \ker \partial_k$.

Proposition 2.24. Let U, V, W be K -vector spaces and let $\varphi: U \rightarrow V$ and $\psi: V \rightarrow W$ be linear maps, such that $\psi \circ \varphi = 0$. Moreover, let L, M and N be linear self-maps of U, V and W , respectively, which are invertible. Then

(a) $N\psi M^{-1} \circ M\varphi L = 0$.

(b) $\ker(\psi) / \text{im}(\varphi) \cong \ker(N\psi M^{-1}) / \text{im}(M\varphi L)$ as K -vector spaces.

Proof. For (a): Since composition of maps is associative, we may rewrite

$$N\psi M^{-1} \circ M\varphi L = N\psi(MM^{-1})\varphi L = N(\psi \circ \varphi)L = N0L = 0.$$

For (b): Since L, M and N are all invertible linear maps, we have that

$$\ker(\psi) \cong \ker(N\psi M^{-1}) \quad \text{and} \quad \text{im}(\varphi) \cong \text{im}(M\varphi L).$$

This directly implies $\ker(\psi) / \text{im}(\varphi) \cong \ker(N\psi M^{-1}) / \text{im}(M\varphi L)$. □

In the setting of a chain complex induced by a simplicial complex \mathcal{K}

$$\dots \xrightarrow{\partial_{k+1}} C_k(\mathcal{K}) \xrightarrow{\partial_k} C_{k-1}(\mathcal{K}) \xrightarrow{\partial_{k-1}} \dots \xrightarrow{\partial_1} C_0(\mathcal{K}) \xrightarrow{\partial_0} 0,$$

Proposition 2.24 enables us to construct an algorithm that makes computing homology groups much easier. The core idea is to use elementary matrices to change the matrices ∂_p so that we can read off the homology groups very easily.

Definition 2.25 (Elementary Matrices). For positive integers $i, j \in \mathbb{N}$, we define the elementary matrix $e(i, j)$ by

$$e_{k,l}(i, j) = \begin{cases} 1 & (k, l) = (i, j) \\ 1 & k = l \\ 0 & \text{else.} \end{cases}$$

We will apply the following three operations to a pair of boundary maps $(\partial_k, \partial_{k+1})$.

- (a) Left multiplication of ∂_k with an elementary matrix $e(i, j)$, i.e. a row operation adding the i -th row to the j -th (this corresponds to the map L in Proposition 2.24).
- (b) Right multiplication of ∂_{k+1} with an elementary matrix $e(i, j)$, i.e. a column operation which adds the j -th columns to the i -th (this corresponds to the map N in Proposition 2.24).
- (c) Any column operation on ∂_k , while applying the inverse row operation to ∂_{k+1} (this corresponds to the maps M and M^{-1} in Proposition 2.24).

We observe that the inverse matrix for $e(i, j)$ is the exact same elementary matrix $e(i, j)$, since $K = \mathbb{F}_2$. Using these matrices, we formulate the following algorithm.

Algorithm 2.26. Input: A pair $(\partial_k, \partial_{k+1})$ of boundary matrices.

Output: A pair

$$\left(\begin{array}{c} \begin{bmatrix} I_k & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I_m \end{bmatrix} \end{array} \right),$$

Where I_k and I_m are the $k \times k$ and $m \times m$ identity matrices, respectively.

Step 1: Apply row and column operations to ∂_k until it is of the form

$$\partial'_k = \begin{bmatrix} I_k & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Whenever we use a column operation, we apply the inverse row operation to the matrix ∂_{k+1} . The new matrix ∂'_{k+1} then looks like

$$\begin{bmatrix} B_{1,1} & B_{1,2} & B_{1,3} \\ B_{2,1} & B_{2,2} & B_{2,3} \\ B_{3,1} & B_{3,2} & B_{3,3} \end{bmatrix},$$

where $B_{1,1}$ has k rows, and we set the number of rows of $B_{2,1}$ and $B_{3,1}$ to be l and m , respectively. Since the relation $\partial_k \circ \partial_{k+1} = 0$ is preserved under the operations performed above by Proposition 2.24, we deduce that the blocks $B_{1,1}, B_{1,2}$ and $B_{1,3}$ are zero already (by the rules of matrix multiplication).

Step 2: Apply row and column operations to the $l + m$ non-zero rows of ∂'_{k+1} and transpose rows and columns until it is of the form

$$\partial''_{k+1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I_m \end{bmatrix}.$$

For the row operations, we have to apply the inverse operations to ∂'_k . But this has no effect, since the $l + m$ last rows of ∂'_k are zero.

When performing the operations described above, it is useful to keep track of the basis elements corresponding to the rows and columns. This allows us to find the homology group easily. Using Proposition 2.24,

$$H_k(\mathcal{K}) = \ker \partial_k / \text{im } \partial_{k+1} \cong \ker \partial'_k / \text{im } \partial''_{k+1}.$$

By linear algebra, the kernel of ∂'_k is spanned by the basis elements corresponding to its zero columns. The image of ∂''_{k+1} is spanned by the basis elements corresponding to its non-zero rows.

So, we restrict ∂''_{k+1} to $\ker(\partial'_k)$, i.e. we look at the $l + m$ last rows of ∂'_{k+1} and we check if there is a zero row. A zero row means that we found a basis element which is in $\ker(\partial'_k) \setminus \text{im}(\partial''_{k+1})$. H_k is spanned by exactly those elements.

Example 2.27. Consider the simplicial complex \mathcal{K} in Figure 2.7. The chain vector spaces are

$$\begin{aligned} C_0 &= \text{span}_K(v_1, \dots, v_5) \\ C_1 &= \text{span}_K(e_1, \dots, e_6) \\ C_2 &= \text{span}_K(f) \\ C_k &= \text{span}_K(\emptyset) = 0, \end{aligned}$$

and the boundary maps ∂_1 and ∂_2 are given by

$$(\partial_1, \partial_2) = \left(\begin{array}{c} e_1 \quad e_2 \quad e_3 \quad e_4 \quad e_5 \quad e_6 \\ \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \begin{array}{c} f \\ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \end{array} \right).$$

We first add the first row to the second in ∂_1 , then the first column to the fourth. We apply the reverse row operation to ∂_2 , which means we add the fourth row to the first.

The resulting pair is

$$(\partial_1, \partial_2) = \left(\begin{array}{cccccc} e_1 & e_2 & e_3 & e_1 + e_4 & e_5 & e_6 \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} & , & \begin{array}{c} f \\ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \end{array} \right).$$

We continue to proceed like in the algorithm and obtain the final pair

$$(\partial_1, \partial_2) = \left(\begin{array}{cccccc} e_1 & e_2 & e_3 & e_3 + e_5 & e_1 + e_2 + e_3 + e_4 & e_3 + e_5 + e_6 \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} & , & \begin{array}{c} f \\ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \end{array} \right).$$

We can now directly read off that $\ker(\partial_1) = \text{span}_K(e_1 + e_2 + e_3 + e_4, e_3 + e_5 + e_6)$ and if we restrict ∂_2 to its last two rows and compare that with the last two columns of ∂_1 , we see that $e_3 + e_5 + e_6$ is in the image of ∂_2 (since there is a 1 in the last row), but $e_1 + e_2 + e_3 + e_4$ is not. Therefore, $e_1 + e_2 + e_3 + e_4$ is a generator for the first homology group $H_1(\mathcal{K})$. If we look back to Figure 2.7, we see that this chain corresponds exactly to the (hollow) cycle on the left side of the complex.

Next, we want to be able to translate maps between simplicial complexes to maps between their homology groups. To do that, we define maps between chain complexes as intermediate step.

Definition 2.28 (Chain Map). Let C_\bullet and C'_\bullet be chain complexes. Then a collection of linear maps $\{f_k: C_k \rightarrow C'_k\}_{k \in \mathbb{N}_0}$ such that all squares 2.1 commute is called a **chain map**.

$$\begin{array}{ccc} C_k & \xrightarrow{\partial_k} & C_{k-1} \\ f_k \downarrow & & \downarrow f_{k-1} \\ C'_k & \xrightarrow{\partial'_k} & C'_{k-1} \end{array} \quad (2.1)$$

The following theorem tells us that maps between simplicial complexes induce chain maps, which in turn induce linear maps between homology groups (i.e. vector spaces).

Theorem 2.29 (Functoriality). Let \mathcal{K} and \mathcal{K}' be simplicial complexes and let $f: \mathcal{K} \rightarrow \mathcal{K}'$ be a simplicial map. Then

(a) the chain map $f_\bullet: C_\bullet(\mathcal{K}) \longrightarrow C_\bullet(\mathcal{K}')$ given by

$$f_k: C_k(\mathcal{K}) \longrightarrow C_k(\mathcal{K}')$$

$$\sigma \longmapsto \begin{cases} f(\sigma) & \dim(f(\sigma)) = k \\ 0 & \dim(f(\sigma)) < k \end{cases} \quad \text{for all } k,$$

satisfies $f_{k-1} \circ \partial_k = \partial'_k \circ f_k$ for all k . In particular, any map of simplicial complexes induces a chain map.

(b) there are induced homomorphisms $H_k(f): H_k(\mathcal{K}) \longrightarrow H_k(\mathcal{K}')$ for all k .

The map (b) is a result of fact that a chain map between \mathcal{K} and \mathcal{L} sends cycles in \mathcal{K} to cycles in \mathcal{K}' , and boundaries to boundaries (by the commutative diagram 2.1). This is what allows us to construct the map

$$H_k(f): \ker(\partial_k) / \text{im}(\partial_{k+1}) \longrightarrow \ker(\partial'_k) / \text{im}(\partial'_{k+1}).$$

For a full proof of Theorem 2.29, consult [EH10], Section IV.1.

2.4 Cubical Complexes and Homology

Next, we introduce cubical complexes. They are useful when we want to put a set into a collection of boxes, and thus make it more combinatorial for computation. In this thesis, they will play an important role in the computation of comparison spaces for the cycling signatures. We already observe that cubes are polytopes, so it makes sense to talk about their faces. We follow [Bau+23b], [KMM04], and [Col82].

Definition 2.30 (Cube). Let $p \in \mathbb{R}^d$ and let $r > 0$. We denote the d -dimensional **cube** with centre $p \in \mathbb{R}^d$ and side length r by

$$Q_r(p) = \overline{B_\infty(p, r/2)} = \prod_{i=1}^d [p_i - r/2, p_i + r/2].$$

A cube $Q_1(p)$ is called an **elementary cube**. Moreover, we denote the set of d -dimensional cubes in \mathbb{R}^d with side length r and vertices on the integer grid \mathbb{Z}^d by

$$\mathcal{Q}_r^d(\mathbb{R}^d) := \{Q_r(p) : p \in (r - 1/2)\mathbb{Z}^d\} = \left\{ \prod_{i=1}^d [v_i, v_i + r], v_i \in r\mathbb{Z}^d \right\},$$

and more generally for $k \leq d$,

$$\mathcal{Q}_r^k(\mathbb{R}^d) = \left\{ \prod_{i=1}^d [v_i, v_i + rw_i], v_i \in r\mathbb{Z}^d, w_i \in \{0, 1\}, \sum_{i=1}^d w_i = k \right\}.$$

For any cube $Q \in \mathcal{Q}_r^k(\mathbb{R}^d)$, we write $\downarrow Q$ to denote the collection of all faces of Q of dimension $\leq k$. For a collection of cubes $K \subseteq \downarrow \mathcal{Q}_r^d(\mathbb{R}^d)$, we write

$$\downarrow K = \bigcup_{Q \in K} \downarrow Q.$$

Moreover, we denote the **geometric realisation** of such a collection K by

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

Using cubes, we can define cubical complexes. The construction of cubical complexes resembles that of simplicial complexes, but instead of triangles and tetrahedra, we glue squares and cubes together along their common boundaries.

Definition 2.31 (Cubical Complex). Consider the set

$$\mathcal{Q}(\mathbb{R}^d) := \downarrow \mathcal{Q}_1^d(\mathbb{R}^d)$$

of all elementary cubes of dimension $\leq d$ with vertices on the integer grid \mathbb{Z}^d .

(a) A **cubical complex** is a set $K \subseteq \mathcal{Q}(\mathbb{R}^d)$ for which $\downarrow K = K$. We write K^k to denote the subcomplex of K containing all k -dimensional cubes of K .

(b) Let $X \subseteq \mathbb{R}^d$ be a subset. The cubical complex

$$K(X) = \downarrow \{Q \in \mathcal{Q}^d(\mathbb{R}^d) : Q \cap X \neq \emptyset\}$$

is called the **outer cubical complex of X** . In analogy to (a), we write $K^k(X)$ to denote the k -skeleton of $K(X)$.

A sketch of an outer cubical complex can be found in Figure 2.8.

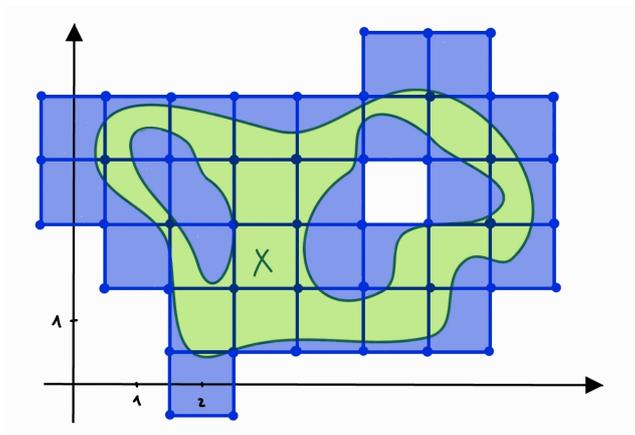


Figure 2.8: Geometric realisation $|K(X)|$ of an outer cubical complex (blue) of a set $X \subset \mathbb{R}^2$ (green).

The next definition relaxes Definition 2.31 in the sense that it allows us to scale and rotate the grid which defines the position and orientation of the cubes. If we want to construct the outer cubical complex for a very small set X , for example, this allows us to use a much finer grid to approximate the set.

Definition 2.32. Let $A \in GL_d(\mathbb{R})$ be a linear, invertible map. We define the collection

$$AQ(\mathbb{R}^d) = \{AQ : Q \in \mathcal{Q}(\mathbb{R}^d)\} \quad \text{where } AQ = \{Aq : q \in Q\}$$

Let $X \subseteq \mathbb{R}^d$ be a subset. The cubical complex

$$AK(X) = \downarrow \{Q \in AQ_1^d(\mathbb{R}^d) : Q \cap X \neq \emptyset\}$$

is called the **outer cubical cover of X transformed by A** .

In Section 2.2, we subdivided simplicial complexes into smaller simplicial complexes. Similarly, we will now subdivide cubical complexes into simplicial complexes.

Definition 2.33 (Simplicial Subdivision). Let K be a cubical complex. A **simplicial subdivision** of K is a simplicial complex \mathcal{K} such that $|K| = |\mathcal{K}|$ and every simplex of \mathcal{K} is contained in a cube of K .

We illustrate a simplicial subdivision of a cubical complex with a sketch in Figure 2.9.

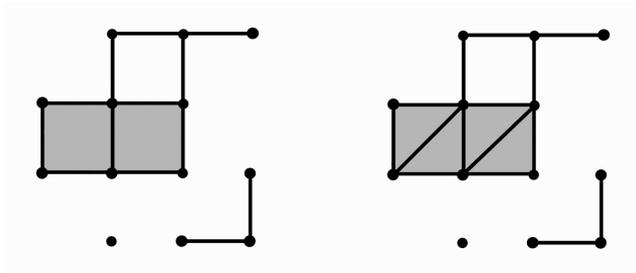


Figure 2.9: Simplicial subdivision (right) of a cubical complex (left).

Finally, we state a proposition that relates cubical complexes to simplicial complexes with the same vertex set, using subdivisions as introduced in Definition 2.33.

Proposition 2.34. Let K be a cubical complex. There exists a simplicial subdivision \mathcal{K} of K , such that $V(\mathcal{K}) = V(K)$.

Proof. We proceed by induction on the maximal dimension of cubes in K .

If K only consists of 0- or 1-dimensional cubes, then it is just a collection of points or edges, so K is already simplicial.

Assume now that K has a cube of maximal dimension $k > 1$. By induction, there exists a simplicial subdivision of $K_{(k-1)}$ with no added vertices. Let Q be a cube of dimension k and let v be a vertex of it. Consider the union $Q_{\hat{v}}$ of all faces of the cube Q that do not contain v . By induction, $Q_{\hat{v}}$ has a subdivision into a simplicial complex $\mathcal{K}_{\hat{v}}$ with equal vertex set. The cube Q is homeomorphic to the cone

$$vQ_{\hat{v}} = \{\lambda v + \mu Q' : Q' \in \mathcal{K}_{\hat{v}}, \lambda + \mu = 1\}.$$

A visualisation of this can be seen in Figure 2.10. We construct the simplicial subdivision of \mathcal{K} of Q is by adding to the collection of simplices of $\mathcal{K}_{\hat{v}}$ the simplices $\{v \cup |\sigma| : \sigma \in \mathcal{K}_{\hat{v}}\}$. \square



Figure 2.10: Visualisation of the induction in the proof of Proposition 2.34.

Using cubical complexes, we could now replicate everything we did in Section 2.3 and introduce cubical homology. We will not look at cubical homology in this thesis, but we refer to [KMM04] for an in-depth study of the topic.

Metric Spaces and Persistent Homology

3.1 Metric Spaces to Simplicial Complexes

In this section, we define constructions that allow us to transform finite metric spaces (a collection of points for which a distance is defined) into simplicial complexes, and then compute homology. This means that we will be able to infer the underlying topological features from only finitely many points. We follow [Car14], section 3.6., [Ada20], and [Bau+23a]. The first construction we present is called a Vietoris-Rips complex.

Definition 3.1 (Vietoris-Rips Complex). *Let (X, d) be a finite metric space and let $r \in \mathbb{R}_{\geq 0}$. The **Vietoris-Rips complex** $\text{VR}(X, r)$ has vertex set*

$$V(\text{VR}(X, r)) = X = \{x_1, \dots, x_n\}$$

and contains a simplex $[x_{i_0}, \dots, x_{i_k}]$ if and only if $d(x_{i_j}, x_{i_l}) \leq r$ for all $j, l \in \{0, \dots, k\}$.

To understand what this brings us, we look at the following example.

Example 3.2. *Consider the metric space (X, d) , where $X = \{v_1, \dots, v_5\} \subset \mathbb{R}^2$ is the set containing the five points depicted on the left in Figure 3.1 and d is the Euclidean metric on \mathbb{R}^2 . Using the radius r as drawn on the right side of Figure 3.1, we see how the Vietoris-Rips complex connects these points, which previously had no interesting topology, to a simplicial complex which, in this case, contains one connected component.*



Figure 3.1: On the left side, we represent the set $X = \{v_1, \dots, v_5\}$. On the right, we drew the Vietoris-Rips complex $\text{VR}(X, r)$, where r is represented by the gray circles.

We introduce another way to construct simplicial complexes, but this time, we only require a topological space and no metric.

Definition 3.3 (Nerve Complex). *Let X be a topological space and let $\mathcal{U} = \{U_i\}_{i \in I}$ be a cover of X . The **nerve complex of \mathcal{U}** is the simplicial complex $N(\mathcal{U})$ on with vertex set I that contains a simplex $\sigma(J)$, $J \subseteq I$ if and only if*

$$\bigcap_{i \in J} U_i \neq \emptyset.$$

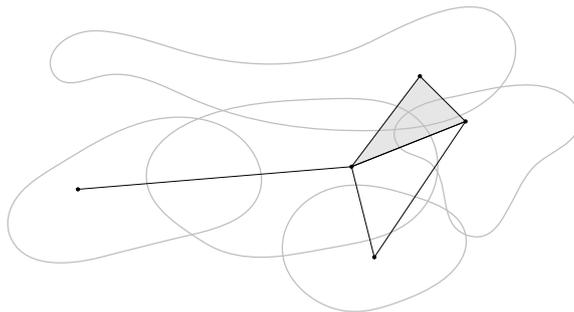


Figure 3.2: Example of a Nerve complex.

Intuitively, we choose one point per subset U_i and draw simplices between points whenever the intersection of the corresponding sets is non-empty. A visual example can be seen in Figure 3.2. Lastly, we introduce the Čech complex, which is a special case of a nerve complex.

Definition 3.4 (Čech Complex). Let (X, d) be a finite metric space and let $r \geq 0$. We write $X = \{x_1, \dots, x_n\}$. The Čech complex $\check{C}(X, r)$ has vertex set X and contains the simplex $[x_{i_0}, \dots, x_{i_k}]$ if and only if

$$\bigcap_{j=0}^k B(x_{i_j}, r/2) \neq \emptyset.$$

To make the connection to the nerve complex precise: for $\mathcal{U}_r = \{\overline{B(x, r/2)}\}_{x \in X}$, we have $N(\mathcal{U}_r) = \check{C}(X, r)$. Moreover, the Čech complex $\check{C}(X, r)$ is a subcomplex of the Vietoris-Rips complex $\text{VR}(X, r)$. In particular, the 1-simplices of the Vietoris-Rips complex and the Čech complex on the same metric space (X, d) are identical. This makes sense, as pairwise intersections and intersections are the same thing if we only have two sets.

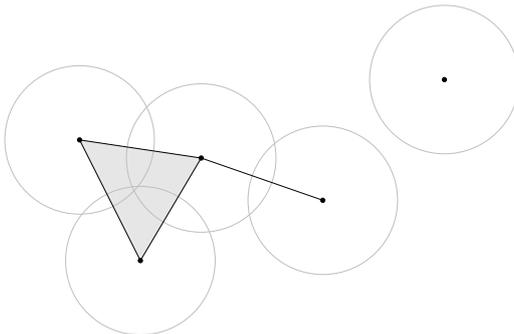


Figure 3.3: Example of a Čech complex.

3.2 Čech Complexes and Thickenings

As we already mentioned in the introduction, an easy way to find topological features in a finite set of points $X \subset \mathbb{R}^d$ is to thicken the points. We will see that the

thickening of the points X by r is homotopy equivalent to the Čech complex $\check{C}(X, r)$. In particular, this means that their homology groups are equal. The references for this section are [Bau+23a], [Bau+23b] and [Hat01].

We define the thickening of a subset of \mathbb{R}^d . This construction relates to the Vietoris-Rips and Čech complexes, in the sense that it also creates a connection between points that are within a certain distance r from each other.

Definition 3.5 (Thickening). *Let $X \subseteq \mathbb{R}^d$ be a subset. Let $r \geq 0$. The r -thickening of X is the open set*

$$U(X, r) = \bigcup_{p \in X} B(p, r/2).$$

Since we are interested in finding topological features of thickenings computationally, it makes sense that we would try to relate $U(X, r)$ to a simplicial complex. This is because simplicial complexes are combinatorial objects, for which computations are usually much more efficient. We will use subdivisions of Čech complexes to prove a version of the nerve theorem (Theorem 3.6, and see [Bau+23a] for different versions), which states that there is a homotopy equivalence between the closed thickening $\overline{U(X, r)}$ and the geometric simplicial complex $|\check{C}(X, r)|$.

One key ingredient for the proof of this result is a barycentric subdivision (as we introduced in Section 2.2). It is important to keep in mind that by definition, subdivisions preserve the geometric realisation of simplicial complexes. Let $r \geq 0$, and set

$$\mathcal{K}_r: \text{eqqs}d\check{C}(X, r).$$

We define the map

$$f_r: |\mathcal{K}_r| \longrightarrow \overline{U(X, r)}, \quad (3.1)$$

as follows. Let v be a vertex of \mathcal{K}_r . By definition, this means that v is the barycentre of a simplex $[p_0, \dots, p_k]$ of $\check{C}(X, r)$. In particular,

$$\bigcap_{i=0}^k \overline{B(p_i, r/2)} \neq \emptyset.$$

Let x be a point in this intersection. We set $f_r(v) = x$. We repeat this process for all vertices v in \mathcal{K}_r . This is a finite procedure, as the set X is finite by assumption.

Next, consider a simplex $\sigma = [v_0, \dots, v_k]$ of \mathcal{K}_r . We set

$$f_r(\lambda_0 v_0 + \dots + \lambda_n v_n): \text{eqq}\lambda_0 f_r(v_0) + \dots + \lambda_n f_r(v_n).$$

A visual representation of this map can be found in Figure 3.4. We state the nerve theorem.

Theorem 3.6 (Nerve Theorem). *Let $X \subset \mathbb{R}^d$ be a finite set of points. The map*

$$f_r: |\mathcal{K}_r| = |\check{C}(X, r)| \longrightarrow \overline{U(X, r)}$$

is a homotopy equivalence for all $r \geq 0$.

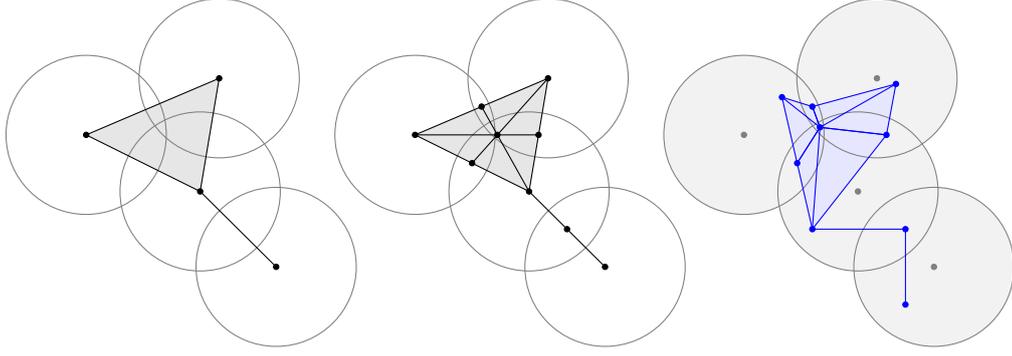


Figure 3.4: A Čech complex $\check{C}(X, r)$ with visible circles of radius $r/2$, its barycentric subdivision and the image of a possible choice for f_r drawn in blue inside the closed thickening $\overline{U(X, r)}$.

Proof. To prove that f_r is a homotopy equivalence for all $r \geq 0$, we are going to construct a homotopy inverse, i.e. a map

$$f_r^{-1}: \overline{U(X, r)} \longrightarrow |\mathcal{K}_r|$$

such that $f_r \circ f_r^{-1}$ and $f_r^{-1} \circ f_r$ are homotopic to the identity on $\overline{U(X, r)}$ and $|\mathcal{K}_r|$, respectively.

Let $p, q \in X$ be points for which

$$\overline{B(p, r/2)} \cap \overline{B(q, r/2)} = \emptyset.$$

Since \mathbb{R}^d is normal (or T_4), there exists a $\varepsilon > 0$ such that the open balls $V_p: \text{eqq}B(p, r/2 + \varepsilon)$ and $V_q: \text{eqq}B(q, r/2 + \varepsilon)$ are disjoint. This property inductively extends to the finite family of closed balls $\{\overline{B(p, r/2)}\}_{p \in X}$. We call the resulting open cover of $\overline{U(X, r)}$

$$\mathcal{C}: \text{eqq}\{V_p\}_{p \in X} = \{B(p, r/2 + \varepsilon)\}_{p \in X} \quad \text{for some } \varepsilon \text{ small enough.}$$

By construction, the new cover preserves empty intersections, which implies

$$|\mathcal{K}_r| = |\check{C}(X, r)| = |N(\mathcal{C})|.$$

Consider the continuous map

$$\varphi_p: \mathbb{R}^d \longrightarrow [0, 1], \quad x \longmapsto \frac{d(x, \mathbb{R}^d \setminus V_p)}{d(x, \overline{B(p, r/2)}) + d(x, \mathbb{R}^d \setminus V_p)}.$$

For $x \in \overline{B(p, r/2)}$, it has value one, and for $x \notin V_p$, it is zero. We normalise φ_p , restrict it to $\overline{U(X, r)}$ and denote the resulting map by

$$\psi_p: \text{eqq}\varphi_p / \sum_{q \in X} \varphi_q: \overline{U(X, r)} \longrightarrow [0, 1].$$

We define the function

$$f_r^{-1}: \overline{U(X, r)} \longrightarrow |\check{C}(X, r)|, \quad x \longmapsto \sum_{p \in X} \psi_p(x) \cdot p,$$

So, the function f_r^{-1} indirectly provides us with barycentric coordinates on $\check{C}(X, r)$. We claim that f_r^{-1} is a homotopy inverse of f_r .

First, let us show that the pair

$$(f_r^{-1}, \text{id}_X): (\overline{U(X, r)}, \mathcal{C}) \longrightarrow (|\mathcal{K}_r|, \{\text{bst}(v_p, \mathcal{K}_r)\}_{p \in X})$$

is a morphism of covered spaces. Here, we use id_X to denote the function which sends the open set V_p to $\text{bst}(p, \check{C}(X, r))$, as well as its inverse map. Let x be in the open set $V_p \in \mathcal{C}$. Then by definition, $\varphi_p(x) = 1$, thus, $\psi_p(x) = \max_{q \in X} \psi_q(x)$ is maximal. Using Lemma 2.12, this implies $f_r^{-1}(x) \in \text{bst}(p, \check{C}(X, r))$.

Next, we show that the pair

$$(f_r, \text{id}_X): (|\mathcal{K}_r| = |\check{C}(X, r)|, \{\text{bst}(p, \check{C}(X, r))\}_{p \in X}) \longrightarrow (\overline{U(X, r)}, \mathcal{C})$$

is also a morphism of covered spaces. By construction, f_r sends the vertices of any simplex $|\sigma| \subseteq \text{bst}(v_p, \check{C}(X, r))$ to V_p . For the construction of V_p , we extended $\overline{B(p, r/2)}$ equally into all directions, so the convexity is preserved. Since f_r is affine linear on $|\sigma|$, $f_r(\sigma)$ is fully contained in V_p . So,

$$f_r(\text{bst}(p, \check{C}(X, r))) \subseteq V_p \quad \text{for all } p \in X.$$

Since both (f_r, id_X) and (f_r^{-1}, id_X) are morphisms of covered spaces, so are $(f_r \circ f_r^{-1}, \text{id}_X)$ and $(f_r^{-1} \circ f_r, \text{id}_X)$. For $x \in V_p$, we obtain $(f_r \circ f_r^{-1})(x) \in V_p$. The set V_p is convex, so the straight line from x to $(f_r \circ f_r^{-1})(x)$ is entirely contained in V_p . This means that we can construct a straight line homotopy between $\text{id}_{\overline{U(X, r)}}$ and $f_r \circ f_r^{-1}$. For $f_r^{-1} \circ f_r$, we use the fact that any two maps into a good cover (all intersections empty or contractible) are homotopic (Proposition 3.8 in [Bau+23a]). We have shown in Lemma 2.9 that barycentric stars are contractible, so f_r^{-1} is indeed a homotopy inverse for f_r . \square

Corollary 3.7. $H_1(\check{C}(X, r)) = H_1(\overline{U(X, r)})$ for all $r \geq 0$.

Proof. This follows directly from the fact that subdivisions preserve geometric representations, Theorem 3.6 and applying homotopy invariance (see e.g. [Hat01], Section 2.1, Homotopy Invariance). \square

3.3 Persistent Homology and its Representation

Assume we are looking at a finite metric space (X, d) , where X is a set of points in \mathbb{R}^d , and d is the Euclidean metric. The construction of a Vietoris-Rips complex $\text{VR}(X, r)$ can help us determine the underlying topological features. However,

choosing a good radius r is not trivial. If we choose it to be small, we might miss global, relevant topological structures. If we choose r to be rather large, we might overlook smaller, but equally interesting features. One approach to this issue is to find a way to capture topological features for all values of r at once.

Example 3.8. Recall the Vietoris-Rips complex from Example 3.2. In Figure 3.5, we constructed the same Vietoris-Rips complex but for 4 different radii to illustrate how the complex evolves when increasing the radius. It is clear from the picture that for $r < r'$, we have an inclusion $\text{VR}(X, r) \subseteq \text{VR}(X, r')$. We will see that these inclusions give the Vietoris-Rips complexes the structure of a \mathbb{R}_+ -filtered simplicial complex; the filtered set S being the set of simplices (including vertices), and the map ρ sending every simplex to the radius r of its first appearance.

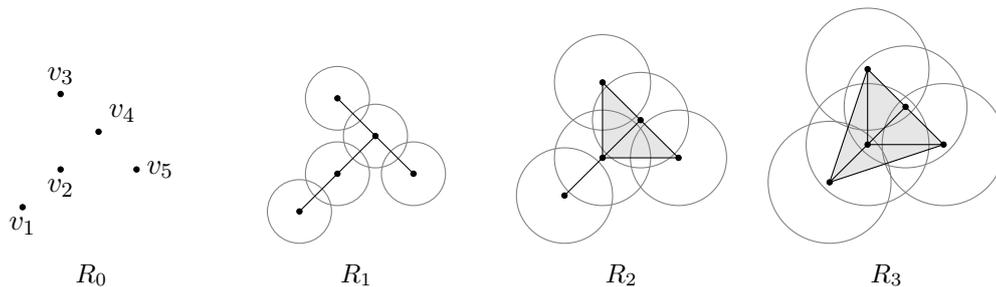


Figure 3.5: Example the Vietoris-Rips complexes $\text{VR}(X, r)$ arising from taking increasing radii $r \in \{R_0, \dots, R_3\}$, $0 = R_0 < R_1 < R_2 < R_3$. The set of vertices is $X = \{v_1, \dots, v_5\} \subset \mathbb{R}^2$.

As we hinted at in Example 3.8, we will add a continuous parameter (e.g. a radius) to everything we did in Section 2.3. Instead of simplicial complexes, we will now look at persistent simplicial complexes, which are simplicial complexes that change (usually grow) over time, like the Vietoris-Rips complex in Figure 3.5. This allows us to see how the homology groups change and what structures appear and disappear as the complex evolves. We follow chapter 3 of [Car14].

Definition 3.9 (Persistent Object). Let \mathcal{C} be a category.

(a) A **persistent object** is a family $\{X_r\}_{r \in \mathbb{R}}$ of objects X_r of \mathcal{C} with morphisms

$$\varphi_r^{r'} : X_r \longrightarrow X_{r'} \quad \text{for all } r \leq r'$$

satisfying

$$\varphi_{r'}^{r''} \circ \varphi_r^{r'} = \varphi_r^{r''} \quad \text{whenever } r \leq r' \leq r''$$

(b) A **morphism of persistent objects** $\{X_r\}_{r \in \mathbb{R}}$ and $\{Y_r\}_{r \in \mathbb{R}}$ is a family $\{f_r\}_{r \in \mathbb{R}}$ of morphisms of \mathcal{C} , such that

$$f_{r'} \circ \varphi_r^{r'} = \psi_r^{r'} \circ f_r,$$

where $\varphi_r^{r'}$ and $\psi_r^{r'}$ are the maps from (a) corresponding to $\{X_r\}_{r \in \mathbb{R}}$ and $\{Y_r\}_{r \in \mathbb{R}}$, respectively. So we have commutative diagrams

$$\begin{array}{ccc}
X_r & \xrightarrow{\varphi_r^{r'}} & X_{r'} \\
f_r \downarrow & & \downarrow f_{r'} \\
Y_r & \xrightarrow{\psi_r^{r'}} & Y_{r'}
\end{array} \cdot$$

Example 3.10 (Persistent Set). A persistent set is a family $\{X_r\}_{r \in \mathbb{R}}$ of sets with maps

$$\varphi_r^{r'} : X_r \longrightarrow X_{r'}$$

that satisfy

$$\varphi_{r'}^{r''} \circ \varphi_r^{r'} = \varphi_r^{r''}.$$

In particular, if we are given a set X and a map $\rho : X \longrightarrow \mathbb{R}$, we can construct the subsets $X_r = \{x \in X : \rho(x) \leq r\}$ and obtain a persistent set, where $\varphi_r^{r'}$ are inclusion maps. We will call such a pair (X, ρ) an \mathbb{R} -**filtered set**. Any structure with an underlying set can be filtered that way. For example, what we see in Figure 3.5 has the structure of an \mathbb{R} -filtered simplicial complex, where for a radius R such that $\text{VR}(X, r)$ is the full complex (which contains all possible simplices), ρ is defined by

$$\rho : \text{VR}(X, R) \longrightarrow \mathbb{R}, \quad \sigma \longmapsto \min\{r \geq 0 : \sigma \in \text{VR}(X, r)\}.$$

In order to introduce persistent homology, the starting point is to observe that for a persistent simplicial complex like $\{\text{VR}(X, r)\}_{r \geq 0}$, we obtain persistent vector spaces $\{C_k(\text{VR}(X, r))\}_{r \geq 0}$ of k -chains and homology groups $\{H_i(\text{VR}(X, r))\}_{r \geq 0}$ by functoriality (Theorem 2.29). Since we will encounter persistent vector spaces very often, we define them separately.

Definition 3.11 (Persistent Vector Space). A persistent vector space is a collection $V = \{V_r\}_{r \in \mathbb{R}}$ of vector spaces with homomorphisms

$$L_V(r, r') : V_r \longrightarrow V_{r'} \quad \text{for all } r \leq r',$$

such that

$$L_V(r', r'') \circ L_V(r, r') = L_V(r, r'') \quad \text{for all } r \leq r' \leq r''.$$

Depending on the author, persistent vector spaces might also be called **persistence modules** or **persistence vector spaces**.

As for usual vector spaces, we can define sub-persistent vector spaces $U = \{U_r\}_{r \in \mathbb{R}}$ of $V = \{V_r\}_{r \in \mathbb{R}}$ by requiring U_r to be a subspace of V_r for all r and $L_V(r, r')(U_r) \subseteq U_{r'}$. Moreover, L_U is the restriction of L_V to U . A special case is the image of a linear transformation $f : V \longrightarrow W$ of persistent vector spaces, which is defined as the sub-persistent vector space that collects all images of $f_r, r \in \mathbb{R}$ (which are subspaces of W_r , respectively). Analogously to quotient vector spaces, we can define persistent quotient vector spaces $V/U = \{V_r/U_r\}_{r \in \mathbb{R}}$ with maps

$$\begin{aligned}
L_{V/U}(r, r') : V_r/U_r &\longrightarrow V_{r'}/U_{r'} \\
[v] &\longmapsto [L_V(r, r')(v)].
\end{aligned}$$

Example 3.12. Consider the filtered Vietoris-Rips complex $\mathcal{K} = \{\mathcal{K}_i\}_{i=0,1,2}$ in Figure 3.6. We denote the j -th boundary map of \mathcal{K}_i by $\partial_{j,\mathcal{K}_i}$, write $C_j(\mathcal{K}) = \{C_j(\mathcal{K}_i)\}_{i=0,1,2}$ to denote the persistent vector space of chains and denote the corresponding linear maps (Definition 3.11) by

$$L_{C_k(\mathcal{K})}(i, j): C_k(\mathcal{K}_i) \longrightarrow C_k(\mathcal{K}_j) \quad \text{for } 0 \leq i < j \leq 2 \text{ and } k \in \mathbb{N}.$$

We obtain the following commutative diagram

$$\begin{array}{ccccccc}
\dots & \longrightarrow & 0 & \xrightarrow{\partial_{1,\mathcal{K}_0}} & C_0(\mathcal{K}_0) & \longrightarrow & 0 \\
& & \downarrow L_{C_1(\mathcal{K})}(0,1) & & \downarrow L_{C_0(\mathcal{K})}(0,1) & & \\
\dots & \longrightarrow & C_1(\mathcal{K}_1) & \xrightarrow{\partial_{1,\mathcal{K}_1}} & C_0(\mathcal{K}_1) & \longrightarrow & 0 \\
& & \downarrow L_{C_1(\mathcal{K})}(1,2) & & \downarrow L_{C_0(\mathcal{K})}(1,2) & & \\
\dots & \longrightarrow & C_1(\mathcal{K}_2) & \xrightarrow{\partial_{1,\mathcal{K}_2}} & C_0(\mathcal{K}_2) & \longrightarrow & 0
\end{array} \tag{3.2}$$

We have all necessary ingredients to compute the persistent vector space $H_0(\mathcal{K})$: $\text{eqq}\{H_0(\mathcal{K}_i)\}_{i=0,1,2}$. We first compute the spaces $H_0(\mathcal{K}_i)$ for $i = 0, 1, 2$ and then explain what the corresponding linear maps (Definition 3.11) are.

($i=0$): We have $\partial_{1,\mathcal{K}_0} = 0$ and $C_0(\mathcal{K}_0) = \text{span}_K(v_1, v_2, v_2)$, thus $\text{im}(\partial_{1,\mathcal{K}_0}) = 0$ and $\ker(\partial_{0,\mathcal{K}_0}) = \text{span}_K(v_1, v_2, v_2)$, which implies

$$H_0(\mathcal{K}_0) = \ker(\partial_0) / \text{im}(\partial_1) \cong \text{span}_K(v_1, v_2, v_3).$$

This reflects the fact that \mathcal{K}_0 has three connected components.

($i=1$): This time, the vector space of 1-chains is more interesting; $C_1(\mathcal{K}_1) = \text{span}_K([v_1, v_2])$ but $C_0(\mathcal{K}_1) = \text{span}_K(v_1, v_2, v_2)$ stays the same. We have

$$\text{im}(\partial_{1,\mathcal{K}_1}) = \text{span}_K(\partial_{1,\mathcal{K}_1}([v_1, v_2])) = \text{span}_K(v_1 + v_2).$$

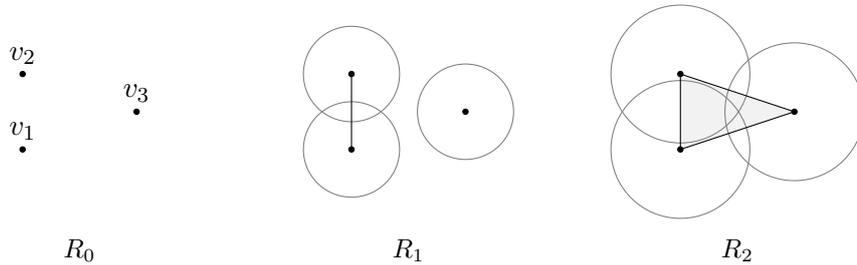


Figure 3.6: A filtered Vietoris-Rips complex $\mathcal{K} = \{\mathcal{K}_i\}_{i=0,1,2}$, $\mathcal{K}_i = \text{VR}(X, R_i)$ with vertex set $X = \{v_1, v_2, v_3\}$ and increasing radii $0 = R_0 < R_1 < R_2$ as indicated by the grey circles.

Therefore, we obtain the homology group

$$\begin{aligned}
H_0(\mathcal{K}_1) &= \ker(\partial_{0,\mathcal{K}_1}) / \text{im}(\partial_1) \\
&= \text{span}_K(v_1, v_2, v_3) / \text{span}_K(v_1 + v_2) \\
&\cong \text{span}_K(v_1 + v_2, v_2, v_3) / \text{span}_K(v_1 + v_2) \\
&\cong \text{span}_K(v_2, v_3).
\end{aligned}$$

This reflects the fact that \mathcal{K}_1 has two connected components.

(i=2): For the last homology group in the filtration, we have the vector space of 1-chains $C_1(\mathcal{K}_2) = \text{span}_K([v_1, v_2], [v_1, v_3], [v_2, v_3])$ and as before, $C_0(\mathcal{K}_2) = \text{span}_K(v_1, v_2, v_3)$. This time, we apply Algorithm 2.26 to the pair $(\partial_{0,\mathcal{K}_2}, \partial_{1,\mathcal{K}_2})$ where $\partial_0 = 0$ and

$$\partial_{1,\mathcal{K}_2} = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \end{array} \begin{array}{ccc} [v_1, v_2] & [v_1, v_3] & [v_2, v_3] \\ \left[\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{array} \right] \end{array}.$$

We obtain the matrix

$$P\partial_{1,\mathcal{K}_2}Q = \begin{array}{c} v_1 + v_2 \\ v_2 + v_3 \\ v_3 \end{array} \begin{array}{ccc} [v_1, v_2] & [v_1, v_2] + [v_1, v_3] & [v_1, v_2] + [v_1, v_3] + [v_2, v_3] \\ \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array} \right] \end{array},$$

where P and Q are products of elementary matrices. So we see that $\text{im}(P\partial_{1,\mathcal{K}_2}Q) = \text{span}_K(v_1 + v_2, v_2 + v_3)$, thus with Proposition 2.24,

$$\begin{aligned}
H_0(\mathcal{K}_2) &= \ker(\partial_{1,\mathcal{K}_2}) / \text{im}(\partial_{1,\mathcal{K}_2}) \\
&= \text{span}_K(v_1, v_2, v_3) / \text{span}_K(v_1 + v_2, v_1 + v_3, v_2 + v_3) \\
&\cong \text{span}_K(v_1, v_2, v_3) / \text{span}_K(v_1 + v_2, v_2 + v_3) \\
&\cong \text{span}_K(v_1 + v_2, v_2 + v_3, v_3) / \text{span}_K(v_1 + v_2, v_2 + v_3) \\
&\cong \text{span}_K(v_3),
\end{aligned}$$

which reflects the fact that \mathcal{K}_3 has only one connected component.

We now explain the maps in homology that are induced by the chain maps from the commutative diagram 3.2.

In $H_0(\mathcal{K}_0)$ each of the vertices v_1, v_2, v_3 is a generator. When passing to $H_0(\mathcal{K}_1)$, we see that

$$[v_1] = [v_1] + [v_1 + v_2] = [v_1 + v_1] + [v_2] = [v_2].$$

So the induced map $H_0(L_C(\mathcal{K})(0, 1))$ is given by

$$\begin{aligned}
H_0(L_C(\mathcal{K})(0, 1)): C_0(\mathcal{K}_0) / \text{im}(\partial_{1,\mathcal{K}_0}) &\cong C_0(\mathcal{K}_0) \longrightarrow C_0(\mathcal{K}_1) / \text{span}_K(v_1 + v_2) \\
v_1 &\longmapsto [v_1] = [v_2] \\
v_3 &\longmapsto [v_3].
\end{aligned}$$

For the map $H_0(L_{C(\mathcal{K})}(1, 2))$, we observe that

$$[v_1] = [v_1] + [v_1 + v_3] = [v_1 + v_1] + [v_3]$$

and

$$[v_2] = [v_2] + [v_2 + v_3] = [v_2 + v_2] + [v_3] = [v_3]$$

in $H_0(\mathcal{K}_2) = \text{span}_K(v_1, v_2, v_3) / \text{span}_K(v_1 + v_2, v_1 + v_3, v_2 + v_3)$. Therefore,

$$\begin{aligned} H_0(L_{C(\mathcal{K})}(1, 2)): C_0(\mathcal{K}_1) / \text{im}(\partial_{1, \mathcal{K}_1}) &\longrightarrow C_0(\mathcal{K}_2) / \text{span}_K(v_1 + v_2, v_1 + v_3, v_2 + v_3) \\ [v_1] = [v_2] &\longmapsto [v_1] = [v_2] = [v_3] \\ [v_3] &\longmapsto [v_1] = [v_2] = [v_3]. \end{aligned}$$

This completes the description of the persistent vector space $H_0(\mathcal{K}) = \{H_0(\mathcal{K}_i)\}_{i=0,1,2}$.

Definition 3.13 (Free, Finitely Generated Persistent Vector Space). Let (X, ρ) be a \mathbb{R}_+ -filtered set.

- (a) The **free persistent vector space** on (X, ρ) over a field K is the persistent vector space $\{V_K(X, \rho)_r\}_{r \in \mathbb{R}}$, where for all $r > 0$, $V_K(X, \rho)_r$ is the free K -vector space spanned by the set $\{x \in X : \rho(x) \leq r\}$. In particular, this means

$$V_K(X, \rho)_r \subseteq V_K(X, \rho)_{r'} \subseteq V_K(X) \text{ for all } r \leq r'.$$

- (b) A persistent vector space $\{V_r\}_{r \in \mathbb{R}}$ is **free**, if it is isomorphic to a space $\{V_K(X, \rho)_r\}_{r \in \mathbb{R}}$. If X can be chosen to be finite, $\{V_r\}_{r \in \mathbb{R}}$ is **finitely generated**.

Example 3.14. Consider the persistent simplicial complex \mathcal{K} from Example 3.12. Then $C_k(\mathcal{K})$ is a free and finitely generated persistent vector space for all $k \in \mathbb{N}$. In fact, the \mathbb{R}_+ -filtered set (X, ρ) from Definition 3.13 is given by the set

$$\Sigma_k(\mathcal{K}) = \bigcup_{i=0,1,2} \Sigma(\mathcal{K}_i),$$

and the map ρ , which assigns to every k -simplex σ the radius

$$\rho(\sigma) = \min\{i \in \{0, 1, 2\} : \sigma \in \Sigma_k(\mathcal{K}_i)\}$$

of its first appearance.

Lemma 3.15. An element $\sum_{x \in X} a_x x$ of $V_K(X)$ is contained in $V_K(X, \rho)_r$ if and only if $a_x = 0$ for all $x \in X$ with $\rho(x) > r$.

Proof. Recall that

$$V_K(X, \rho)_r = \text{span}_K(\{x \in X : \rho(x) \leq r\}),$$

so the lemma follows directly from the definition of the linear span. \square

Definition 3.16 (Finitely Presented Persistent Vector Space). A *finitely presented* persistent vector space is one isomorphic to a persistent vector space of the form

$$\{W_r\}_{r \in \mathbb{R}} / \text{im}(f)$$

for a linear transformation of finitely generated, free persistent vector spaces

$$f: \{V_r\}_{r \in \mathbb{R}} \longrightarrow \{W_r\}_{r \in \mathbb{R}}.$$

Example 3.17. The persistent vector spaces $\{H_k(\mathcal{K}_i)\}_{i=0,1,2}$ from Example 3.12 are finitely presented. Consider the linear transformation of finitely generated, free persistent vector spaces

$$\{\partial_{k,\mathcal{K}_i}\}_{i=0,1,2}: \{C_{k,\mathcal{K}_i}\}_{i=0,1,2} \longrightarrow \{\ker(\partial_{k-1,\mathcal{K}_i})\}_{i=0,1,2} \subseteq \{C_{k-1}(\mathcal{K}_i)\}_{i=0,1,2}$$

Recall that the subscript i refers to the radius R_i from Figure 3.6. Moreover, the above map makes sense, as $\text{im}(\partial_{k,\mathcal{K}_i}) \subseteq \ker(\partial_{k-1,\mathcal{K}_i})$ for $i = 0, 1, 2$. Setting $W_i: \text{eqq ker}(\partial_{k-1,\mathcal{K}_i})$ for all i and $f: \text{eqq}\{\partial_{k,\mathcal{K}_i}\}_{i=0,1,2}$, we see that indeed,

$$\{H_k(\mathcal{K}_i)\}_{i=0,1,2} \cong \{W_i\}_{i=0,1,2} / \text{im}(f)$$

is finitely presented.

Another example for finitely presented persistent vector spaces are the interval spaces defined below. We will use them later on to characterise all finitely presented persistent vector spaces.

Definition 3.18 (Persistent Interval Vector Spaces). Let $a < b$ be both positive real numbers, b is allowed to be ∞ . The persistent vector space $P(a, b) = \{P(a, b)_r\}_{r \in \mathbb{R}}$ is given by

$$P(a, b)_r = \begin{cases} K & r \in [a, b) \\ \{0\} & r \notin [a, b), \end{cases}$$

and is called a *persistent interval vector space*.

Lemma 3.19. The spaces $P(a, b)$ are finitely presented.

Proof. Let (X, ρ) and (Y, σ) be \mathbb{R}_+ -filtered sets containing one element each, i.e. $X = \{x\}$ and $Y = \{y\}$. Let $\rho(x) = a$ and $\sigma(y) = b$.

First, assume $b < \infty$. Since $a < b$, the 1×1 identity matrix $[1]$ is a (ρ, σ) -adapted (X, Y) -matrix. Moreover,

$$\theta([1])_R = \begin{cases} \{0\} & R < a \\ K/\{0\} = K & a \leq R \leq b, \\ K/K = \{0\} & a < b \leq R \end{cases}$$

which coincides with the definition of $P(a, b)_R$.

Let now $b = \infty$. Then

$$\begin{aligned} P(a, b)_R &= \begin{cases} K & R \geq a \\ 0 & \text{else} \end{cases} \\ &\cong V_k(X, \rho)_R \\ &= \theta([0])_R. \end{aligned}$$

In both cases, we have shown that $P(a, b)$ is of the form $\theta(\cdot)$, so finitely presented. \square

As we did in Section 2.3 for the homology groups of simplicial complexes, it turns out that we can represent maps between finitely generated, free persistent vector spaces using matrices. The crucial observation to do so, is that when choosing a radius R_∞ large enough, it turns out that

$$V_K(X, \rho)_{R_\infty} = V_K(X),$$

because X is a finite set. This means that a linear transformation of finitely generated, free persistent vector spaces

$$\{f_r\}_{r \in \mathbb{R}}: \{V_K(Y, \sigma)_r\}_{r \in \mathbb{R}} \longrightarrow \{V_K(X, \rho)_r\}_{r \in \mathbb{R}}$$

induces a linear transformation of finite-dimensional, free vector spaces

$$f_\infty: V_K(Y) \longrightarrow V_K(X).$$

We can represent such maps using matrices with rows corresponding to elements of X and columns corresponding to elements of Y . The following definition formally introduces this type of matrix.

Definition 3.20 ((X, Y)-Matrix). *Let X and Y be finite sets. A (X, Y) -matrix is a matrix with rows indexed by the set X and columns indexed by Y . We denote the row corresponding to $x \in X$ by $\text{row}(x)$ and the column corresponding to $y \in Y$ by $\text{col}(y)$.*

Remark 3.21. *A (X, Y) -matrix always has size $|X| \times |Y|$. Moreover, we will label every element in $x \in X$ and $y \in Y$ with their corresponding values $\rho(x)$ and $\sigma(y)$ when we annotate rows and column.*

We denote the (X, Y) -matrix arising from a linear transformation $f = \{f_r\}_{r \in \mathbb{R}}$ of finitely generated, free persistent vector spaces by $A(f)$. We state an important property of such matrices $A(f)$.

Proposition 3.22.

- (a) *Let $f = \{f_r\}_{r \geq 0}$ be a linear transformation of finitely generated, free persistent vector spaces. Then the matrix $A(f)$ has entries $a_{x,y} = 0$ whenever $\rho(x) > \sigma(y)$.*
- (b) *Any (X, Y) -matrix A satisfying the property in (a) uniquely determines a linear transformation f_A of persistent vector spaces $\{V_k(Y, \sigma)_r\}_{r \geq 0}$ and $\{V_k(X, \rho)_r\}_{r \geq 0}$.*

Moreover, the maps $f \mapsto A(f)$ and $A \mapsto f_A$ are inverse to each other.

Proof. Let $f: V_K(Y, \sigma) \rightarrow V_K(X, \rho)$ be a linear transformation of finitely generated persistent vector spaces. Let $y \in Y$, so in particular, y is a basis element of $V_K(Y, \sigma)_{\sigma(y)}$. Applying the linear transformation $f_{\sigma(y)}$ to y yields

$$f_{\sigma(y)}(y) = \sum_{x \in X} a_{x,y}x,$$

which by Lemma 3.15 lies in $V_K(X, \rho)_{\sigma(y)}$ if and only if $a_{x,y} = 0$ for all $x \in X$ with $\rho(x) > \sigma(y)$.

For (b): If we have $\mathbb{R}_{\geq 0}$ -filtered stes (X, ρ) and (Y, σ) , as well as a (X, Y) -matrix with zero entries whenever $\rho(x) > \sigma(y)$, we can uniquely define

$$\begin{aligned} f: V_K(Y, \sigma) &\rightarrow V_K(X, \rho) \\ f_{\sigma(y)}: V_K(Y, \sigma)_{\sigma(y)} &\rightarrow V_K(X, \rho)_{\sigma(y)} \\ y &\mapsto \sum_{x \in X} a_{x,y}x. \end{aligned}$$

□

Definition 3.23 (Adapted Matrix). Let (X, ρ) and (Y, σ) be \mathbb{R}_+ -filtered sets. A (X, Y) -matrix A satisfying $a_{x,y} = 0$ whenever $\rho(x) > \sigma(y)$ is called (ρ, σ) -**adapted**.

Corollary 3.24. There is a one-to-one correspondence:

$$\begin{aligned} \{(\rho, \sigma)\text{-adapted } (X, Y)\text{-matrices}\} &\longleftrightarrow \left\{ \begin{array}{l} \text{linear transformations of finitely} \\ \text{generated, free persistent vector} \\ \text{spaces} \end{array} \right\} \\ A &\longmapsto f_A \\ A(f) &\longleftarrow f \end{aligned}$$

For a (ρ, σ) -adapted (X, Y) -matrix, we define a special, finitely presented persistent vector space

$$\theta(A): \text{eq}V_K(X, \sigma) / \text{im}(f_A).$$

Corollary 3.25. Let (X, ρ) and (Y, σ) be \mathbb{R}_+ -filtered sets and let A be a (ρ, σ) -adapted (X, Y) -matrix.

- (a) $\theta(A)$ is a finitely presented persistent vector space.
- (b) Any finitely presented vector space is isomorphic to a persistent vector space of the form $\theta(A)$ for some A as above.

Corollary 3.26. Let (X, ρ) be a \mathbb{R}_+ -filtered set. The automorphisms of $V_K(X, \rho)$ are in one-to-one correspondence with the invertible (ρ, ρ) -adapted (X, X) -matrices.

Proof. Both corollaries directly follow from the one-to-one correspondence established in Proposition 3.22. □

Example 3.27. We continue analysing the persistent Vietoris-Rips complex from Figure 3.6. We already wrote down all maps $\partial_{1,\mathcal{K}_i}$ for $i = 0, 1, 2$, and are ready to construct the map ∂_1^∞ , which is the boundary map on the full complex, i.e. the complex with no missing faces. In our case this is the complex \mathcal{K}_2 .

$$\partial_1^\infty = \begin{array}{c} v_{1,R_0} \\ v_{2,R_0} \\ v_{3,R_0} \end{array} \begin{array}{ccc} [v_1, v_2]_{R_1} & [v_1, v_3]_{R_2} & [v_2, v_3]_{R_2} \\ \left[\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{array} \right] \end{array}.$$

This is a (ρ, σ) -adapted (X, Y) -matrix for

$$\begin{aligned} X &= \Sigma_0(\mathcal{K}) : \text{eqq} \bigcup_{i=0,1,2} \Sigma_0(\mathcal{K}_i) \\ Y &= \Sigma_1(\mathcal{K}) : \text{eqq} \bigcup_{i=0,1,2} \Sigma_1(\mathcal{K}_i), \end{aligned}$$

ρ and σ assigning to every simplex in X and Y , respectively, the radius of its first appearance. Formally,

$$\begin{aligned} \rho : X &\longrightarrow \mathbb{R}_{\geq 0} \\ x &\longmapsto \min\{R_i : x \in \Sigma_0(\mathcal{K}_i)\} \end{aligned}$$

and

$$\begin{aligned} \sigma : Y &\longrightarrow \mathbb{R}_{\geq 0} \\ y &\longmapsto \min\{R_i : y \in \Sigma_1(\mathcal{K}_i)\}. \end{aligned}$$

The fact that ∂_1^∞ is (ρ, σ) -adapted essentially comes from the definition of simplicial complexes (Definition 2.3): If a simplex is contained in the Vietoris-Rips complex with radius R_i , then all of its faces must appear at latest at radius R_i as well, otherwise the resulting construct is not a well-defined simplicial complex.

Proposition 3.28. Let (X, ρ) and (Y, σ) be \mathbb{R}_+ -filtered sets and let A be a (ρ, σ) -adapted (X, Y) -matrix. Moreover, let P be a (ρ, ρ) -adapted (X, X) -matrix and Q a (σ, σ) -adapted (Y, Y) -matrix. Then PAQ is still a (ρ, σ) -adapted (X, Y) -matrix and $\theta(A) \cong \theta(PAQ)$.

Proof. Let A be a (ρ, σ) -adapted (X, Y) -matrix. Then its entries $a_{x,y}$ are zero whenever $\rho(x) > \sigma(y)$. We denote the entries of P by $p_{x,x'}$ for $x, x' \in X$ and the entries of Q by $q_{y',y}$ for $y', y \in Y$. Consider the entry

$$(PAQ)_{x,y} = \sum_{y' \in Y} \sum_{x' \in X} p_{x,x'} a_{x',y'} q_{y',y},$$

by the rules of matrix multiplication. We have the following possible cases:

$$p_{x,x'} a_{x',y'} q_{y',y} = \begin{cases} 0 a_{x',y'} q_{y',y} = 0 & \rho(x) > \rho(x') \\ p_{x,x'} 0 q_{y',y} = 0 & \rho(x') > \sigma(y') \\ p_{x,x'} a_{x',y'} 0 = 0 & \sigma(y') > \sigma(y) \end{cases}$$

In fact these are already all possible cases (which might overlap): If they all do not apply, this means

$$\rho(x) \leq \rho(x') \leq \sigma(y') \leq \sigma(y),$$

which contradicts adaptedness of A . The uniqueness from Proposition 3.22, Corollary 3.26 and the one-to-one correspondence from Corollary 3.25 imply

$$\theta(A) = V_K(X, \rho) / \text{im}(f_A) \cong V_K(X, \rho) / \text{im}(f_{PAQ}) = \theta(PAQ).$$

□

This proposition will prove to be very useful in our endeavour to replicate Algorithm 2.26 for matrices corresponding to persistent simplicial complexes. The last ingredient we need are row and column operations, with a twist.

Definition 3.29 (Adapted Elementary Operations). *Let (X, ρ) and (Y, σ) be \mathbb{R}_+ -filtered sets and let A be a (ρ, σ) -adapted (X, Y) -matrix.*

- (a) *Let $x, x' \in X$. An **adapted row operation** adds $\text{row}(x)$ to $\text{row}(x')$ in A , under the condition $\rho(x) \geq \rho(x')$. This corresponds to the left multiplication of A by an elementary matrix $e(i, j)$.*
- (b) *Let $y, y' \in Y$. An **adapted column operation** adds $\text{ccol}(y)$ to $\text{col}(y')$ in A , under the condition $\sigma(y) \leq \sigma(y')$. This corresponds to the right multiplication of A by an elementary matrix $e(i, j)$.*

Next, we show that we can decompose any finitely presented persistent vector space $\theta(A)$ into a finite direct sum of persistent interval vector spaces $P(a, b)$, as introduced in Definition 3.18. In order to do so, we will use the adapted elementary operations we just introduced (Definition 3.29) to reduce the matrix A .

Theorem 3.30 (Structure Theorem). *Any finitely presented persistent K -vector space is isomorphic to a finite direct sum*

$$\bigoplus_{i=1}^n P(a_i, b_i).$$

Proof. Let $V = \{V_R\}_{R \in \mathbb{R}}$ be a finitely presented persistent K -vector space. By Corollary 3.25, V is isomorphic to a space $\theta(A)$ for some (ρ, σ) -adapted (X, Y) -matrix A . We decompose A into a (ρ, σ) -adapted (X, Y) -matrix PAQ with at most one non-zero element per row and column, which is 1.

First, we find all $x \in X$ for which $\text{row}(x)$ is non-zero and pick one for which $\rho(x)$ is maximal. Now, consider the $y \in Y$ for which $a_{x,y} \neq 0$ and pick one for which $\sigma(y)$ is minimal.

Thanks to our choice of x , we can apply adapted row operations ad libitum ((b) in Definition 3.29), until the only non-zero entry left in $\text{col}(y)$ is $a_{x,y}$. By our choice of y , we can apply adapted column operations ((c) in Definition 3.29) until the only non-zero entry remaining in $\text{row}(x)$ is $a_{x,y}$.

Since $K = \mathbb{F}_2$, all entries are now one or zero. We delete $\text{row}(x)$ and $\text{col}(y)$ and repeat the procedure for the resulting matrix A' . A' is a (ρ', σ') adapted (X', Y') -matrix, where $X' = X \setminus \{x\}$, $Y' = Y \setminus \{y\}$, ρ' and σ' are the maps ρ and σ restricted to X' and Y' , respectively. Repeating this until we only have a zero matrix left provides a decomposition of A into

$$PAQ = \begin{bmatrix} I_n & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \sum_{i=1}^n \begin{matrix} y_i \\ \begin{bmatrix} 0 & \vdots & 0 \\ \cdots & 1 & \cdots \\ 0 & \vdots & 0 \end{bmatrix} x_i \end{matrix},$$

where n denotes the number of the last step in the procedure with a non-zero matrix.

Using Proposition 3.28 and Corollary 3.25, this translates to

$$\theta(A) \cong \theta(PAQ) \\ \cong \bigoplus_{i=1}^n P(\rho(x_i), \sigma(y_i)) \oplus \bigoplus_{x \notin \{x_1, \dots, x_n\}} P(\rho(x), +\infty) \quad (\text{cf. Lemma 3.19}).$$

□

The following algorithm is a reduction algorithm for adapted matrices and extends Algorithm 2.26 to persistent simplicial complexes.

Algorithm 3.31. Input: A pair $(\partial_k^\infty, \partial_{k+1}^\infty)$ of boundary maps for a persistent simplicial complex, with labels for every row and column indicating the minimal radius for which the corresponding simplex appears.

Output: A pair

$$\left(\begin{matrix} \begin{bmatrix} I_k & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I_m \end{bmatrix} \end{matrix} \right),$$

of matrices with labels as well. As before, I_k and I_m denote the $k \times k$ and $m \times m$ identity matrices.

Step 1: Apply adapted row and column operations to ∂_k^∞ as in the proof of Theorem 3.30, until we get a matrix of the form

$$P\partial_k^\infty Q = \begin{bmatrix} I_k & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

For any row and column operation where we add a row or column to one with a different label, we keep the larger label for the changed row or column. This

makes sense, since a chain of simplices can only exist if all of the simplices in the chain do.

Whenever we apply column operations to ∂_k^∞ , we apply the inverse row operation to ∂_{k+1}^∞ . This means that we get a pair $(P\partial_k^\infty Q, Q^{-1}\partial_{k+1}^\infty)$, where $Q^{-1}\partial_{k+1}^\infty$ is of the form

$$Q^{-1}\partial_{k+1}^\infty = \begin{bmatrix} B_{1,1} & B_{1,2} & B_{1,3} \\ B_{2,1} & B_{2,2} & B_{2,3} \\ B_{3,1} & B_{3,2} & B_{3,3} \end{bmatrix}.$$

As in Algorithm 2.26, $B_{1,1}$ has k rows and we set denote the number of rows of $B_{2,1}$ and $B_{3,1}$ by l and m , respectively. Since $(P\partial_k^\infty Q) \circ (Q^{-1}\partial_{k+1}^\infty) = 0$, $B_{1,1}$, $B_{1,2}$ and $B_{1,3}$ must be zero blocks.

Step 2: Finally, we apply row and column operations to $Q^{-1}\partial_{k+1}^\infty$. We only apply row operations to the $l+m$ last rows and apply the inverse column operations to the $l+m$ last columns of $P\partial_k^\infty Q$. This does not have any effect, as the last $l+m$ columns of $P\partial_k^\infty Q$ are already zero. We switch rows and columns around until we arrive at the form (maybe for a slightly different Q)

$$Q^{-1}\partial_k^\infty S = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I_m \end{bmatrix},$$

whilst adapting the labels in the same way as we did in Step 1.

Proposition 3.32. *The decomposition from Proposition 3.30 is unique up to reordering.*

For the proof, we refer to [Car14], Proposition 3.13.

Example 3.33. *We demonstrate how to apply Algorithm 3.31 to the persistent Vietoris-Rips complex from Figure 3.6. Since we already computed the 0-th persistent homology group $\{H_0(\mathcal{K}_i)\}_{i=0,1,2}$ in Example 3.12, we will now compute the first homology group $\{H_1(\mathcal{K}_i)\}_{i=0,1,2}$.*

In Example 3.27, we already constructed the matrix for ∂_1^∞ , and the matrix for ∂_2^∞ looks the same as $\partial_{2,\mathcal{K}_3}$. So we get the pair

$$(\partial_1^\infty, \partial_2^\infty) = \left(\begin{array}{c} [v_1, v_2]_{R_1} \quad [v_1, v_3]_{R_2} \quad [v_2, v_3]_{R_2} \quad [v_1, v_2, v_3]_{R_2} \\ v_{1,R_0} \left[\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{array} \right], \left[\begin{array}{c} 1 \\ 1 \\ 1 \end{array} \right] \begin{array}{c} [v_1, v_2]_{R_1} \\ [v_1, v_3]_{R_2} \\ [v_2, v_3]_{R_2} \end{array} \end{array} \right),$$

with labels (subscripts next to the simplices) to represent for which radius R_i the simplices first appear.

We apply Algorithm 3.31. First, we add row 1 to row 2 in $\partial_{1,\infty}$. This is allowed, since v_1 and v_2 have the same radius of first appearance. We are left with only one non-zero entry in the first column, so we are happy. Next, we add the first column to the second. This is allowed, because $[v_1, v_2]$ is the first edge to appear. We apply the inverse row operation to $\partial_{2,\infty}$ (which is also a legitimate adapted row operation) and get

$$(e(1,2)\partial_{1,\infty}e(2,1), e(2,1)\partial_{2,\infty}) = \left(\begin{array}{ccc|ccc} [v_1, v_2]_{R_1} & ([v_1, v_3] + [v_1, v_2])_{R_2} & [v_2, v_3]_{R_3} & [v_1, v_2, v_3]_{R_2} & & \\ \hline 1 & 0 & 0 & 0 & & \\ 0 & 1 & 1 & 1 & & \\ 0 & 1 & 1 & 1 & & \end{array} \right).$$

We add the second to the third row in $\partial_{1,\infty}$, then the second column to the third whilst adding the third row to the second in $\partial_{2,\infty}$ and already get the form

$$(P\partial_{1,\infty}Q, Q^{-1}\partial_{2,\infty}) = \left(\begin{array}{ccc|ccc} [v_1, v_2]_{R_1} & ([v_1, v_3] + [v_1, v_2])_{R_2} & (\sum_{i,j}[v_i, v_j])_{R_2} & [v_1, v_2, v_3]_{R_2} & & \\ \hline 1 & 0 & 0 & 0 & & \\ 0 & 1 & 0 & 0 & & \\ 0 & 0 & 0 & 1 & & \end{array} \right),$$

where $P = e(2,3)e(1,2)$ and $Q = e(2,1)e(3,2)$ are products of elementary matrices.

We see that there is one zero column in $P\partial_{1,\infty}Q$, which means that its kernel is spanned by the sum of all three edges, which appears at radius R_2 . Moreover, this sum lies in the image of $Q^{-1}\partial_{2,\infty}$, as the third entry is 1. Since this combination also appears at radius R_2 , the first homology group is zero for all radii. One advantage of the algorithm, is that maps $H_1(L_C(\mathcal{K})(i, j))$, which we computed by hand in Example 3.12 for H_0 are encoded as well in these matrices.

Intuitively, this indicates that the cycle $\sum_{i,j}[v_i, v_j]$ which appears at radius R_2 disappears simultaneously, because of the birth of the face $[v_1, v_2, v_3]$ at the same time.

We will now introduce two ways to represent decompositions (Theorem 3.30) of finitely presented persistent vector spaces in a way which makes it easier to analyse and compare them.

Definition 3.34 (Persistence Diagram). A *persistence diagram* is a multiset of pairs

$$D = \{(a_i, b_i) : i = 1, \dots, n \text{ and } b_i \geq a_i \text{ for all } i\} \cup \{(a, b) : a = b\} \subset \mathbb{R}^2.$$

We denote the multiplicity of $(x, y) \in D$ by $\mu(x, y) \in \mathbb{N} \cup \{\infty\}$, and we set $\mu(x, y) = +\infty$ whenever $x = y$.

Let V be a finitely presented persistent vector space. By Theorem 3.30, we can write V as a direct sum

$$V = \bigoplus_{i=1}^n P(a_i, b_i).$$

Then there is a persistence diagram

$$D = \{(a_i, b_i) : i = 1, \dots, n\} \cup \{(a, b) : a = b\}$$

corresponding to V . The multiplicities in this case are given by

$$\mu_V(x_1, x_2) = \begin{cases} \text{number of occurrences of } (a_i, b_i) \text{ in } \bigoplus_{i=1}^n P(a_i, b_i) & \text{if } x_1 < x_2 \\ +\infty & \text{if } x_1 = x_2. \end{cases}$$

Remark 3.35. *If $V = H_k(\mathcal{K})$ for a persistent simplicial complex \mathcal{K} , then the multiplicity $\mu_V(x)$ of a point $x = (x_1, x_2) \in D(V)$ tells us how many topological features get born at radius (or another continuous parameter) x_1 and die at radius x_2 .*

Another way to represent (isomorphism classes of) finitely presented persistent vector spaces is with persistence barcodes. Starting from the decomposition given by Theorem 3.30, we can gather all intervals $\{(x_1, x_2) : (x_1, x_2) \in D\}$ (the first pair of brackets indicates an open interval, the second a point in \mathbb{R}^2) and we get a persistence barcode.

Definition 3.36 (Persistence Barcode). *A persistence barcode is a multiset of intervals*

$$\mathcal{I} = \{(a_i, b_i) \subset \mathbb{R} : i = 1, \dots, n\},$$

such that every interval has finite multiplicity.

Example 3.37. *In Figure 3.7, we plot the persistence diagram and barcode for the persistent Vietoris-Rips complex \mathcal{K} from Figure 3.6. There are no points/bars for*

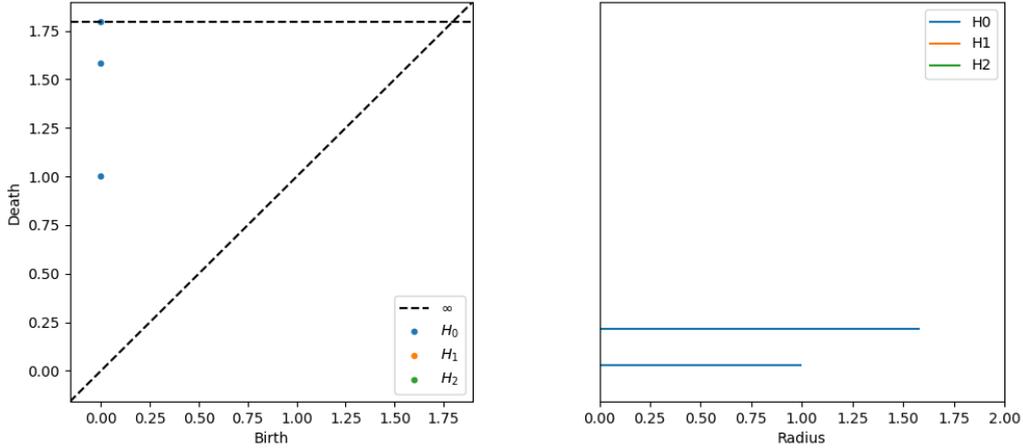


Figure 3.7: In the first subfigure, we can see the persistence diagram corresponding to the Vietoris-Rips filtration $\mathcal{K} = \{\mathcal{K}_i\}_{i=0,1,2}$ from Figure 3.2. In the second subfigure, we added the corresponding barcode.

$H_1(\mathcal{K})$, since there are no one-dimensional topological features which persists at all. In fact, as we explained in Example 3.27, the cycle $[v_1, v_2] + [v_1, v_3] + [v_2, v_3]$ appears at the exact same time as the face $[v_1, v_2, v_3]$, so the first homology group $H_1(\mathcal{K})$ is

trivial at all times. However, the diagram detects the three connected components at the start, two of which merge at R_1 and later merge again, so that we only have one connected component left, which lives on eternally. This is represented by the top point in the persistence diagram and omitted in the barcode, since this feature always exists (if we have at least 1 sample in the data) and therefore has no meaning at all.

Example 3.38. We do another example in pictures. Consider the point cloud in Figure 3.8 on the left. We construct a persistent Vietoris-Rips complex by considering increasing radii $r \geq 0$, compute persistent homology as introduced earlier in this section and obtain the persistence diagram in the middle.

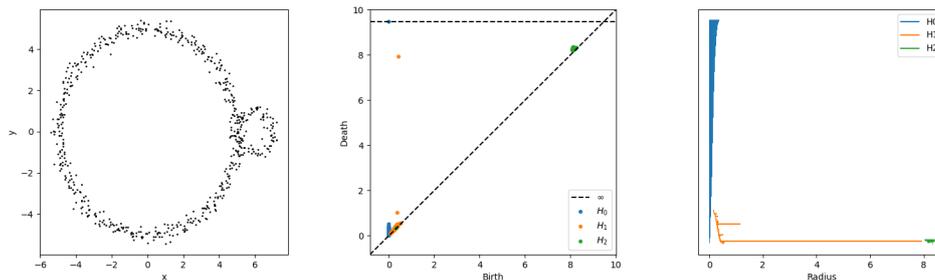


Figure 3.8: In the left plot, we see a point cloud sampled from two circles glued together with added noise. In the middle plot, we show the persistence diagram corresponding to the persistent Vietoris-Rips complex constructed on the point cloud from the left plot. In the rightmost plot, one can see the corresponding persistence barcode.

The persistence diagram has one blue point at infinity, indicating that there is a connected component appearing early on and which does not disappear for larger radii. The two orange dots that are a bit further from the diagonal than the rest of the orange dots reflect the fact that there is a smaller and a larger circle which persist for a wider range of radii, but eventually disappear as well. This is the point where thickening the point cloud just merges everything into one connected component. The points close to the diagonal represent topological features which persist only over a small range of radii and are therefore less relevant. The green dots hint at the fact that there are no 2-dimensional structures of interest in the data.

On the right side, we see an example of a persistence barcode. It corresponds to the persistence diagram and point cloud left of it. The two longer orange bars correspond to the two points in the diagram which are a bit further away from the diagonal, and thus arise from two approximate circles in the point cloud.

3.4 Distances between Persistence Diagrams

In this section, we introduce two notions of distance on the set of persistence diagrams. This can, for example, help us to compare and classify diagrams we obtain

from the construction of persistent Vietoris-Rips complexes on different point clouds, which indirectly provides us with a comparison of the point clouds themselves.

Definition 3.39 (Bottleneck Distance). Let $D \subset \mathbb{R}^2$ and $D' \subset \mathbb{R}^2$ be persistence diagrams and let

$$\Theta = \{\theta: A \rightarrow A' : \theta \text{ is bijective, } A \subseteq D \setminus \{(x, y) : x = y\} \text{ and } A' \subseteq D' \setminus \{(x, y) : x = y\}\}$$

be the collection of bijections between subsets of D and D' . Moreover, set

$$\lambda(x): \text{eqq} \frac{x_2 - x_1}{2} \quad \text{for all } x = (x_1, x_2) \in D \cup D'.$$

Recall that $x_2 \geq x_1$ by definition of persistence diagrams.

The **bottleneck distance** between D and D' is defined as

$$d_\infty(D, D') = \min_{\theta \in \Theta} \max \left(\max_{x \in A} \|x - \theta(x)\|_\infty, \max_{x \in D \setminus A} \lambda(x), \max_{y \in D' \setminus A'} \lambda(y) \right),$$

Where $\|b - a\|_\infty = \max(|b_1 - a_1|, |b_2 - a_2|)$ denotes the l_∞ -distance on \mathbb{R}^2 .

Remark 3.40. $\lambda(x)$ measures the distance between x and the diagonal $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 = x_2\}$. The bottleneck distance thus minimises the maximum l_∞ -norm over all bijections, including topological features that appear and vanish at the same time (this set has continuous cardinality). A sketch to illustrate the bottleneck distance can be found in Figure 3.9.

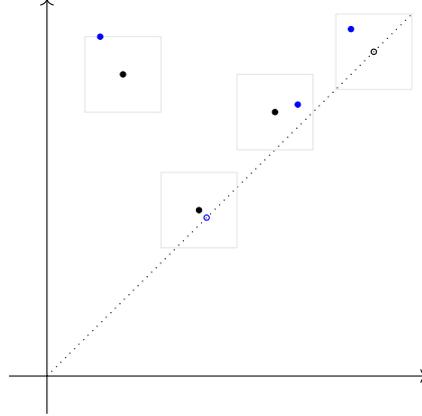


Figure 3.9: Visualisation of the bottleneck distance between two persistence diagrams containing 3 points each (filled black and blue dots). In this case, the distance is given by the l_∞ -distance between the two points on the top left, represented by the gray boxes. The hollow dots on the diagonal correspond to the component $\lambda(x)$ from the definition. They determine the distance between the points from the persistence diagrams and the diagonal (dotted line).

Example 3.41. We consider the two persistence diagrams

$$D = \{(0.3, 1.3), (1.3, 1.6), (3.2, 7.0)\} \cup \{(x, y) : x = y\} \quad \text{and}$$

$$D' = \{(1.0, 1.2), (1.8, 5.5), (4.0, 4.2), 4.4, 5.2\} \cup \{(x, y) : x = y\}.$$

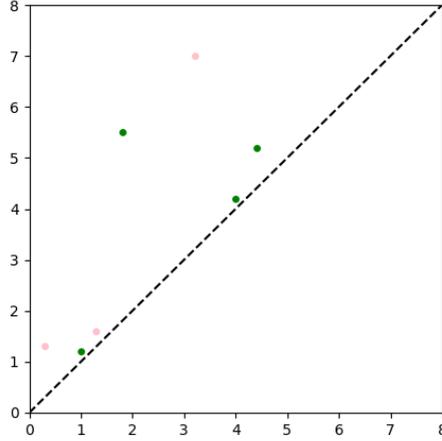


Figure 3.10: The two persistence diagrams D in pink and D' in green from Example 3.41.

Both diagrams are illustrated in Figure 3.10. We define two bijections $\theta_1, \theta_2 \in \Theta$ to illustrate how the bottleneck distance is constructed.

$$\begin{aligned} \theta_1: \quad & (0.3, 1.3) \mapsto (1.0, 1.2) \\ & (1.3, 1.6) \mapsto (4.0, 4.2) \\ & (3.2, 7.0) \mapsto (1.8, 5.5) \\ \theta_2: \quad & (1.3, 1.6) \mapsto (1.0, 1.2) \\ & (3.2, 7.0) \mapsto (1.8, 5.5) \end{aligned}$$

We see that $\max \|x - \theta_1(x)\|_\infty = 2.6$ which comes from the pair $(1.3, 1.6)$ and $(4.0, 4.2)$. Moreover, $\lambda(4.4, 5.2) = 0.4$. This has no impact, since it is smaller than 2.6.

On the other hand, $\max \|x - \theta_2(x)\|_\infty = 1.5$, which comes from the pair $(3.2, 7.0)$ and $(1.8, 5.5)$. As we can guess from the coordinates and Figure 3.10, those points are the ones which determine the bottleneck distance, as they are both far away from all other points and the diagonal. However, $\lambda(3.2, 7.0) = 1.6$ and $\lambda(1.8, 5.5) = 1.85$, so θ_2 finds the smallest distance for those two points. The points x which are not in the range or image of θ_2 have much smaller $\lambda(x)$, so they do not play a significant role in the computation of the bottleneck distance for this example.

Comparing θ_1 and θ_2 makes it clear that θ_2 provides the smaller distance between the two diagrams. In fact, θ_2 determines the bottleneck distance in this case, so

$$d_\infty(D, D') = \max \left(\max_{x \in A} \|x - \theta_1(x)\|_\infty, \max_{x \in D \setminus A} \lambda(x), \max_{y \in D' \setminus A'} \lambda(y) \right),$$

where $A = \{(1.3, 1.6), (1.0, 1.2)\}$ and $A' = \{(1.0, 1.2), (1.8, 5.5)\}$.

Next, we observe that the bottleneck distance arises as the ∞ -version of a class of metrics called **Wasserstein distance**.

Definition 3.42 (Wasserstein Distance). Let D , D' , Θ and $\lambda(\cdot)$ be as in Definition 3.39, and let $p \geq 1$. The p -**Wasserstein distance** between D and D' is given by

$$d_p(D, D') = \left(\min_{\theta \in \Theta} \left(\sum_{x \in A} \|x - \theta(x)\|_\infty^p + \sum_{x \in D \setminus A} \lambda(x)^p + \sum_{y \in D' \setminus A'} \bar{y}^p \right) \right)^{\frac{1}{p}}.$$

Differential Geometry, Dynamical Systems and Time Series

Time series are sequences of measurements sampled from some dynamical system. We assume that the dynamical system describes the smooth evolution of some subspace of \mathbb{R}^d over time. We start this chapter by introducing smooth manifolds, and more importantly, smooth maps.

4.1 Smooth Manifolds and Maps

For the preliminaries in differential geometry, we follow the textbook [Lee12]. Instead of Euclidean space \mathbb{R}^d , we now consider only a topological space. However, since Euclidean space is the environment in which we work best, we aim at relating topological spaces, at least locally, with some subspace of \mathbb{R}^d .

Definition 4.1 (Topological Manifold). *Let M be a topological space. M is called a **topological manifold of dimension d** , if the following conditions are satisfied.*

- (a) *M is a Hausdorff space (also referred to as T_2). This means that for any distinct points $p, q \in M$, there exist disjoint open subsets U, V of M with $p \in U$ and $q \in V$.*
- (b) *M is second-countable, i.e. there is a countable basis of open subsets for the topology on M .*
- (c) *M is locally Euclidean of dimension d . That is, for any point $p \in M$, there exists a neighbourhood of p which is homeomorphic to some open subset in \mathbb{R}^d .*

We clarify the notion of being locally Euclidean by defining charts. By Definition 4.1, a topological manifold can be completely covered with charts.

Definition 4.2 (Chart). *Let M be a topological manifold. A **(coordinate) chart on M** is a pair (U, φ) consisting of an open set $U \subseteq M$ and a homeomorphism $\varphi: U \rightarrow \varphi(U)$, where $\varphi(U) \subset \mathbb{R}^d$ is an open subset. The map φ is often called a **(local) coordinate map**. The components $(\varphi_1, \dots, \varphi_d)$ of φ given by $\varphi(p) = (\varphi_1(p), \dots, \varphi_d(p))$ are called **local coordinates on U** .*

Example 4.3. *We consider the two curves defined as follows. First, a flower-like shape as depicted in Figure 4.1 on the left. It is given by*

$$\gamma_1: [0, 2\pi] \longrightarrow \mathbb{R}^2, \quad t \longmapsto (x_1(t), y_1(t)) : eqqr(t)(\cos(t), \sin(t)),$$

where the radius $r(t)$ is given by the triangle wave

$$r(t) = 1 + 2 \left| \frac{5t}{2\pi} - \left\lfloor \frac{5t}{2\pi} + \frac{1}{2} \right\rfloor \right|.$$

Moreover, we consider the circle with centre $(0, 0)$ and radius 1 (see Figure 4.1 on the right),

$$\gamma_2: [0, 2\pi] \longrightarrow \mathbb{R}^2, \quad t \longmapsto (x_2(t), y_2(t)) = (\cos(t), \sin(t)).$$

We define the map

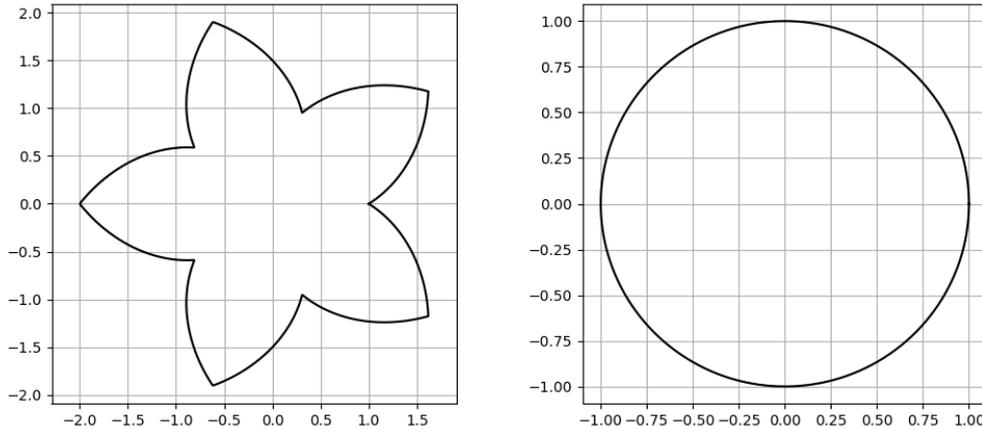


Figure 4.1: Two topological manifolds: the flower-like curve defined in Example 4.3 and a circle with radius $r = 1$.

$$f: \text{im}(\gamma_1) \longrightarrow \text{im}(\gamma_2), \quad (x_1(t), y_1(t)) \longmapsto r(t)(x_1(t), y_1(t)).$$

The function f is continuous and has a continuous inverse

$$f^{-1}: \text{im}(\gamma_2) \longrightarrow \text{im}(\gamma_1), \quad (x_2(t), y_2(t)) \longmapsto \frac{1}{r(t)}(x_2(t), y_2(t)).$$

This is well-defined, because $r(t) > 0$ for all t . So, the flower and the circle are homeomorphic. We define the charts

$$U = \text{im}(\gamma_1) \setminus \{(1, 0)\} \quad \text{and} \quad V = \text{im}(\gamma_2) \setminus \{(0, 1)\}$$

and we the continuous maps

$$\varphi_U: U \longrightarrow (-1, 1), \quad (x, y) \longmapsto \left(\frac{y}{1-x} \right)$$

$$\varphi_V: V \longrightarrow (-1, 1), \quad (x, y) \longmapsto \left(\frac{x}{1-y} \right).$$

These maps are so-called stereographic projections. This shows that the circle is a topological manifold. Pulling back the charts (U, φ_U) and (V, φ_V) through f gives us another pair of charts $(f^{-1}(U), \varphi_U \circ f)$ and $(f^{-1}(V), \varphi_V \circ f)$ for the flower, so it is a topological manifold as well.

Definition 4.4 (Smooth Maps: Euclidean Space). Let $U \subseteq \mathbb{R}^d$ and $V \subseteq \mathbb{R}^e$ be open sets. A map $f: U \longrightarrow V$ is **smooth** (also denoted by C^∞), if all partial derivatives of all components are well-defined and continuous. If f is bijective with smooth inverse map f^{-1} , we say that f is a **diffeomorphism**.

The next restriction we may require, is for the charts to overlap in a way that makes transitioning between different charts from the open cover nicer.

Definition 4.5 (Compatible Charts, Atlas). Let M be a topological manifold and let $(U, \varphi), (V, \psi)$ be two charts on M .

(a) Assume $U \cap V \neq \emptyset$. The map

$$\psi \circ \varphi^{-1}: \varphi(U \cap V) \longrightarrow \psi(U \cap V)$$

is called the **transition map from φ to ψ** .

(b) If $U \cap V = \emptyset$ or $\psi \circ \varphi^{-1}$ is a diffeomorphism, then the charts (U, φ) and (V, ψ) are **smoothly compatible**.

(c) An **atlas** for M is a collection of charts $\mathcal{A} = \{(U_i, \varphi_i)\}_{i \in I}$ such that $M = \bigcup_{i \in I} U_i$. If for every $i, j \in I$, (U_i, φ_i) and (U_j, φ_j) are smoothly compatible, \mathcal{A} is called a **smooth atlas**.

(d) A smooth atlas \mathcal{A} is **maximal**, if it is not strictly included in any other smooth atlas for M . In that case, \mathcal{A} is also called a **smooth structure on M** .

Definition 4.6 (Smooth Manifold). A pair (M, \mathcal{A}) , where M is a topological manifold and \mathcal{A} is a smooth structure on M , is called a **smooth manifold**. The charts contained in the atlas \mathcal{A} are called **smooth charts**.

Example 4.7. The circle from Example 4.3 is a smooth manifold. The charts (U, φ_U) and (V, φ_V) are smoothly compatible on the intersection $U \cap V$. For instance, we have the inverse map of the stereographic projection,

$$\varphi_V^{-1}: t \longmapsto \left(\frac{2t}{1+t^2}, \frac{-1+t^2}{1+t^2} \right),$$

so we can compute

$$\varphi_U \circ \varphi_V^{-1}: t \longmapsto \frac{t^2 - 1}{(t - 1)^2}.$$

Clearly, this is well-defined on $\varphi_V(U \cap V)$. The charts $(f^{-1}(U), \varphi_U \circ f)$ and $(f^{-1}(V), \varphi_V \circ f)$ are not smooth charts, since the function r is a triangle wave, which is not differentiable at the peaks.

Using their the smooth structure, we can extend the definition of smooth maps from Euclidean space to general smooth manifolds.

Definition 4.8 (Smooth Maps). Let M and N be smooth manifolds.

(a) A map $F: M \longrightarrow \mathbb{R}^n$ is **smooth**, if for all $p \in M$, there is a smooth chart (U, φ) , such that $F \circ \varphi^{-1}$ is smooth on $\varphi(U) \subseteq \mathbb{R}^n$.

(b) A map $G: M \longrightarrow N$ is **smooth**, if for any $p \in M$, there are charts (U, ϕ) and (V, ψ) with $p \in U$ and $G(p) \in V$ such that $\varphi(U) \subseteq V$ and $\psi \circ G \circ \varphi^{-1}: \varphi(U) \longrightarrow \psi(V)$ is smooth.

(c) $G: M \longrightarrow N$ is called a **diffeomorphism**, if it is smooth and bijective with smooth inverse function G^{-1} . In that case, M and N are said to be **diffeomorphic**.

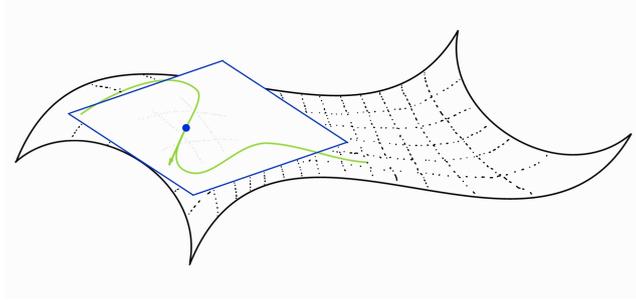


Figure 4.2: A smooth manifold with a highlighted point, its tangent space and a curve determining one of its tangent vectors in that tangent space.

4.2 Tangent Spaces

For any point p of a smooth manifold, the tangent space collects all possible ways for a curve to go through this point. A tangent vector essentially records the direction and the speed at which a certain curve passes through p . We introduce tangent spaces, since when we will look at time series, we will not only be interested in what the samples are per se, but we will want to compare them with the measurements recorded before and after, in order to get an idea of the direction in which the underlying dynamical system flows. Such directional information is best captured with tangent vectors. A sketch of a tangent space to gain some intuition can be found in Figure 4.2. We continue following the textbook [Lee12].

We will only formally introduce tangent spaces for Euclidean space \mathbb{R}^d , as this is enough to understand the results in the later chapters. For a general introduction to tangent spaces of smooth manifolds M , we recommend studying [Lee12], Chapter 3.

Definition 4.9 (Curve, Velocity). Let $I \subset \mathbb{R}$ be an interval. A **curve** in \mathbb{R}^d is a continuous map

$$\gamma: I \longrightarrow \mathbb{R}^d.$$

If γ is differentiable at t_0 , its **velocity** at $t_0 \in I$ is given by

$$\gamma'(t_0) = \left. \frac{d}{dt} \gamma(t) \right|_{t=t_0}.$$

The value $\|\gamma'(t_0)\|_2$ can be interpreted as the speed of the curve γ at the time t_0 without taking direction into account, and the unit vector $\gamma'(t_0)/\|\gamma'(t_0)\|_2$ indicates the direction of the curve at t_0 while ignoring its speed. A tangent vector at a point $p \in \mathbb{R}^d$ is nothing other than the velocity of a curve through the point p .

Example 4.10. In Figure 4.3, we can see the plot of a curve in \mathbb{R}^2 and some of its tangent vectors. The curve is defined by

$$\gamma: [0, 10] \longrightarrow \mathbb{R}^2, \quad t \longmapsto (\sin(2t)/(t+4), \cos(2t)/(t^2+4)).$$

We compute its tangent vector at $t = 0$.

$$\begin{aligned}\gamma'(0) &= \frac{d}{dt} \left(\frac{\sin(2t)}{t+4}, \frac{\cos(2t)}{t^2+4} \right) \Big|_{t=0} \\ &= \left(\frac{2\cos(2t)}{t+4} - \frac{\sin(2t)}{(t+4)^2}, \frac{-2\sin(2t)}{t^2+4} - \frac{2t\cos(2t)}{(t^2+4)^2} \right) \Big|_{t=0} \\ &= (0.5, 0).\end{aligned}$$

So, this tangent vector is parallel to the x -axis, which can also be observed in Figure 4.3.

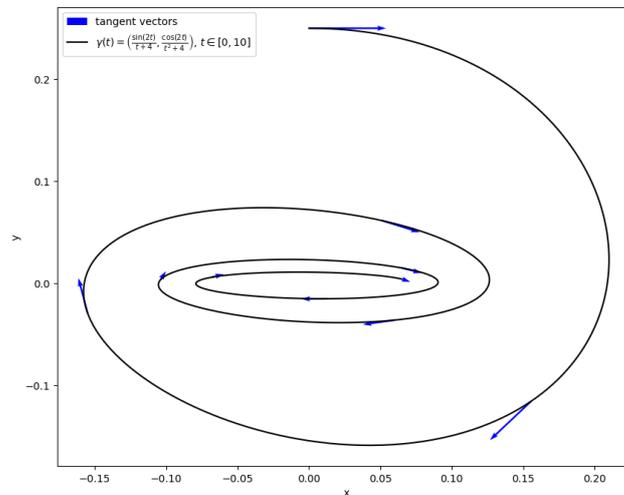


Figure 4.3: The curve $\gamma(t) = (\sin(2t)/(t+4), \cos(2t)/(t^2+4))$ for $t \in [0, 10]$ with a few tangent vectors (blue arrows).

Tangent spaces collect the tangent vectors for all curves through one point. In Figure 4.4, we can see a sketch with a point $p \in \mathbb{R}^2$ and three different curves going through p with tangent vectors at the point p . This should already motivate the fact that for a point in \mathbb{R}^d , there are tangent vectors in all directions and of all lengths.

Definition 4.11 (Tangent Space). Let $p \in \mathbb{R}^d$ be a point. The *tangent space* $T_p\mathbb{R}^d$ of \mathbb{R}^d at p is given by

$$T_p\mathbb{R}^d = \{v \in \mathbb{R}^d : p = \gamma(t_0) \text{ and } v = \gamma'(t_0) \text{ for some curve } \gamma: I \rightarrow \mathbb{R}^d \text{ and } t_0 \in I\}.$$

We go one step further and collect the tangent spaces for every point $p \in \mathbb{R}^d$. The resulting set is called the tangent bundle.

Definition 4.12 (Tangent Bundle and Unit Tangent Bundle).

(a) The *tangent bundle* of \mathbb{R}^d is given by

$$T\mathbb{R}^d = \{(p, v) : p \in M, v \in T_pM\}.$$

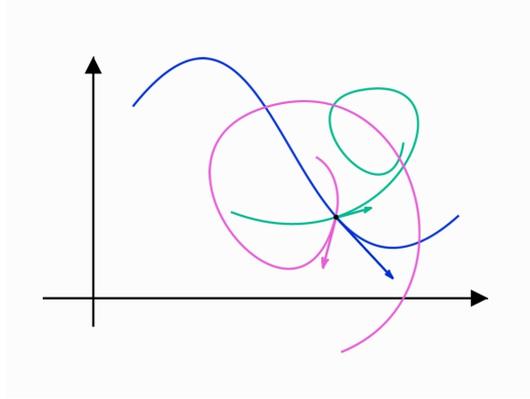


Figure 4.4: Three different curves going through a point $p \in \mathbb{R}^2$, with corresponding tangent vectors at p .

(b) The **unit tangent bundle** of M is a subset of the tangent bundle, which only takes into account tangent vectors of length one. Formally,

$$UT\mathbb{R}^d = \{(p, v) : p \in M, v \in T_pM \text{ and } \|v\|_2 = 1\}.$$

The following theorem formalises what we already hinted at. The tangent spaces $T_p\mathbb{R}^d$ contain vectors of all lengths and directions, so it makes sense that $T_p\mathbb{R}^d \cong \mathbb{R}^d$. Since in the unit tangent bundle, we restrict the tangent vectors to have length one, it makes sense that the tangent spaces contained in it are spheres S^{d-1} . If we collect the (unit) tangent spaces for all points in \mathbb{R}^d , we get the following result.

Theorem 4.13 (Real Tangent Bundles).

(a) The tangent bundle $T\mathbb{R}^d$ is diffeomorphic to $\mathbb{R}^d \times \mathbb{R}^d$.

(b) The unit tangent bundle $UT\mathbb{R}^d$ is diffeomorphic to $\mathbb{R}^d \times S^{d-1}$.

We refer to chapter 3 in [Lee12] for a proof.

The definition of the unit tangent bundle requires a norm. So far, we have always used the Euclidean or 2-norm, which is standard for spaces \mathbb{R}^d . We will introduce unit tangent bundles with respect to the ∞ -norm as well, as the ∞ -norm is better suited for computation. We recall the definitions of both norms.

Definition 4.14 (2- and ∞ -Norm). Let $x = (x_1, \dots, x_d)$ be an element of \mathbb{R}^d .

(a) The **2-norm** of x is given by

$$\|x\|_2 = \sqrt{x_1^2 + \dots + x_d^2}.$$

(b) The **∞ -norm** of x is given by

$$\|x\|_\infty = \max_{i \in \{1, \dots, d\}} |x_i|.$$

It is clear from the definition that ∞ -norms are easier to compute. Computation of the 2-norm requires squares and roots, while we merely need to find a maximal element in a vector for the ∞ -norm. The unit tangent bundle with respect to the ∞ -norm is given by

$$UT_{\infty}\mathbb{R}^d = \{(p, v) \in T\mathbb{R}^d : \|v\|_{\infty} = 1\}.$$

The following lemma links the two unit tangent bundles and will be part of the justification why we can use ∞ -norms for computation without any loss of significance.

Lemma 4.15. *There is a homeomorphism*

$$\eta: UT\mathbb{R}^d \longrightarrow UT_{\infty}\mathbb{R}^d, \quad (p, v) \longmapsto \left(p, \frac{v}{\|v\|_{\infty}}\right).$$

Proof. For the proof, we note that

$$\eta^{-1}: UT_{\infty}\mathbb{R}^d \longrightarrow UT\mathbb{R}^d, \quad (p, v) \longmapsto \left(p, \frac{v}{\|v\|_2}\right)$$

is an inverse map to η . Moreover, both η and η^{-1} are continuous. \square

4.3 Dynamical Systems

As already mentioned in the introduction, dynamical systems are a central object of interest of this thesis. A time series is a sequence of data with time stamps sampled from such a dynamical system. In order to understand how to handle such data, we introduce a few basic notions from the field of dynamical systems to get a feeling of what they are. We follow [BS02] and [Lee12].

Definition 4.16 (Dynamical System).

- (a) A **discrete-time dynamical system** is a pair (X, φ) , where X is a non-empty set called the **phase space** and $\varphi: X \rightarrow X$ a self-map. For any $n \in \mathbb{N}$, we call the n -fold composition $\varphi^n = \varphi \circ \varphi \circ \dots \circ \varphi$ the **n -th iterate** of φ , with φ^0 set to be the identity map. If φ is invertible, we can also choose n to be a negative integer by iterating the inverse map φ^{-1} .
- (b) A **continuous-time dynamical system** is a pair (X, φ) consisting of a set X and a map $\varphi: \mathbb{R} \times X \rightarrow X$, such that $\varphi(0, x) = x$ and $\varphi(s+t, x) = \varphi(s, \varphi(t, x))$.

Definition 4.17 (Orbit). Let (X, φ) be a dynamical system and let $x \in X$ be a point. Then the **orbit through** x is the set

$$\mathcal{O}_{\varphi}(x) = \{\varphi(t, x) : t \in I\} \subseteq X,$$

where $I \subseteq \mathbb{N}$ or $I \subseteq \mathbb{Z}$ if (X, φ) is discrete-time and $I \subseteq \mathbb{R}$, if (X, φ) is continuous-time. There are the following possible special cases.

(a) If there is a $T > 0$, such that $\varphi(T, x) = x$, then x is called a **periodic point of period** (or **length**) T . In that case, the orbit through x is called a **periodic orbit**.

(b) A point $x \in X$ is called a **fixed point** of (X, φ) , if $\varphi(t, x) = x$ for all t .

Example 4.18 (Double Well). An example of a dynamical system we will come back to a few times is the so-called (**perturbed**) **double well**.

In order to write down the differential equation giving rise to it, we define Wiener processes.

A **d -dimensional Wiener Process** is a stochastic process

$$\{W_t = (W_t^{(1)}, \dots, W_t^{(d)})\}_{t \geq 0}$$

with the following properties.

(a) $W_0 = 0$

(b) The map $t \mapsto W_t$ is continuous in t almost surely.

(c) The process $\{W_t\}_{t \geq 0}$ has stationary, independent increments.

(d) The increment $W_{t+s} - W_s$ has normal distribution $\mathcal{N}(0, t)$.

Wiener processes arise as limits of random walks, but also in other settings.

Consider the following stochastic differential equation. The constant σ is a noise amplitude coefficient which controls the random perturbation introduced by the Wiener process W .

$$dx = f(x)dt + \sigma dW$$

The solution to this differential equation is a so-called **double well Hamiltonian system** given by

$$f(x) = \begin{pmatrix} H_p(x) \\ -H_q(x) \end{pmatrix} - ah(H(x)) \begin{pmatrix} H_q(x) \\ H_p(x) \end{pmatrix},$$

where $h(z) = (z^3 - z)/2$, H is the Hamiltonian function

$$H(q, p) = p^2/2 + q^4/8 - q^2/2 - q^3/15 - q/10,$$

and H_p, H_q denote the partial derivatives. We note that this system is special, as it does not depend on time. This just means that the speed of its motion does not change. It is nevertheless a well-defined dynamical system. A plot of an orbit of this dynamical system in its phase space \mathbb{R}^2 can be found in Figure 4.5.

There are two standard ways to plot dynamical systems. The first is to plot the orbits $\varphi^n(x)$ component-wise against $n \in \mathbb{N}_0$ or $n \in \mathbb{Z}$, or in the case of a continuous-time dynamical system, $\varphi(t, x)$ against $t \in \mathbb{R}$. The second is to plot the phase space X and the curve $\gamma(t): eqq\varphi(t, x)$. The second version is often called a phase space

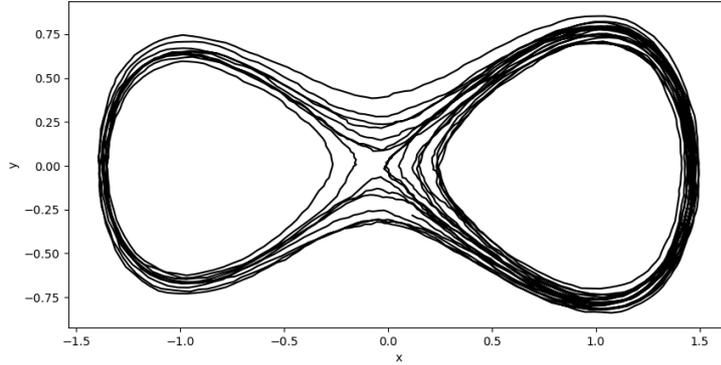


Figure 4.5: Plot of an orbit of the perturbed double well system introduced in Example 4.18.

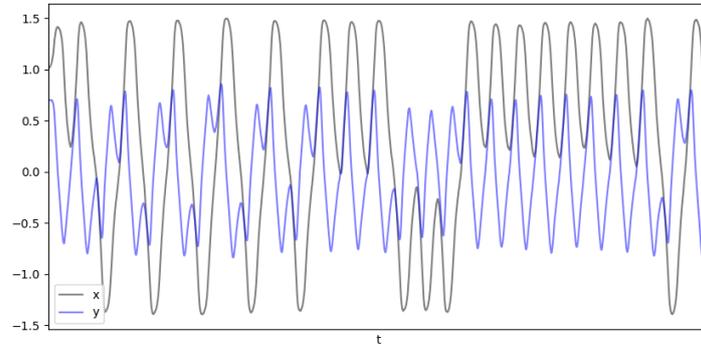


Figure 4.6: Component-wise plot of the perturbed double well system from Example 4.18 to illustrate the two types of standard plots for dynamical systems.

plot. The plot in Figure 4.5 is a phase space plot, and we demonstrate in Figure 4.6 how the other version looks for this example.

In this thesis, we will be dealing with a special type of continuous-time dynamical systems, namely flows. They form a crossing point between the fields of dynamical systems and differential geometry.

Definition 4.19 (Flow). *Let M be a manifold. A (global) flow on M is a continuous map $\Phi: \mathbb{R} \times M \rightarrow M$ that satisfies the following two conditions.*

(a) $\Phi(s + t, p) = \Phi(s, \Phi(t, p))$ for all $s, t \in \mathbb{R}$ and $p \in M$.

(b) $\Phi(0, p) = p$ for all $p \in M$.

Example 4.20. *The perturbed double well system introduced in Example 4.18 is a smooth flow, since the functions h and H are smooth maps.*

4.4 Time Series

Now that we know what a dynamical system is, we will introduce time series. In principle, we just take samples from a dynamical system at certain time stamps and this already gives us a time series. We follow [BD16], [Tak81] and [Bau+23b].

Definition 4.21 (Time Series). *A **time series** is a sequence $\{x_t\}_{t \in I}$ of observations recorded at time t from some dynamical system. It is called **discrete-time**, if the index set I is discrete and **continuous-time**, if I is continuous.*

In real life, we are always given finite time series, so we will always assume $I = \{0, \dots, T\}$ for some $T \in \mathbb{N}$.

The most simple types of time series occur when $x_t \in \mathbb{R}$ are just scalars. But often, the sample might come from a higher-dimensional space, or there might be several factors observed at the same time, making the time series multivariate. So x_t could also be a tuple or even a whole table of data, all corresponding to one single point in time.

When we have a given time series, we want to be able to extract as much information as possible about the underlying dynamical system from it. The following result was proved in 1981 by Floris Takens, see [Tak81] and provides us with a first approach to extract said information.

Theorem 4.22 (Takens' Embedding). *Let M be a compact manifold of dimension m , and let (Φ, β) be a pair consisting of a smooth diffeomorphism $\Phi: M \rightarrow M$ and a smooth function $\beta: M \rightarrow \mathbb{R}$. Then the map*

$$\Psi_{(\Phi, \beta)}: M \longrightarrow \mathbb{R}^{2m+1}, \quad x \longmapsto (\beta(x), \beta(\Phi(x)), \dots, \beta(\Phi^{2m}(x)))$$

is an embedding.

Remark 4.23. β is often called **observation function**.

Remark 4.24. $\Psi_{(\Phi, \beta)}$ being an embedding means that its domain M is diffeomorphic to $\text{im}(\Psi_{(\Phi, \beta)})$. So the whole topological (and differential) structure is preserved.

What the theorem essentially tells us in the setting of finite time series, is that if we have a given real time series

$$\{x_t\}_{t \in \{0, \dots, T\}} = \{\beta(\Phi^t(x))\}_{t \in \{0, \dots, T\}},$$

we can extract all information about the underlying dynamical system (M, Φ) , at least for this finite time span, by considering the observation at time t together with a certain number of following observations.

One difficulty in the application of Theorem 4.22 is to find a good embedding dimension $d \geq 2m + 1$ (note that we do not know m). A problem that might arise is that points that are not actually neighbours in a higher dimension might look like they are when projected to a lower dimension. One possibility to avoid this issue

(mentioned in [RC19]) is to pick the smallest positive integer d for which the nearest neighbours as well as the corresponding distances stay the same when looking at the embedding in dimension $d + 1$.

Moreover, in practice, we often choose a time delay parameter τ , and then apply Takens' theorem to the modified time series. Its purpose is to make the computation stable when we have noisy data (by Theorem 4.22, there is no time lag needed if the data is precise). It is usually chosen to be the smallest time lag for which x_t and $x_{t+\tau}$ seem to correlate very little.

There are several other methods to find d and τ , but for now, let us assume we know their ideal values. The theorem then tells us that having all points of M is equivalent to having all tuples

$$(x_t, x_{t+\tau}, \dots, x_{t+(d-1)\tau}).$$

Example 4.25. *We will apply Takens' theorem to the Lotka-Volterra system given by the following differential equations.*

$$\begin{aligned} \frac{dx}{dt} &= x - 0.4xy \\ \frac{dy}{dt} &= -y + 0.3xy, \end{aligned}$$

where x represents the population size of a prey species (e.g. rabbits) and y the population size of a predator species (e.g. foxes). Moreover, we set the initial populations to be 10 for both species. A visualisation can be seen in Figure 4.7.

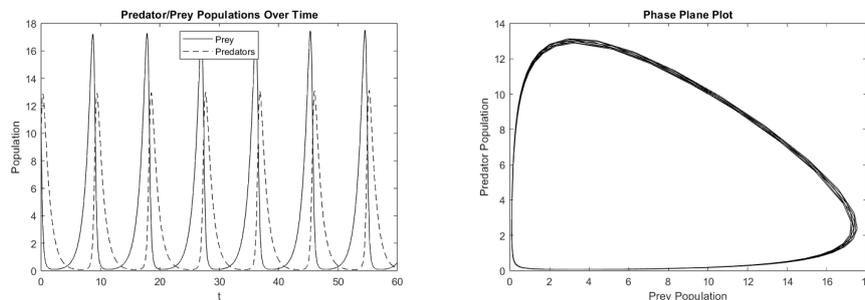


Figure 4.7: Population dynamics and phase space plots for the Lotka-Volterra Predator-Prey model stated in Example 4.25.

For this example's sake, we took samples from the prey population and then reconstructed the phase space from it, which allowed to compute persistent homology, as can be seen in Figure 4.8. In this example, persistent homology helps us capture the cyclic motion in the phase space, which means we detected periodicity in the system. In Chapter 5, we will build on this idea and study a more refined method to detect cycles in time series.

Let $\Gamma = \{x_t\}_{t=0, \dots, T}$ be a time series sampled from a smooth flow (X, Φ) , where $X \subseteq \mathbb{R}^d$. Then there is an increasing sequence $\{t_i\}_{i=1, \dots, T}$ and an initial point x_0 ,

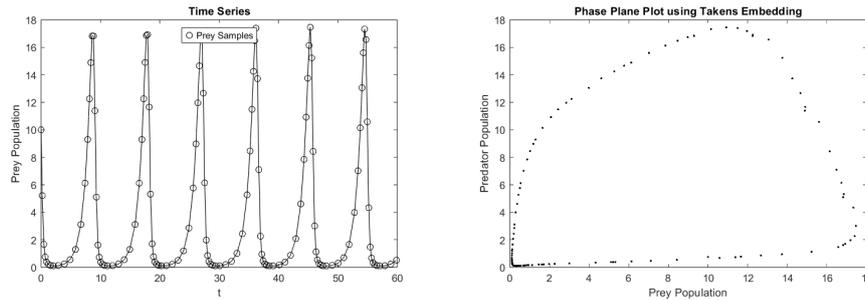


Figure 4.8: The left plot shows a time series sampled from the prey population of the Lotka-Volterra system stated in Example 4.25 consisting of 126 equidistant observations (distance 3). The plot on the right shows the phase space reconstruction with the help of Takens' Embedding, using $d = 2$ and $\tau = 3$.

such that

$$x_i = x(t_i) = \Phi(x_0, t_i) = \Phi(x_{i-1}, t_i - t_{i-1}).$$

In practice, we encounter time series with noise, so the above equations are only approximate.

Definition 4.26 (Segment of a Time Series). A *segment* of Γ is a consecutive sub-time series x_k, x_{k+1}, \dots, x_l for some $k, l \in \{0, \dots, T\}$ and $k \leq l$. Its **time span** is given by $t_l - t_k$.

We now consider a fragment $\gamma = \Phi([a, b], x)$ of the orbit of Φ that goes through some $x \in X$. The following lemma is a simple but crucial observation.

Lemma 4.27. *The following are equivalent.*

- (a) γ contains a periodic orbit.
- (b) $H_1(\gamma) \neq 0$.

Proof. A simplicial complex homeomorphic to a curve is always one-dimensional, connected and contains either zero or two loose edges (i.e. which are attached to the complex on one side only) by default.

γ contains a periodic orbit if and only if it is homeomorphic to a connected simplicial complex γ^Δ containing at least one cycle, which is the case exactly when the first homology group $H_1(\gamma^\Delta) = H_1(\gamma)$ is non-zero. \square

In order to make use of this lemma in the context of time series, we consider a segment $\gamma \subseteq \Gamma$ of length $l - k$. This gives us a point cloud with points on the fragment $\Phi([t_k, t_l], x_k)$ of the orbit $\mathcal{O}_\Phi(x_k)$. In order to detect cycles, we will make use of persistent homology methods.

Cycling Signatures

We have gathered all preliminaries needed to finally introduce cycling signatures. Their study is the main focus of this thesis. They were introduced by Bauer et al. ([Bau+23b]) in 2023. In one sentence, the idea behind cycling signatures is to use persistent homology to detect cycles in time series data, and to distinguish between qualitatively different cycles by computing their homology as subsets of a larger comparison space.

Let Γ be a time series in \mathbb{R}^d , which we assume is sampled from a flow (X, Φ) . The first thing we do is to compute thickenings of the points contained in this time series. We make use of the fact that the time series comes from a flow by adapting the metric used to compute the thickenings.

Example 5.1. *Consider the flow illustrated in Figure 5.1. Clearly, the points are sampled from one single loop. But using balls in \mathbb{R}^2 for the thickenings as in the picture, we see that they merge in the middle, since the points there are a little closer. So, we detect two cycles instead of one. This does not make much sense, as the merged balls correspond to points that are at very different places in the flow (depicted on the left), and in particular, they evolve in opposite directions. This illustrates why it makes sense to take directional information into account when deciding on what metric to use for the thickenings. The metric we choose is then the one we use to construct the open balls from Definition 3.5.*

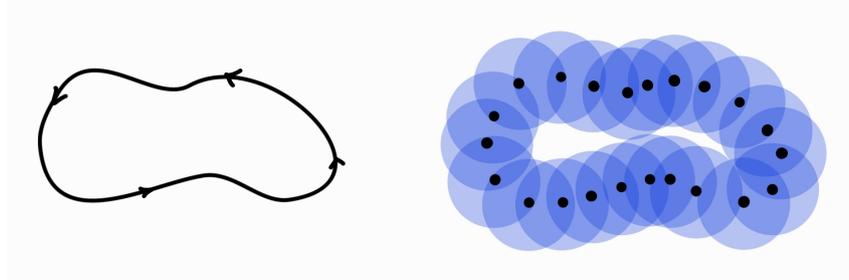


Figure 5.1: On the left side, we see the plot of a flow with arrows to indicate direction. The black dots in the right subfigure are samples taken from the flow on the left, and the transparent purple balls represent a thickening of the samples.

5.1 Finding a Good Metric

The goal of this section is to find a metric that takes both the position of a point and the direction given by the flow Φ (in practice, we do an approximation using a time series) into account. For this, we make use of tangent vectors (Section 4.2).

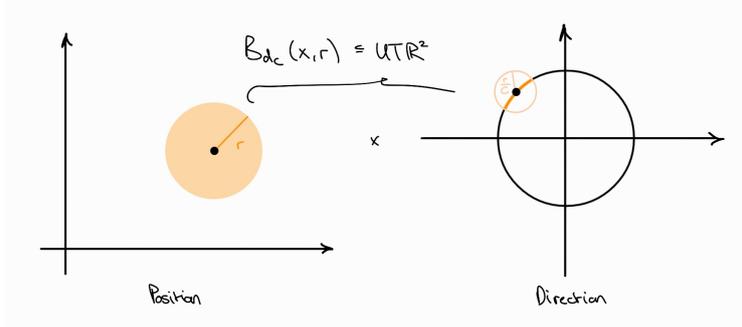


Figure 5.2: Illustration of a ball of radius r with respect to the metric d_C .

We recall that elements of the tangent bundle $(p, v) \in \mathbb{R}^d \times \mathbb{R}^d$ can be thought of as a pair containing the point $p \in \mathbb{R}^d$ and the velocity v given by some curve through p . In our scenario, the curves of interest are fragments of orbits of Φ . We follow [Bau+23b], Section 2.1.

For any point $x \in X$ which is not a fixed point of Φ , we define the non-zero tangent vector

$$v(x) = \left. \frac{d}{dt} \Phi(t, x) \right|_{t=0}.$$

Let $X_{fix} \subseteq X$ denote the set of fixed points in (X, Φ) . We define the map

$$\rho: X \setminus X_{fix} \longrightarrow UTX, x \longmapsto \left(x, \frac{v(x)}{\|v(x)\|_2} \right). \quad (5.1)$$

We consider the following distance function d_C on $\mathbb{R}^d \times S^{d-1}$, which induces a metric on $\text{im}(\rho)$ by restriction. For (p, v) and (q, w) in $UT\mathbb{R}^d$, set

$$d_C((p, v), (q, w)) := \max\{\|p - q\|_2, C\|v - w\|_2\}.$$

See Figure 5.2 for a sketch of how a ball of radius r looks with respect to the metric d_C , and Figure 5.3 for a Čech complex constructed with respect to this metric, next to the Čech complex without modification. We observe that the two points with opposed directions are not connected in the new Čech complex, while they are connected by an edge to the third point which is further away in its spatial component. We will discuss in Section 5.5 how to choose C in practice.

5.2 Cycling Segments and Classification

Using the metric found in the first section, we proceed to the computation of persistent homology groups. Since cycles are one-dimensional objects, we only look at the first homology groups $\{H_1(U(X, r))\}_{r \geq 0}$. We have all ingredients to define cycling segments, as introduced in Sections 2.1 and 2.2 in [Bau+23b]. Recall from Lemma 4.27 that a fragment $\gamma = \Phi([a, b], x)$ contains an orbit if and only if it has non-trivial first homology. Since in our case, γ is a segment of a time series, thus not a continuous curve, we rely on thickenings to detect periodic orbits.

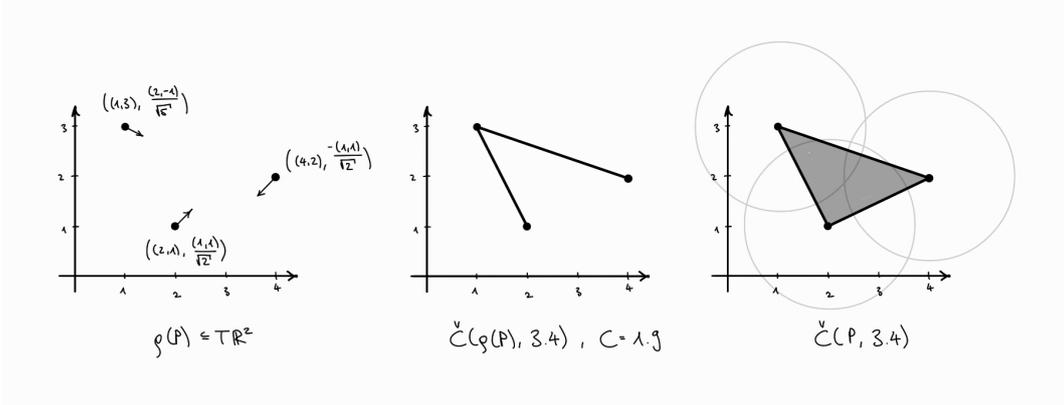


Figure 5.3: In the left coordinate system, we see three points P with unit tangent vectors, so it is a sketch of $\rho(P) \in \mathbb{R}^4$. In the middle, we show the Čech complex $\check{C}(\rho(P), 3.4)$ with respect to the distance d_C for $C = 1.9$. For comparison, we added the standard Čech complex on the three points with the same radius as a third subfigure.

Definition 5.2 (Cycling). A segment $\gamma \subseteq \Gamma = \{x_t\}_{t=0, \dots, T}$ is *r-cycling*, if

$$H_1(U(\gamma, r)) \neq 0,$$

where the construction of the thickening is done with a suitable metric d_C .

Once we find cycling segments in a time series Γ , the next step is to compare and classify them. We call a collection of similar cycling segments in Γ an oscillation. The comparison of oscillations requires a few additional notions.

Definition 5.3 (Comparison space). Let Γ be a time series with values in $X \subseteq \mathbb{R}^d$. A comparison space for Γ is a neighbourhood $Y \subseteq UTX$ of $\rho(\Gamma)$, where ρ is the map from (5.1).

A sketch to illustrate comparison spaces (first component only) can be seen in Figure 5.4.

Assume we have a segment $\gamma \subseteq \Gamma$ and a given radius $r > 0$. For r small enough, there is an inclusion map

$$i_{\gamma, r}: U(\rho(\gamma), r) \longrightarrow Y,$$

which by functoriality induces a map

$$H_1(i_{\gamma, r}): H_1(U(\rho(\gamma), r)) \longrightarrow H_1(Y).$$

The image of this inclusion defines a subspace $\text{im}(H_1(i_{\gamma, r}))$ of $H_1(Y)$.

Collecting the inclusions for all $r > 0$ and passing to homology like we just demonstrated, we get a linear transformation $\{H_1(i_{\gamma, r})\}_{r>0}$ of persistent vector spaces. This allows us to define cycling signatures.

Definition 5.4 (Cycling Signature). Let γ be a segment of a time series Γ , and let Y be a comparison space of Γ .

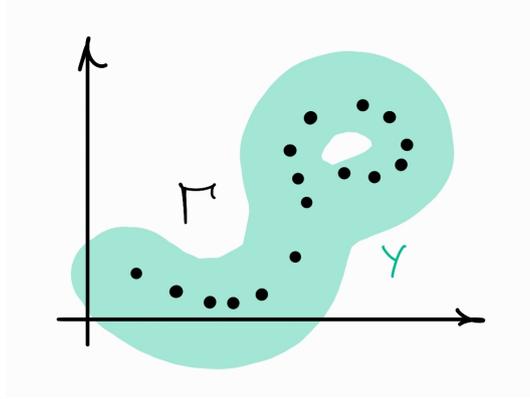


Figure 5.4: A time series (black dots) in \mathbb{R}^2 and a comparison space Y . We note that the sketch only captures the spatial component of the comparison space.

(a) The *r-cycling space* of γ with respect to Y is given by

$$\text{Cyc}_r(\gamma, Y) = \text{im}(H_1(i_{\gamma,r})).$$

(b) The *r-cycling rank* of γ is $\dim(\text{Cyc}_r(\gamma, Y))$.

(c) $H_1(Y)$ is called a **homological comparison space** for Γ .

(d) The *cycling signature* of γ in Y is the persistent vector space

$$\text{Cyc}(\gamma, Y) = \{\text{Cyc}_r(\gamma, Y)\}_{r \in [0, r_0(Y))},$$

where r_0 is a choice of maximal radius that makes sense in the comparison space Y . This will be discussed in Section 5.5.

In practice, we will see that we can compute a basis of the cycling signature using Proposition 5.24.

Example 5.5. We consider the double well system introduced in Example 4.18. Figure 5.5 shows a time series sampled from said system, which contains 3000 points. The data are provided together with the Julia code used in [Bau+23b] on the GitHub repository [Hie23]. The next figure, Figure 5.6, illustrates why for this example, we obtain three different rank 1 signatures. The blue marking represents all cycles going around the left loop of the comparison space once, the green marking represents those which go around the right loop, and the red one represents the segments that go around both.

Let Y be a comparison space which is fine enough that it admits the two obvious voids from the time series. We see that this is the case in Figure 5.6, although the picture shows a cubical comparison space Y_∞ , as we will introduce later. However, it is still useful to encourage intuition. Let g_b and g_g be the generators of $H_1(Y)$ corresponding to the voids inside the blue and green loops. Then the cycling signature of the blue segment is the subspace of $H_1(Y)$ generated by g_b , the cycling signature of the green segment is generated by g_g , and the red segment has a cycling signature generated by the sum of both.

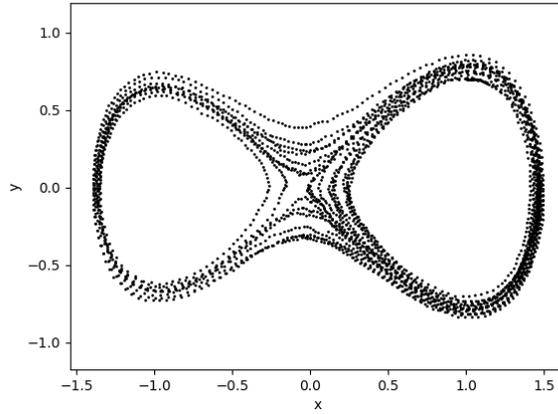


Figure 5.5: 3000 samples from a double well system as described in Example 4.18.

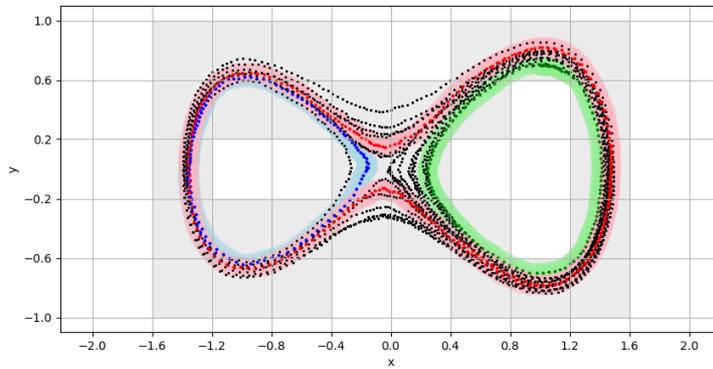


Figure 5.6: The blue, green and red markings indicate the three types of rank 1 cycles that arise in the double well time series. The grey boxes correspond to the spatial component of a cubical comparison space Y_∞ (see Section 5.4).

The higher-rank signatures correspond to segments which are combinations of rank 1 cycles. A longer sequence of the time series which goes along a blue cycle and then a red one has a cycling signature of rank 2 that is generated by both g_b and g_r . This also motivates that the rank 1 signatures are subspaces of the rank 2 signatures. Finally, since the vector space $H_1(Y)$ has dimension 2 (two voids) in this example, there cannot be cycling spaces with cycling rank > 2 .

5.3 Internal and External Thickenings

We continue following [Bau+23b]. It is not clear yet how we can compute cycling signatures in practice. The following diagram shows the detour we are going to take.

$$\begin{array}{ccc}
 \{H_1(U(\rho(\gamma), r))\}_{r \geq 0} & \xrightarrow{H_1(i_\gamma)} & H_1(Y) \\
 \cong \uparrow & & \uparrow \cong \\
 \{H_1(\check{C}(\rho(\gamma), r))\}_{r \geq 0} & \xrightarrow{H_1(\phi)} & H_1(Y_\infty)
 \end{array} \tag{5.2}$$

We already know what happens in the top row. We will first address the left vertical arrow. The Čech complex introduced in Definition 3.4 is a simplicial complex, which is a combinatorial object and thus computable. With the nerve theorem (Theorem 3.6), we have shown that there is topological equivalence between Čech complexes and thickenings.

So far, we have not paid much attention to the fact that our thickenings are taken in a unit tangent space. We recall the map

$$\rho: X \setminus X_{fix} \subseteq \mathbb{R}^d \longrightarrow UTX, \quad x \longmapsto \left(x, \frac{v(x)}{\|v(x)\|_2} \right).$$

There are two possible ways to define thickenings, which we will use interchangeably after providing justification in this section.

Definition 5.6 (Internal and External Thickenings). *Let γ be a segment of a time series Γ .*

(a) The **internal ball** of radius r around a point $x \in UTX$ is given by

$$B_C^I(x, r) = \{x' \in UTX : d_C(x, x') \leq r\}.$$

The **internal r -thickening** of $\rho(\gamma) \subseteq \Gamma$ is the union of internal balls

$$U_C^I(\rho(\gamma), r) = \{p \in UTX : d_C(p, \rho(\gamma)) \leq r\}.$$

(b) The **external ball** of radius r around a point $(x, v(x)) \in UTX$ is given by

$$B_C^E(x, r) = \{x' \in TX : d_C(x, x') \leq r\}.$$

The **external r -thickening** of $\rho(\gamma) \subseteq \Gamma$ is the union of external balls

$$U_C^E(\rho(\gamma), r) = \{p \in TX : d_C(p, \rho(\gamma)) \leq r\}.$$

We refer to Figure 5.7 for a visual comparison of an internal and external ball. The next proposition provides justification to use internal and external balls interchangeably.

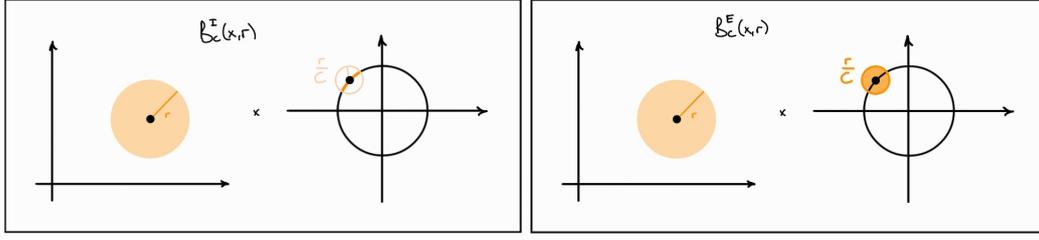


Figure 5.7: On the left, we sketch internal ball, and on the left we sketch the same ball, but external.

Proposition 5.7. *Let γ be a segment of some time series Γ . In particular, γ is a finite subset of \mathbb{R}^d . Moreover, let $C \geq 0$ and $r \in [0, C)$. Then, for*

$$\theta(x) = \sqrt{\left(1 - \left(1 - \frac{x^2}{2}\right)^2\right)} \text{ and } D_r: \text{eq} \frac{r}{C\theta\left(\frac{r}{C}\right)},$$

the map

$$\alpha_r: U_{D_r}^E(\rho(\gamma), r) \longrightarrow U_C^I(\rho(\gamma), r), \quad (p, v) \longmapsto \left(p, \frac{v}{\|v\|_2}\right)$$

is a homotopy equivalence. The collection $\{\alpha_r\}_{r \in [0, C)}$ is a map of filtered topological spaces.

Proof. The proof is constructive, and we will only state an overview of the steps. For full details, see Appendix F and G of [Bau+23b]. In the first part, we only concentrate on the tangent component.

Let $B^S(p, r) = \{q \in S^{d-1} : \|p - q\|_2 < r\} \subset B(p, r)$ be the internal open ball of p in $S^{d-1} \subset \mathbb{R}^d$, and consider the maps

$$\begin{aligned} \pi: \overline{B(p, \theta(r))} &\longrightarrow \overline{B^S(p, r)}, & v &\longmapsto \frac{v}{\|v\|_2} \\ \psi: \overline{B^S(p, r)} &\longrightarrow \overline{B(p, \theta(r))}, & w &\longmapsto \frac{2 - r^2}{2}w. \end{aligned}$$

The first step of the proof is to show that π and ψ are homotopy equivalences and that they are homotopy inverses of each other. In fact, $\pi \circ \psi$ is already the identity and $\psi \circ \pi$ can be homotoped to the identity using a straight-line homotopy.

Next, the maps π and ψ can be extended to unions of closed balls, such that they are still homotopy inverses of each other. Then, we note that

$$\alpha_r(p, v) = (p, \pi(v)),$$

and we define the new map

$$\beta_r: U_C^I(\rho(\gamma), r) \longrightarrow U_{D_r}^E(\rho(\gamma), r), \quad (p, v) \longmapsto (p, \psi(v)).$$

After showing that the maps α_r and β_r are homotopy inverses to each other, the proof is complete. \square

Corollary 5.8. *The intersection*

$$\bigcap_{x \in \gamma} \overline{B_{D_r}^E(x, r)}$$

is non-empty if and only if

$$\bigcap_{x \in \gamma} \overline{B_C^I(x, r)}$$

is. In particular, the corresponding nerve complexes are identical.

Proof. Let $x \in \gamma$. By construction of the maps α and $\beta = \{\beta_r\}_{r \in [0, C]}$, if a point p is inside $\overline{B_{D_r}^E(x, r)}$ then $\alpha(p)$ is in $\overline{B_C^I(x, r)}$. On the other hand, if $q \in \overline{B_C^I(x, r)}$, then $\beta(q)$ is in $\overline{B_{D_r}^E(x, r)}$. This property translates to intersections as well: if p is in the intersection

$$\bigcap_{x \in \gamma} \overline{B_{D_r}^E(x, r)},$$

it is in all of the closed balls, thus the image $\alpha_r(p)$ is in all balls $\overline{B_C^I(x, r)}$, so in particular, $\alpha_r(p)$ is in the intersection of all of them. The same argument goes for the map β_r . \square

Corollary 5.9. *Let γ be as before. There is a commutative diagram of $[0, C]$ -filtered topological spaces*

$$\begin{array}{ccc} U_C^I(\rho(\gamma)) & \xleftarrow{\alpha} & U_{D_\bullet}^E(\rho(\gamma)) \\ g \uparrow & & \uparrow f \\ |\check{C}^I(\rho(\gamma), d_C)| & \xrightarrow{h} & |\check{C}^E(\rho(\gamma), d_{D_\bullet})| \end{array},$$

Where the I and E parameter in the Čech complexes indicate whether the nerve complex is taken with respect to internal or external balls. For a fixed filtration value r , all maps are homotopy equivalences.

Proof. From the last corollary, we know that

$$\check{C}^I(\rho(\gamma), d_C) = \check{C}^E(\rho(\gamma), d_{D_\bullet}).$$

So, we can choose the same geometric representation for both and thus assume that h_r is the identity map for all r (so h is an identity map between filtered topological spaces). The nerve theorem (Theorem 3.6) provides us with the map f , which is a collection of homotopy equivalences. The map α is the one from Proposition 5.7, so it is also a collection of homotopy equivalences.

Thus, the collection g of maps

$$g_r = \alpha_r \circ f_r \circ h_r$$

is a collection of homotopy equivalences as well, and a morphism of filtered topological spaces. \square

This ends the discussion on the left arrow in the diagram 5.2.

5.4 Construction of Comparison Spaces

The next ingredient for the computation of cycling signatures we want elaborate on are the comparison spaces. The easiest way to construct a comparison space is by drawing a grid in dimension d over $X \subseteq \mathbb{R}^d$ and then deleting all the boxes the time series Γ does not touch. For this, we will make use of cubical complexes similar to those we introduced in Section 2.4. The reference for this section is Section 4.4 in [Bau+23b].

We recall two maps from Section 4.2.

$$\begin{aligned}\rho: X \setminus X_{fix} &\longrightarrow UTX, & x &\longmapsto \left(x, \frac{v(x)}{\|v(x)\|_2}\right) \\ \eta: UT\mathbb{R}^d &\longrightarrow UT_\infty\mathbb{R}^d, & (p, v) &\longmapsto \left(p, \frac{v}{\|v\|_\infty}\right)\end{aligned}$$

and introduce the new map

$$\vartheta_\infty: T\mathbb{R}^d \setminus (\mathbb{R}^d \times \{0\}) \longrightarrow UT_\infty\mathbb{R}^d, \quad (p, v) \longmapsto \left(p, \frac{v}{\|v\|_\infty}\right).$$

The following commutative diagram serves as an overview of how they are related.

$$\begin{array}{ccc} X \setminus X_{fix} & \xrightarrow{\rho} & UTX \\ p \mapsto (p, v) \downarrow & & \eta \downarrow \uparrow \eta^{-1} \\ TX \setminus (X \times \{0\}) & \xrightarrow{\vartheta_\infty} & UT_\infty X \end{array}$$

We will choose a neighbourhood U of $\eta(\rho(\Gamma))$ and then pull it back to a comparison space (as in Definition 5.3)

$$Y = \eta^{-1}(U).$$

This allows us to do the computation with boxes S_∞^{d-1} in the tangent component, and then come back to the standard unit tangent space introduced in section 4.2.

We proceed to the explicit construction of the neighbourhood U . To construct U , we consider products of cubes (Definition 2.30)

$$Q_{r,k}(p, q) = Q_r(p) \times Q_{1/k}(q) \subseteq T\mathbb{R}^d,$$

where $r > 0$ and $k \in \mathbb{N}$.

Using the sets $Q_{r,k}(p, q)$, we can construct the following cover of $UT_\infty\mathbb{R}^d$ (seen as a subset of $T\mathbb{R}^d$).

$$\mathcal{Q}_{r,k}^d = \left\{ Q_{r,k}(p, q) : p \in \mathbb{Z}^d, q \in \frac{1}{k}\mathbb{Z}^d, \|q\|_\infty = 1 \right\}$$

We define the cubical cover

$$Y_\infty = \{Q \in \mathcal{Q}_{r,k}^d : Q \cap \eta(U(\rho(\Gamma), \varepsilon)) \neq \emptyset\} \text{ for some } \varepsilon \geq 0. \quad (5.3)$$

So, we remove all boxes from the cover $\mathcal{Q}_{r,k}^d$ that have empty intersection with the thickening of Γ . Its geometric realisation is

$$|Y_\infty| = \bigcup_{\mathcal{Q}_{r,k}(p,q) \in Y_\infty} \mathcal{Q}_{r,k}(p,q) \subseteq TX \setminus (X \times \{0\}),$$

and we can finally define the neighbourhood

$$U: eq\vartheta_\infty(|Y_\infty|) \subseteq UT_\infty X \quad (5.4)$$

of $\eta(\rho(\Gamma))$.

We now formally introduce this second, more computable form of comparison spaces. An illustration to understand how the two kinds of comparison spaces are related can be found in Figure 5.8.

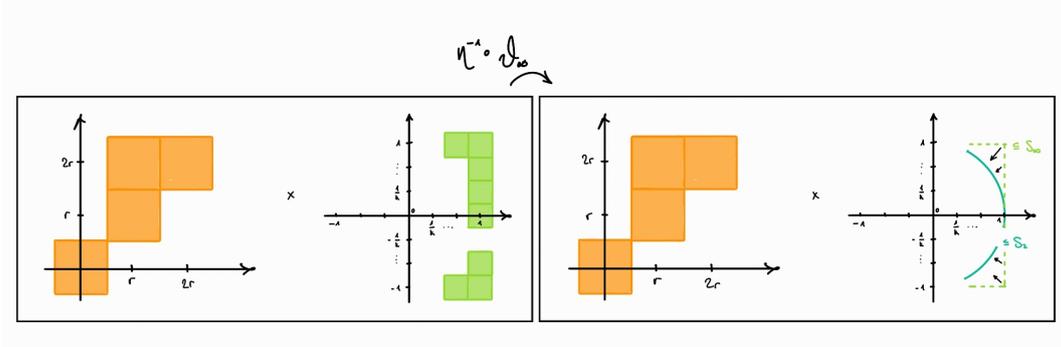


Figure 5.8: Illustration of the map $\eta^{-1} \circ \vartheta_\infty$ and how it relates cubical comparison spaces to comparison spaces as in Definition 5.3.

Definition 5.10. Let Γ be a time series in some ambient space $X \subseteq \mathbb{R}^d$. A **cubical comparison space** for Γ is a set $Y_\infty \subseteq \mathcal{Q}_{r,k}^d$ for some $r > 0$ and $k \in \mathbb{N}$, such that

$$Y = \eta^{-1}(\vartheta_\infty(|Y_\infty|))$$

is a comparison space (see Definition 5.3) for Γ .

5.5 Choice of the Parameters r_0 and C

Next, we address the question of how to choose r_0 , the maximal thickening radius for the computation of the Čech complexes and the parameter C for the definition of the metric d_C . What we do in the following differs a little from what is done in [Bau+23b], Section 5.5 and Proposition E.1.

Let $\Gamma \subset \mathbb{R}^d$ be a time series and $Y_\infty \subseteq \mathcal{Q}_{r,k}^d$ be a cubical comparison space. We recall that we defined the metric

$$d_C((p, v), (q, w)) = \max\{\|p - q\|_2, C\|v - w\|_2\}$$

on the unit tangent bundle UTX .

Proposition 5.11. *Let $r > 0$. Consider a collection $K \subseteq \mathcal{Q}_r^d$ of d -dimensional cubes with side length r and vertices on the grid $r\mathbb{Z}^d$. Moreover, consider the set*

$$U(K, s): \text{eqq}\{U(Q, s) : Q \in K\},$$

where the thickenings are taken with respect to Euclidean metric. For $s \in (0, r/2)$, the spaces $|U(K, s)| = \bigcup_{Q \in K} U(Q, s)$ and $|K| = \bigcup_{Q \in K} Q$ are homotopy equivalent.

Proof. In K , two boxes intersect if and only if they are adjacent. We need to check that there are no intersections added when passing to $U(K, s)$. So, we take two disjoint boxes in K and consider their thickenings in $U(K, s)$. The bound $s < r/2$ guarantees that they do not intersect in $U(K, s)$, either. This implies that the nerve complexes of the cover $U(K, s)$ of $|U(K, s)|$ and K of $|K|$ are isomorphic. Applying the nerve theorem for closed, convex sets (e.g. Theorem 3.1. in [Bau+23a]) yields that $|U(K, s)|$ and $|K|$ are homotopy equivalent. \square

Consider the set

$$U_{d_C}(Y_\infty, s): \text{eqq}\{U_{d_C}(Q, s) : Q \in Y_\infty\}$$

of thickenings of boxes in Y_∞ . The cubes Q are of the form $Q_r(p) \times Q_{1/k}(q)$, so in order to preserve intersections, the first component can only be thickened by less than $r/2$ and the second by less than $1/(2k)$ for Proposition 5.11 to apply. This implies the conditions $s < r/2$ and $Cs < 1/(2k)$ by definition of d_C . Setting $C = 1/(rk)$, and $r_0 = r/2$, the inequalities holds true for all $s < r_0$.

5.6 Acyclic Covers and Carriers

We have already introduced Čech complexes and cubical comparison spaces. In this section, we construct the cycling signature using the framework we established. So, we tackle the bottom arrow in the commutative diagram (5.2). This means that we want to construct a function from a simplicial to a cubical complex. Mapping any simplex to a cycle would be bad, so we will make sure this function maps simplices to connected subcomplexes. We follow [Col82],[MM18], [Hat01] and [Bau+23b].

As in the very first section, we want to make sure to avoid pathological cases. The following definition captures what this means in the context of this section.

Definition 5.12. *Let $\sigma \subset \mathbb{R}^d$ be a convex, compact set. If $\sigma \cap Q^\circ \neq \emptyset$ for all $Q \in K^d(\sigma)$, then σ is said to be **in general position** with respect to its outer cubical cover.*

An illustration with a non-example and an example can be found in Figure 5.9.

Remark 5.13. *Although we use the same word, this definition is somewhat different from Definition 2.3. In this thesis, it should be clear from context when we mean which definition. In both cases, it is a definition we make to avoid any strange cases.*

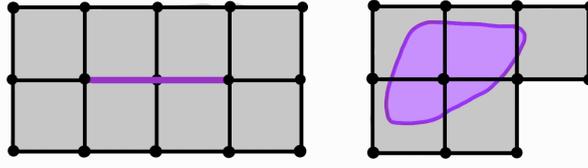


Figure 5.9: On the left side, we sketched a convex, compact set which is not in general position with respect to its outer cubical cover. On the right side, we see a convex, compact set that is.

We show that we can thicken convex, compact sets a little bit, while preserving the same outer cubical cover. We recall the definition, Definition 3.5, of thickenings.

Lemma 5.14. *Let $\sigma \subset \mathbb{R}^d$ be a convex, compact set. There exists a $\varepsilon_0 > 0$ such that for all $\varepsilon \in [0, \varepsilon_0)$,*

$$K(\sigma) = K(U(\sigma, \varepsilon)).$$

Proof. Let $L = \downarrow (K(\mathbb{R}^d) \setminus K(\sigma))$ be the cubical complex containing all boxes of all dimensions that do not touch σ . By construction, this implies that the Euclidean distance $d(Q, \sigma)$ is positive for all $Q \in L$. Moreover, for any $C > 0$, there exist only finitely many boxes $Q \in L$ for which $d(Q, \sigma) < C$. Thus, the minimum

$$\varepsilon_0 = \min_{Q \in L} d(Q, \sigma)$$

is well-defined.

Let $\varepsilon < \varepsilon_0$. It follows from the triangle inequality, that $d(Q, U(\sigma, \varepsilon)) > 0$. Thus, $K(\sigma) = K(U(\sigma, \varepsilon))$. \square

Lemma 5.15. *Let $\sigma \subset \mathbb{R}^d$ be a convex, compact set. There exists an $\varepsilon > 0$ for which $K(\sigma) = K(U(\sigma, \varepsilon))$, and $U(\sigma, \varepsilon)$ is in general position.*

Proof. Using the last lemma, we may fix an $\varepsilon > 0$ for which $K(\sigma) = K(U(\sigma, \varepsilon))$. Let $Q \in K(\sigma)$ be an arbitrary cube, and pick some $x \in \sigma \cap Q$. By construction, $B(x, \varepsilon) \cap Q^\circ \neq \emptyset$, and $B(x, \varepsilon) \cap Q^\circ \subseteq U(\sigma, \varepsilon) \cap Q^\circ$. Hence, $U(\sigma, \varepsilon)$ is in general position. \square

Lemma 5.16. *Let $\sigma \subset \mathbb{R}^d$ be a convex, compact set. Then $K(\sigma)$ is acyclic.*

Proof. By the preceding lemma, we may assume that σ is in general position with respect to its outer cubical cover $K(\sigma)$. We proceed by induction on the number n of cubes in $K(\sigma)$ which have maximal dimension.

If we have a single cube of top dimension, it is clearly contractible, thus acyclic. We now assume the claim holds for some $n > 1$.

We now assume that $K(\sigma)$ contains $n + 1$ cubes of top dimension d_{\max} . Since by assumption, $n > 1$, there exist at least two "adjacent" cubes, i.e. there exists a positive

integer i , such that the projection of $|K(\sigma)| = \bigcup_{Q \in K(\sigma)} Q$ to the i -th coordinate satisfies

$$\pi_i(|K(\sigma)|) = [m_1, m_2]$$

for some integers m_1, m_2 which satisfy $m_2 - m_1 \geq 2$. Set $l = m_2 - 1$.

Consider the hyperplane $H = \pi_i^{-1}(l)$, and define the two sets

$$\begin{aligned} A &= \downarrow \{Q \in K_{d_{\max}}(\sigma) : \pi_i(Q) \subseteq (-\infty, l]\} \\ B &= \downarrow \{Q \in K_{d_{\max}}(\sigma) : \pi_i(Q) \subseteq [l, +\infty)\}. \end{aligned}$$

We write $|A| = \bigcup_{Q \in A} Q$ and $|B| = \bigcup_{Q \in B} Q$ to be the geometric realisations of A and B and will use the Mayer-Vietoris theorem (see [Hat01], Mayer-Vietoris Sequences in Chapter 2). We cannot directly apply the standard Mayer-Vietoris theorem, as the interiors of $|A|$ and $|B|$ do not cover all of $|K_{d_{\max}}(\sigma)|$. In fact, the hyperplane H is not contained in $|A|^\circ \cup |B|^\circ$.

Instead, we construct open neighbourhoods $U_A = U(|A|, \varepsilon)$, $U_B = U(|B|, \varepsilon)$ for some $\varepsilon > 0$ small enough. Then, $|A|$ is a deformation retract of U_A , $|B|$ is a deformation retract of U_B and moreover, $H \cap |K(\sigma)| = |A| \cap |B|$ is a deformation retract of $U_A \cap U_B$. A sketch to gain some intuition for this construction can be found in Figure 5.10 on the left.

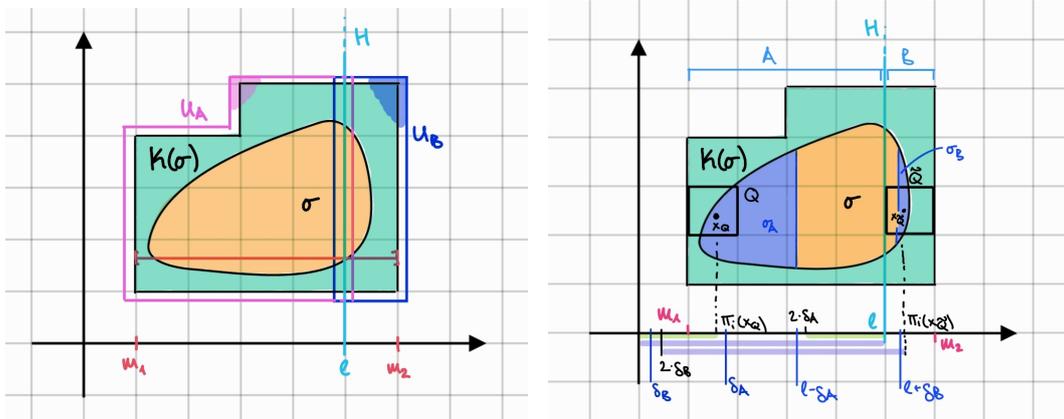


Figure 5.10: Sketches of the constructions for the proof of Lemma 5.16 in two dimensions for some convex compact set σ .

Since the interiors of U_A and U_B now clearly cover $|K_{d_{\max}}(\sigma)|$, we may apply an adapted version of the Mayer-Vietoris theorem (also explained in [Hat01], same chapter) to get the short exact sequence of chain complexes

$$0 \longrightarrow C_\bullet(A \cap B) \longrightarrow C_\bullet(A) \oplus C_\bullet(B) \longrightarrow C_\bullet(A \cup B) = C_\bullet(K_{d_{\max}}(\sigma)) \longrightarrow 0,$$

and a long exact sequence

$$\cdots \longrightarrow H_k(A \cap B) \longrightarrow H_k(A) \oplus H_k(B) \longrightarrow H_k(A \cup B) \longrightarrow H_{k-1}(A \cap B) \longrightarrow \cdots$$

in homology. We will show that $H_k(A \cap B) = H_k(A) = H_k(B) = 0$ for all k .

For any cube $Q \in K_{d_{\max}}(\sigma)$, we can fix a point $x_Q \in \sigma \cap Q^\circ$, since σ is in general position. We define the two constants

$$\begin{aligned}\delta_A &= \frac{1}{2} \max\{l - \pi_i(x_Q) : Q \in A\} \\ \delta_B &= \frac{1}{2} \min\{\pi_i(x_Q) - l : Q \in B\}.\end{aligned}$$

Since x_Q lies in the interior of some cube of dimension d_{\max} , we know that $\pi_i(x_Q) \notin \mathbb{Z}$. So, δ_A and δ_B are strictly positive.

We now set

$$\begin{aligned}\sigma_A &= \sigma \cap \pi_i^{-1}((-\infty, l - \delta_A]) \\ \sigma_B &= \sigma \cap \pi_i^{-1}([l + \delta_B, +\infty))\end{aligned}$$

An attempt to sketch how these sets are constructed can be found in Figure 5.10 on the right. We show that $A = K(\sigma_A)$ and $B = K(\sigma_B)$, thus showing that A and B are outer cubical complexes of compact, convex sets.

Let $Q \in A$. Since $x_Q \in \sigma \cap Q^\circ$, we can deduce $\pi_i(x_Q) < l - \delta_A$, thus $x_Q \in \sigma_A$. By construction, there is no other cube than Q for which $x_Q \in Q^\circ$, so $Q \in K(\sigma_A)$.

If $Q \in K(\sigma_A)$, then $Q \in K(\sigma)$ as well by definition. Moreover, $\sigma_A \in \pi_i^{-1}((-\infty, l - \delta_A])$, so in particular, $Q \subseteq \pi_i^{-1}((-\infty, l])$, which means $Q \in A$. Finally, this implies $K(\sigma_A) \subseteq A$ and therefore $A = K(\sigma_A)$.

We have shown that A and B are outer cubical covers of the convex, compact sets σ_A and σ_B . By construction, they both contain less than $n + 1$ d_{\max} -dimensional cubes, so A and B are acyclic (i.e. $H_k(A) = H_k(B) = 0$ for all k).

We are left with showing that $A \cap B$ is acyclic as well. The set $\sigma_H = \sigma \cap H$ is compact and convex. Moreover, it is fully contained in $A \cap B$. Let $\pi_{\hat{i}}$ be the projection deleting the i -th coordinate. $\pi_{\hat{i}}(\sigma_H)$ is compact and convex, and admits a cover $C = \{\pi_{\hat{i}}(Q) : Q \in A \cap B\}$. We show that $C = K(\pi_{\hat{i}}(\sigma_H))$. Let $Q \in K(C)$. The cube Q corresponds to a unique $(d_{\max} - 1)$ -dimensional cube P of side length 1 in \mathbb{R}^d which is contained in $A \cap B$. In particular, there are cubes $Q_A \in A$ and $Q_B \in B$, which have P as a face in common. We pick points $x_A \in \sigma \cap Q_A$ and $x_B \in \sigma \cap Q_B$. The line connecting x_A and x_B intersects P in a point x . For this point, $\pi_{\hat{i}}(x) \in Q$, so $C = K(\pi_{\hat{i}}(\sigma_H))$.

Finally, since a $(d_{\max} - 1)$ -dimensional cube of $K(C)$ is always adjacent to a d_{\max} -dimensional cube in both A and B , we know that $2\#K_{d_{\max}-1}(C) \leq n$. Thus, C is acyclic by induction. Since C is homeomorphic to $A \cap B$ by construction, $A \cap B$ is acyclic as well.

The proposition follows using exactness of the Mayer-Vietoris long exact sequence in homology. \square

We claim that the same property holds after scaling and turning the grid.

Corollary 5.17. *Let $A \in GL_d(\mathbb{R})$ be an invertible linear map, and let $\sigma \in \mathbb{R}^d$ be convex and compact. Then $AK(\sigma)$ is acyclic.*

Proof. Applying Lemma 5.16 and using the fact that A preserves compactness, convexity and homology yields the result. \square

The next corollary translates what we just established for outer cubical complexes (as introduced in Section 2.4) to cubical comparison spaces.

Corollary 5.18. *Let σ be a compact and convex subset of $UT_\infty \mathbb{R}^d$. For any $r \in \mathbb{R}$ and $k \in \mathbb{Z}$, the set*

$$L = \downarrow \{Q \in \mathcal{Q}_{r,k}^d : Q \cap \sigma \neq \emptyset\}$$

is acyclic.

Proof. Let $v \in \mathbb{R}^{2d}$ be the vector in with $v_i = r$ for $i = 1, \dots, d$ and $v_i = 1/k$ for $i = d+1, \dots, 2d$. Let A be the diagonal matrix with diagonal v . Moreover, consider the set $\tau: eqq\{w - v/2 : w \in \sigma\} \subset UT_\infty \mathbb{R}^d$. This shift ensures that the vertices of the cubical cover lie on the integer grid, since for cubical comparison spaces, we defined the centre of the cubes to be on the integer grid (see Section 5.4). By the previous corollary, $AK(\tau)$ is acyclic. We show that $|L|$ and $|AK(\tau)|$ are the identical up to a translation by $v/2$.

Let $Q \in AK(\tau)$ and define $\tilde{Q} = \{w + v/2 : w \in Q\}$. By construction, $\tilde{Q} \cap \sigma \neq \emptyset$. We show that $\tilde{Q} \in \mathcal{Q}_{r,k}^d$. Clearly, $\tilde{Q} = Q_r(p) \times Q_{\frac{1}{k}}(q)$ for some $p \in r\mathbb{Z}^d$ and $q \in \frac{1}{k}\mathbb{Z}^d$. So, we need to show $\|q\|_\infty = 1$. Whenever $\|q\|_\infty \neq 1$, $Q_{\frac{1}{k}}(q) \cap S_\infty^1 = \emptyset$. The explanation is that the box $Q_{\frac{1}{k}}(q)$ has length $1/k$ and center q , thus

$$q - \frac{1}{2k} \leq \|x\|_\infty \leq q + \frac{1}{2k} \quad \text{for all } x \in Q_{1/k}(q).$$

In particular, $\|x\|_\infty \neq 1$ if $q \in 1/k\mathbb{Z}^d$ is such that $\|q\|_\infty \neq 1$. Translation is a homeomorphism, so $|L|$ is acyclic. \square

We are going to use these results to construct a function that maps simplices to sets of the form L as in Corollary 5.18. Using the upcoming theorem, this will provide us with the map ϕ appearing in the bottom of the diagram (5.2). To understand the theorem, we introduce acyclic carriers from simplicial to cubical complexes.

Definition 5.19 (Acyclic Carrier). *Let \mathcal{K} be a simplicial complex and K a cubical complex. Consider a map*

$$F: \Sigma(\mathcal{K}) \longrightarrow \{\text{Non-empty subcomplexes of } K\}.$$

- (a) *The function F is a **carrier** from \mathcal{K} to K , if for any simplex $\sigma \in \Sigma(\mathcal{K})$ and σ' a face of σ , $F(\sigma')$ is a subcomplex of $F(\sigma)$.*
- (b) *Moreover, F is an **acyclic carrier** from \mathcal{K} to K , if F is a carrier and $\tilde{H}_k(F(\sigma)) = 0$ for all $k \geq 0$, where \tilde{H}_k means reduced homology.*

(c) A map $f: \mathcal{K} \rightarrow K$, which sends simplices to cubes, is said to be **carried** by F , if $f(\sigma)$ is a cube of $F(\sigma)$ for all simplices σ of \mathcal{K} .

When passing to chain complexes, a carrier F from a simplicial complex \mathcal{K} to a cubical complex K induces a function \mathcal{F} which sends basis elements σ of the vector spaces $C_k(\mathcal{K})$ to sub-chain complexes of $C_\bullet(K)$. These induced maps are so-called **algebraic carriers**. In particular, acyclic carriers induce acyclic algebraic carriers, which means the reduced homology of all chain complexes in $\text{im}(\mathcal{F})$ is trivial. Algebraic carriers can carry homomorphisms

$$\varphi: C_k(\mathcal{K}) \rightarrow C_l(K),$$

which means that $\varphi(\sigma)$ is contained in one of the chain vector spaces of $\mathcal{F}(\sigma) \subseteq C_\bullet(K)$. In fact, the next theorem tells us that if \mathcal{F} is acyclic, there always exists a chain map carried by \mathcal{F} . This is the map we will use for the bottom arrow in the commutative diagram (5.2).

For an introduction to reduced homology and more details on acyclic carriers, we refer to [MM18], Sections 7 and 13.

Theorem 5.20 (Acyclic Carrier Theorem). *Let \mathcal{K} be a simplicial complex, K a cubical complex and F an acyclic carrier from \mathcal{K} to K . Moreover, let $(C_\bullet(\mathcal{K}), \varepsilon_{\mathcal{K}})$ and $(C_\bullet(K), \varepsilon_K)$ be the corresponding augmented chain complexes and let \mathcal{F} be the algebraic carrier induced by F . Then there is an augmentation-preserving chain map*

$$\phi: C_\bullet(\mathcal{K}) \rightarrow C_\bullet(K)$$

carried by \mathcal{F} . Moreover, any two such maps are homotopic, with the homotopy being carried by \mathcal{F} as well.

For a proof of this theorem, consult [MM18], Theorem 13.4. As before, let $\Gamma \subset \mathbb{R}^d$ be a time series, γ a segment of Γ and let Y_∞ be a cubical comparison space for Γ . Moreover, let $R \in [0, r_0)$ be the largest critical point of $\{\check{C}_r(\rho(\gamma))\}_{r \in [0, r_0)}$, i.e. the largest radius where a change happens in homology. Our goal is to apply the acyclic carrier theorem (Theorem 5.20) to obtain a chain map between the chain complex $C_\bullet(\check{C}_R(\rho(\gamma)))$ and the chain complex $C_\bullet(Y_\infty)$ of the cubical comparison space. This chain map then induces the map in homology that we can ultimately use for the computation of cycling signatures in real life.

We recall the two maps of filtered topological spaces (for $r \in [0, C)$).

$$\begin{aligned} \eta: UT\mathbb{R}^d &\rightarrow UT_\infty\mathbb{R}^d, & (p, v) &\mapsto \left(p, \frac{v}{\|v\|_\infty}\right) \\ \alpha: U_{D_\bullet}^E(\rho(\gamma)) &\rightarrow U_C^I(\rho(\gamma)), & (p, v) &\mapsto \left(p, \frac{v}{\|v\|_2}\right) \\ f: |\check{C}^E(\gamma, d_{D_\bullet})| &\rightarrow U_{D_\bullet}^E(\rho(\gamma)) & &\text{the map constructed in (3.1).} \end{aligned}$$

We define a new map of filtered topological spaces

$$\Theta: |\text{sd}\check{C}(\rho(\gamma))| \rightarrow Y_\infty, \quad \Theta: eqq\eta \circ \alpha \circ f.$$

The following commutative diagram makes this detour a bit more understandable.

$$\begin{array}{ccc}
|\text{sd}\check{C}(\rho(\gamma))| & \xrightarrow{\Theta} & Y \subseteq UT_\infty\mathbb{R}^d \\
f \downarrow & & \uparrow \eta \\
U_{D_\bullet}^E(\rho(\gamma)) & \xrightarrow{\alpha} & U_C^I(\rho(\gamma))
\end{array}$$

Proposition 5.21. *There exists a simplicial subdivision \mathcal{K}_R of $\check{C}_R(\rho(\gamma))$ such that the map*

$$F: \mathcal{K}_r \longrightarrow Y_\infty, \quad \sigma \longmapsto F(\sigma) = \downarrow \{Q \in Y_\infty : Q \cap \Theta_R(\sigma) \neq \emptyset\}$$

is an acyclic carrier.

Proof. We divide the proof into two steps.

First, we show that there exists a subdivision \mathcal{K}_R of $\check{C}_R(\rho(\gamma))$ such that for any face H of the cube $\overline{B_\infty(0,1)} \subset \mathbb{R}^d$, the set $\Theta_R^{-1}(\mathbb{R}^d \times H)$ is a subcomplex of \mathcal{K}_R . By definition, the set $\overline{U_{D_R}^E(\rho(\gamma), R)}$ is compact, and it does not intersect $\mathbb{R}^d \times \{0\} \subset T\mathbb{R}^d$. This is the case because with respect to the metric d_C , the distance between $\mathbb{R}^d \times \{0\}$ and $UT\mathbb{R}^d$ is the scaled distance in the tangent component, thus exactly C . By assumption, $R < C$. This means that there are radii $r_1, r_2, r_3 > 0$ such that

$$\overline{U_{D_R}^E(\rho(\gamma), R)} \subseteq \overline{B_\infty^d(0, r_1)} \times \left(\overline{B_\infty^d(0, r_2)} \setminus B_\infty^d(0, r_3) \right) =: A \subseteq T\mathbb{R}^d.$$

We show that A is the geometric realisation of a simplicial complex \mathcal{L} . To do so, we decompose A into subsets

$$A_H = A \cap \left(\mathbb{R}^d \times \{rH : r > 0\} \right),$$

where H is a face of $B_\infty^d(0, 1)$. A sketch of how this might look in dimension $d = 2$ can be found in Figure 5.11. By construction, the sets A_H are convex. The intersection of

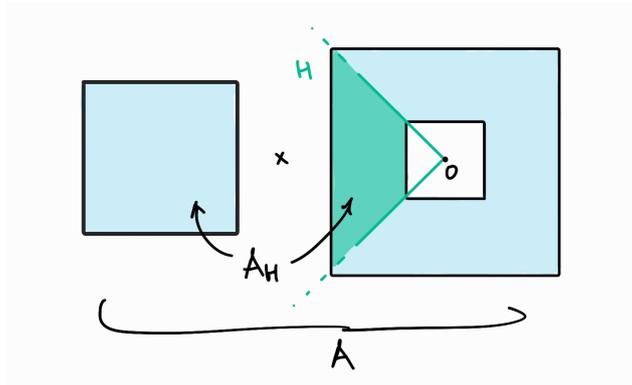


Figure 5.11: Visualisation of a set of the form A_H in dimension $d = 2$ from the proof of Proposition 5.21.

two distinct such subsets is either empty, or a face of both. In fact, A is the geometric

realisation of a cubical complex, which can be subdivided into a simplicial complex \mathcal{L} without adding any vertices by Proposition 2.34. By construction,

$$f_R: |\text{sd}\check{C}_R(\rho(\gamma))| \longrightarrow \overline{U_{D_R}^E(\rho(\gamma), R)} \subseteq |\mathcal{L}|$$

is linear on every simplex of its domain. By Lemma 2.13, there are simplicial subdivisions \mathcal{K}_R of $\check{C}_R(\rho(\gamma))$ and \mathcal{L}' of \mathcal{L} such that

$$f_R: |\mathcal{K}_R| \longrightarrow |\mathcal{L}'|$$

is simplicial. We define the filtered simplicial complex \mathcal{K} as follows.

$$\mathcal{K} = \{\mathcal{K}_r\}_{r < R}, \quad \mathcal{K}_r = \{\sigma \in \mathcal{K}_R : |\sigma| \subseteq |\check{C}_r(\rho(\gamma))|\}.$$

The second step of the proof is to check that F satisfies Definition 5.19. The fact that $F(s) \subseteq F(\sigma)$ for s a face of a simplex σ is true by construction. Let σ be any simplex of \mathcal{K}_R , and write $\tau = \Theta_R(\sigma)$. By construction of the subdivision \mathcal{L}' , τ is contained in $\mathbb{R}^d \times H$ for some face H of the cube $B_\infty^d(0, 1)$. Therefore,

$$\Theta_R \Big|_\sigma = (f_R \circ \eta \circ \alpha_R) \Big|_\sigma.$$

In the tangent component, this is identical to the radial projection onto H . All maps in the above composition preserve compactness and convexity, so τ is compact and convex as well. By Corollary 5.18, $F(\sigma)$ is acyclic. \square

Corollary 5.22. *Let \mathcal{K} be the subdivision of $\text{sd}\check{C}(\rho(\gamma))$ from Proposition 5.21, and let ψ_1 and ψ_2 be the chain maps which induce the isomorphism between singular and cellular homology (for details on these maps, see [Hat01] Chapter 2, Cellular Homology). There is a commutative diagram*

$$\begin{array}{ccc} H_1(|\text{sd}\check{C}(\rho(\gamma))|) & \xrightarrow{H_1(\Theta)} & H_1(|Y_\infty|) \\ \uparrow H_1(\psi_1) & & H_1(\psi_2) \downarrow \\ H_1(\mathcal{K}) & \xrightarrow{H_1(\nu)} & H_1(Y_\infty) \end{array}$$

Proof. We only sketch the proof, as an in-depth approach would be very extensive to add. For details, see [Bau+23b], Corollary 4.15.

Consider the subdivision $\{\mathcal{K}_R\}_{R < C}$ from Proposition 2.34. We define the algebraic acyclic carrier $\mathcal{F}: C_\bullet(\mathcal{K})_R \longrightarrow C_\bullet\{Y_\infty\}$ by

$$\mathcal{F}(\sigma) = C_\bullet(F(\sigma)).$$

Then, we apply the acyclic carrier theorem, (Thm. 5.20) to get a chain map

$$\phi_R: C_\bullet(\mathcal{K}_R) \longrightarrow C_\bullet(Y_\infty)$$

carried by \mathcal{F} . We translate Θ_R to the chain map $C_\bullet(\Theta_R)$ and show that $\psi_2 \circ C_\bullet(\Theta_R) \circ \psi_1$ is carried by \mathcal{F} as well. This is done by showing that $\psi_2 \circ C_\bullet(\Theta_R) \circ \psi_1$ is chain homotopic to ϕ_R , and that the chain homotopy respects the acyclic carrier \mathcal{F} . After applying restrictions, this yields a filtered chain map and proves commutativity of the diagram. \square

This final corollary concludes the construction of the (computable) cycling signature. We will see in the next section that the cycling signature can now be computed comfortably, using a result (Proposition 5.24) from linear algebra.

Corollary 5.23. *Let $Y = \eta^{-1}(\vartheta_\infty(|Y_\infty|))$ be a comparison space. There exists a commutative diagram*

$$\begin{array}{ccc} H_1(U(\rho(\gamma))) & \xrightarrow{H_1(i_\gamma)} & H_1(Y) \\ \uparrow \cong & & \cong \downarrow \\ H_1(\check{C}(\rho(\gamma))) & \xrightarrow{H_1(\phi)} & H_1(Y_\infty) \end{array}$$

Proof. We have seen in this chapter that any of the spaces in this diagram is isomorphic to the corresponding space in the previous commutative diagram. \square

5.7 Summarised Computation Pipeline

In this section, we briefly summarise the computational pipeline used to obtain cycling signatures. The code by Bauer et al. used in [Bau+23b] can be accessed using the following link.

<https://github.com/davidhien/CyclingSignatures.jl>

The first step of the algorithm is to construct the cubical comparison space and compute its homology. As this only depends on the given time series Γ and not on the segments we choose to analyse later, it makes sense to do this computation in advance. We collect all boxes $Q_{r,k}$ which contain a point from the time series Γ . This is done in both components of the tangent space separately. A basis $\{\alpha_i\}_{i=1}^m$ of the resulting cubical comparison space Y_∞ is computed using a reduction algorithm similar to the one we introduced for simplicial complexes, as mentioned in Section 2.4. A visual example of a comparison space can be found in Figure 5.12.

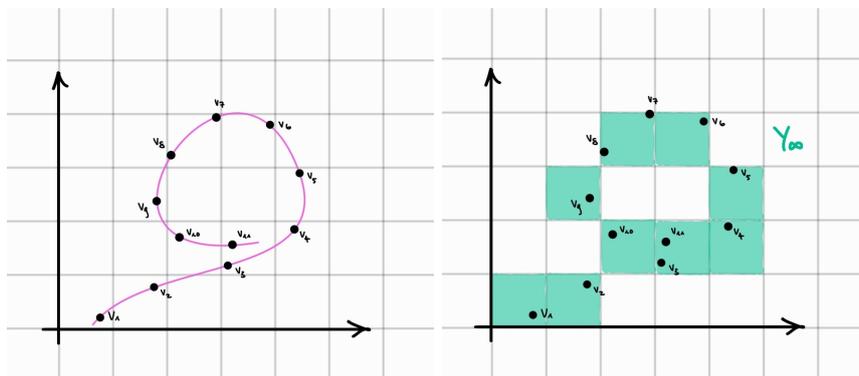


Figure 5.12: On the left, we see a very short time series $\{v_1, \dots, v_{11}\}$, with the underlying orbit drawn in pink. On the right, we illustrate a cubical comparison space for this time series.

Next we consider a collection of segments $\gamma \subseteq \Gamma$. More exactly, we fix a number N of segments and a choice \mathcal{T} of segment lengths and randomly sample N segments of length τ for every $\tau \in \mathcal{T}$.

Although we only ever mentioned Čech complexes in the context of cycling signatures, the persistent homology computations with this algorithm are actually done with Vietoris-Rips complexes. So, we compute persistent homology of persistent Vietoris-Rips complexes using the machinery introduced in Section 3.3. The result is a set of generators $\{c_i\}_{i \in I}$ of the persistent homology group $H_1(\text{VR}(\rho(\gamma)))$. We single out the generators $\{c_1^\infty, \dots, c_n^\infty\}$ which form the basis of the immortal part $H_1^\infty(\text{VR}(\rho(\gamma)))$ given by

$$H_1^\infty(\text{VR}(\rho(\gamma)))_r = \{v \in H_1(\text{VR}(\rho(\gamma)))_r : v = 0 \text{ or } H_1(\varphi_r^{r_0})(v) \neq 0\},$$

where $\varphi_r^{r_0}$ is the inclusion map $\text{VR}(\rho(\gamma), r) \subseteq \text{VR}(\rho(\gamma), r_0)$. We then use those generators as an approximation for the basis of the immortal part $H_1^\infty(\check{C}(\rho(\gamma)))$ defined analogously. They correspond to 1-dimensional topological features that have not disappeared at the maximal radius r_0 .

Next, we show how to compute the map ϕ introduced in the proof of Corollary 5.22. Let c be a chain in $C_1(\check{C}(\rho(\gamma), r))$ for some radius $r \in [0, r_0]$. Then c is a finite formal sum

$$c = \sum_k \lambda_k e_k, \quad \lambda_k \in K \text{ and } e_k \in \Sigma_1(\check{C}(\rho(\gamma), r)) \text{ for all } k.$$

We compute the map ϕ on the edges e_k , the definition for chains then follows by linearity. We first subdivide the edges such that two consecutive vertices in the subdivision lie in adjacent boxes of Y_∞ , and such that all segments lie on the same face of $\partial B_\infty(0, 1)$ in the tangent component.

Instead of an edge, we now have a chain of edges with boundary ∂e_k . We map every vertex v of this chain to a vertex $\phi_r(v)$ that is a face of the cube $Q \in Y_\infty$ for which $v \in Q$. Using this, we map every edge $e = [v, w]$ from the subdivision to a chain of edges in Y_∞ that are faces of the boxes Q containing the vertices, and such that

$$\partial \phi_r(e) = \phi_r(\partial e) = \phi_r(v + w) = \phi_r(v) + \phi_r(w).$$

The image $\phi_r(e_k)$ is then given by the sum of those chains, thus has boundary $\partial \phi_r(e_k) = \phi_r(\partial e_k)$ by construction. A sketch (tangent component neglected) to understand this construction better is provided in Figure 5.13. Doing this for all r gives the map $\phi: \check{C}(\rho(\gamma)) \rightarrow Y_\infty$.

The last step in the algorithm is the computation of a basis for the cycling signature, so for $\text{im}(H_1(\phi))$. To do this, we rewrite the basis elements $\alpha_1, \dots, \alpha_m$ of $H_1(Y_\infty)$ and the images $\phi(c_1^\infty), \dots, \phi(c_n^\infty)$ of the generators of $H_1^\infty(\text{VR}(\rho(\gamma)))$ as vectors in \mathbb{F}_2^N , where N is the number of edges contained in Y_∞ . We set an entry of such a vector to be 1 if the edge corresponding to the entry is contained in the chain, and 0 else. We then define the matrix A with entries $a_{i,j} = \alpha_i^T \phi(c_j)$ and apply Proposition 5.24 to obtain a basis for the cycling signature.

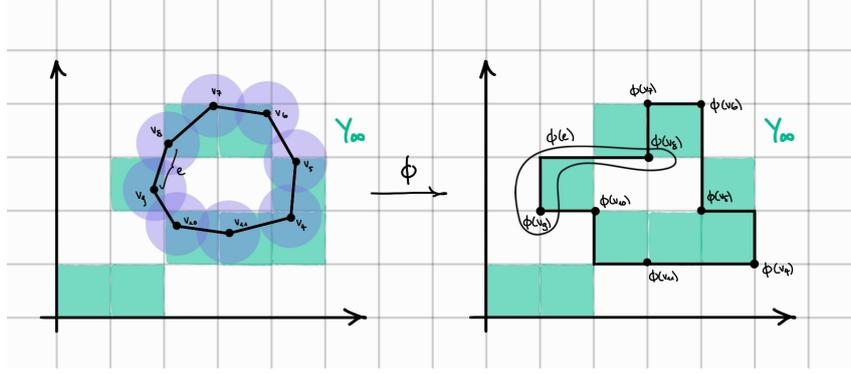


Figure 5.13: Intuition for how the map ϕ works. The function is applied to a Vietoris-Rips complex constructed on a segment of the time series from Figure 5.12 and the radii are represented by the purple discs.

Proposition 5.24. *Let V and W be persistent vector spaces with a finite index set $\{0, \dots, R\}$ and let $\phi: V \rightarrow W$ be a linear transformation. Moreover, let $B_R^V = \{v_1, \dots, v_n\}$ and $B_R^{W^*} = \{\alpha_1, \dots, \alpha_m\}$ be bases of V_R and the dual space W_R^* , respectively, such that they are listed in increasing order with respect to their birth radius. We define the matrix A with entries*

$$a_{i,j} = \alpha_i(\phi_R(v_j)).$$

We consider the matrix $A' = AU$ that results from reducing A . The set

$$\{\phi_i(v_j) : j \leq i \text{ and } \text{pivot}(R_j) \neq 0\}$$

is a basis for $\text{im}(\phi)$.

A proof of this proposition can be found in [Bau+23b], Proposition 4.21.

Bibliography

- [Ada20] Henry Adams. *What is the difference between Vietoris-Rips and Cech complexes?* Accessed: 02/10/2024. 2020. URL: <https://youtu.be/YLXA F8IB9Ng?feature=shared>.
- [Bau+23a] Ulrich Bauer et al. “A unified view on the functorial nerve theorem and its variations”. In: *Expositiones Mathematicae* 41.4 (Dec. 2023), p. 125503. ISSN: 0723-0869. DOI: [10.1016/j.exmath.2023.04.005](https://doi.org/10.1016/j.exmath.2023.04.005). URL: <http://dx.doi.org/10.1016/j.exmath.2023.04.005>.
- [Bau+23b] Ulrich Bauer et al. *Cycling Signatures: Identifying Oscillations from Time Series using Algebraic Topology*. 2023. arXiv: [2312.04734](https://arxiv.org/abs/2312.04734) [math.DS]. URL: <https://arxiv.org/abs/2312.04734>.
- [BD16] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting*. Springer, 2016.
- [BS02] Michael Brin and Garrett Stuck. *Introduction to Dynamical Systems*. Cambridge University Press, 2002.
- [Car14] Gunnar Carlsson. “Topological pattern recognition for point cloud data”. In: *Acta Numerica* 23 (2014), pp. 289–368. DOI: [10.1017/S0962492914000051](https://doi.org/10.1017/S0962492914000051).
- [Col82] Brian J. Sanderson Colin P. Rourke. *Introduction to Piecewise-Linear Topology*. Springer Berlin, Heidelberg, 1982.
- [EH10] Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction*. Jan. 2010. ISBN: 978-0-8218-4925-5. DOI: [10.1007/978-3-540-33259-6_7](https://doi.org/10.1007/978-3-540-33259-6_7).
- [Hat01] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2001.
- [Hie23] David Hien. *CyclingSignatures.jl*. Accessed 10/03/2025. 2023. URL: <https://github.com/davidhien/CyclingSignatures.jl.git>.
- [JLR20] Elena Jahr, Gabriel Laude, and Jeremy O. Richardson. “Instanton theory of tunneling in molecules with asymmetric isotopic substitutions”. In: *The Journal of Chemical Physics* 153.9 (Sept. 2020). ISSN: 1089-7690. DOI: [10.1063/5.0021831](https://doi.org/10.1063/5.0021831). URL: <http://dx.doi.org/10.1063/5.0021831>.
- [KMM04] Tomasz Kaczynski, Konstantin Mischaikow, and Marian Mrozek. “Cubical Homology”. In: *Computational Homology*. New York, NY: Springer New York, 2004, pp. 39–92. ISBN: 978-0-387-21597-6. DOI: [10.1007/0-387-21597-2_2](https://doi.org/10.1007/0-387-21597-2_2). URL: https://doi.org/10.1007/0-387-21597-2_2.
- [Lee12] John M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer New York, NY, 2012. DOI: <https://doi.org/10.1007/978-1-4419-9982-5>.

- [Mel19] Alexandre Melanson. *Effective stochastic models of neuroscientific data with application to weakly electric fish*. 2019. DOI: <http://dx.doi.org/10.20381/ruor-23339>.
- [MM18] J.R. Munkres and J.W. Munkres. *Elements Of Algebraic Topology*. CRC Press, 2018. ISBN: 9780429962462. URL: <https://books.google.ch/books?id=-mdQDwAAQBAJ>.
- [RC19] Nalini Ravishanker and Renjie Chen. *Topological Data Analysis (TDA) for Time Series*. 2019. arXiv: [1909.10604](https://arxiv.org/abs/1909.10604) [stat.AP]. URL: <https://arxiv.org/abs/1909.10604>.
- [Tak81] Floris Takens. “Detecting strange attractors in turbulence”. In: *Dynamical Systems and Turbulence, Warwick 1980*. Ed. by David Rand and Lai-Sang Young. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, pp. 366–381. ISBN: 978-3-540-38945-3.
- [Wof18] Eric Wofsey. *Barycentric subdivision preserves geometric realization*. Mathematics Stack Exchange. Accessed: 05/03/2025. 2018. eprint: <https://math.stackexchange.com/q/3031715>. URL: <https://math.stackexchange.com/q/3031715>.



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every written paper or thesis authored during the course of studies. In consultation with the supervisor, one of the following three options must be selected:

I confirm that I authored the work in question independently and in my own words, i.e. that no one helped me to author it. Suggestions from the supervisor regarding language and content are excepted. I used no generative artificial intelligence technologies¹.

I confirm that I authored the work in question independently and in my own words, i.e. that no one helped me to author it. Suggestions from the supervisor regarding language and content are excepted. I used and cited generative artificial intelligence technologies².

I confirm that I authored the work in question independently and in my own words, i.e. that no one helped me to author it. Suggestions from the supervisor regarding language and content are excepted. I used generative artificial intelligence technologies³. In consultation with the supervisor, I did not cite them.

Title of paper or thesis:

Detecting Cycles in Time Series with Persistent Homology

Authored by:

If the work was compiled in a group, the names of all authors are required.

Last name(s):

Pêtre

First name(s):

Katharina Silvia Ghislaine

With my signature I confirm the following:

- I have adhered to the rules set out in the Citation Guide.
- I have documented all methods, data and processes truthfully and fully.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for originality.

Place, date

Basel, March 5, 2025

Signature(s)

If the work was compiled in a group, the names of all authors are required. Through their signatures they vouch jointly for the entire content of the written work.

¹ E.g. ChatGPT, DALL E 2, Google Bard

² E.g. ChatGPT, DALL E 2, Google Bard

³ E.g. ChatGPT, DALL E 2, Google Bard