Tampere University

HENRI RIIHIMÄKI

Metric Stabilization of Invariants for Topological Persistence

Tampere University Dissertations 88

Tampere University Dissertations 88

HENRI RIIHIMÄKI

Metric Stabilization of Invariants for Topological Persistence

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Engineering and Natural Sciences
of Tampere University,
for public discussion in the auditorium S2
of the Sähkötalo building, Korkeakoulunkatu 3, Tampere,
on 2 August 2019, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Engineering and Natural Sciences
Finland

Responsible supervisor	Professor Mikko Kaasalainen Tampere University Finland	
Supervisor	Professor Wojciech Chachólski KTH Royal Institute of Technology in Stockholm Sweden	
Pre-examiners	Dr Pawel Dlotko Swansea University United Kingdom	Professor Claudia Landi Università di Modena e Reggio Emilia Italy
Opponent	Assistant Professor Magnus Botnan Vrije Universiteit Amsterdam Netherlands	
Custos	Professor Eero Hyry Tampere University Finland	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Cover design: Roihu Inc.

ISBN 978-952-03-1152-0 (print)
ISBN 978-952-03-1153-7 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-1153-7

PunaMusta Oy – Yliopistopaino
Tampere 2019

What I've come to realize...
is that mathematics is a dangerous seductress and a most demanding lover.

Abstract

Rank or the minimal number of generators is a natural invariant attached to any n-dimensional persistent vector space. However, rank is highly unstable. Building an algorithmic framework for stabilizing the rank in one-dimensional persistence and proving its usefulness in concrete data analysis are the main objectives of this thesis. Studied stabilization process relies on choosing a pseudometric between tame persistent vector spaces. This allows to minimize the rank of a persistent vector space in larger and larger neighbourhoods around it with respect to the chosen pseudometric. The result is the stable rank invariant, a simple non-increasing function from non-negative reals to non-negative reals.

We show how the needed pseudometrics arise from so called persistence contours. Contour is a certain function system which can be generated very efficiently and in implementable way by integrating a so called density function from non-negative reals to strictly positive reals. We prove an algorithmic way of computing the stable rank invariant with respect to a chosen contour. The result of the theoretical development is an embedding theorem showing that persistent vector spaces embed into Lebesgue measurable functions through stable rank.

The success of persistent homology in data analysis has been largely due to the barcode decomposition and its efficient computation. One result of this thesis is that the barcode decomposition can be proved using the monotonicity of the rank with respect to taking a subspace of persistent vector space. This property of the rank only holds in one-dimensional case. We claim that rank is more fundamental for persistence and barcode is but a technical artifact of its properties. Even though barcode is a powerful tool, progress in persistence theory requires invariants generalizing to multi-dimensional persistence and not relying on decomposition theorems.

Recent years have seen active research around mapping barcodes to some

representation that enables statistics of results from persistent homology analysis and connects naturally to machine learning algorithms. Our embedding theorem shows that the stable rank provides a connection to machine learning. One of our main results is the full applicability of our pipeline in practical data analysis. We demonstrate how choosing an appropriate contour can enhance results of supervised learning. Contour can also be seen to act as a form of feature selection on the bar decomposition.

Preface

After few years in various research positions in Finland and abroad I returned to Tampere University of Technology, my master degree institute, on August 2014. At that time in my new research group on inverse problems, there was an interest on data analysis related to the applications the group was working on. I got attracted to this and started bouncing around ideas how I could do some interesting mathematics in connection to data analysis. After being granted a position in the university doctoral programme, in early 2015 I came across something called topological data analysis and persistent homology. I was immediately hooked.

Lacking background in algebraic topology and category theory, and since there was no activity around topological data analysis in Finland, I started from scratch to immerse myself into the field. Fortunately I had the possibility to attend international conferences to connect with the research and people working on TDA. In an (late) summer school in Oxford on September 2015, organizer Ulrike Tillmann introduced me to my future supervisor Wojciech Chachólski from KTH Stockholm. Whether she ever reads this or not, I want to thank Ulrike for this introduction, it largely determined my research path. Since the first days of 2016 until now, early summer of 2019, I made quasi-frequent visits to KTH to work with Wojciech and the group there. The pages to follow are the distillation of these research years.

This being said, I want to express my sincere gratitude to Wojciech for taking me as his student. His generosity, enthusiasm and immense knowledge on math have left a deep mark on me. Truly inspirational figure. I've heard the opinion that mathematics is something one ponders in solitude. After years in a young and vibrant field of research and working with Wojciech I have to disagree. I've never learned so much about math and the working of a true mathematician than in meetings with Wojciech.

I've been lucky to meet and work with amazing people. Thanks to the

TDA group at KTH: Ryan, Oliver, Barbara, Alvin, Francesca and Martina. Great thing about research is that even though we are scattered around the world, I can be quite confident our paths will cross due to research activities. Thanks to José, our meeting in the Applied Topology in Bedlewo conference in summer of 2017 resulted in a collaboration and interesting work on atmospheric clouds, some of which is presented in this thesis.
Thanks to the work mates and fellow PhD students (past and present) in the math department in Tampere. Ellu, Juho, Kalle, Miikka, Marko, Petteri, Markku and Konsta. Great that we didn't only know each other from work but also engaged in extracurricular activities.
Pawel Dlotko and Claudia Landi made great work in pre-examining my thesis and suggesting needed enhancements, which are implemented in this final work. Magnus Botnan was willing to act as an opponent in my thesis defence and Eero Hyry promised to be the custos. Mikko Kaasalainen provided financially stable environment for conducting research. I sincerely thank all of them.
My dear friends Taneli, Paju, Pilvi and Sauli gave much needed balance to hard work in the form of playing, hiking, ice hole swimming, visits to Slovenian wine yards and extended weekends in Madrid. With Jaakko we made more than one hikes in the Alps, some also in connection to work related travels. Thank you so much guys!
Kiitos äiteelleni Railalle siitä kasvatuksesta, mikä lopulta johti siihen, että kirjoitan viimeisiä sanojani tähän väitöskirjaan. Kuka olisi uskonut nuoruuden Seinäjoen kaduilla ja Lakeuden luontopolun hiekkateillä, että tähän tullaan. En minä ainakaan. Kiitos muorilleni Siirille, ihailemalleni ihmiselle, joka on aina ollut kiinnostunut siitä, mitä elämääni kuuluu. Kiitos edesmenneelle isseelleni Antille, hänen ansiostaan opin muitakin taitoja kuin näppäimistön paukuttamisen. Kiitos molemmille, äiteelle ja Artolle, kaikesta avusta Riihimäen sukutilan ylläpidossa näiden väitösvuosien aikana.
I want to thank my dear Hanna. The intense final stages of my PhD would have been way more stressful without you putting my mind to something else. And you made me book a three weeks trip to Bali after an hour consideration.
Extended thanks go to Tux and Allu, even though man's best friends understand chicken filee better than these words.

Tampere 17.6.2019
Henri Riihimäki

Notation

$\mathbf{N}$	natural numbers
$\mathbf{Z}$	integers
$\mathbf{Q}$	rational numbers
$\mathbf{R}$	interval $[0, \infty)$ or reals
$\mathbf{R}_{/0}$	interval $(0, \infty)$
$\mathbf{R}_{\infty}$	interval $[0, \infty]$
K	field
$\mathrm{Ob}(\mathbf{C})$	class of objects in a category $\mathbf{C}$
$\mathrm{Hom}_{\mathbf{C}}(A, B)$	set of morphisms between A and B in a category $\mathbf{C}$
$\mathrm{Hom}(\mathbf{C})$	collection of all Hom-sets of a category $\mathbf{C}$
$\mathbb{1}_A$	identity map on A
X	set or finite metric space
$B(x, r)$	metric ball with center x and radius r
$C(X)$	simplicial complex of X
$\mathrm{Vert}(C)$	0-simplices of simplicial complex C
C_k	set of k-simplices of C
$\check{C}_\varepsilon(X)$	Čech complex of X at radius ε
$VR_\varepsilon(X)$	Vietoris-Rips complex of X at radius ε
$\mathrm{Del}(X)$	Delaunay complex of X
$\mathrm{Alpha}_\varepsilon(X)$	alpha complex of X at radius ε
W_L	witness complex on set $L \subset X$
Spaces	category of simplicial complexes
$\mathbf{Vec}_K$	category of vector spaces over field K
$\Delta(C)_k$	K-vector space with basis all k-simplices of C
∂_k	boundary operator/map $\Delta(C)_k \to \Delta(C)_{k-1}$
$H_k(C, K)$	degree k homology of C over K
I	discrete invariant $I\colon T \to \mathbf{N}$ for a set of objects T
V	persistent vector space, a functor $\mathbf{R} \to \mathbf{Vec}_K$
$V_{a \le b}\colon V_a \to V_b$	transition map of V

$\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$	category of tame persistent vector spaces
$K(s,e)$	finite bar in $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$
$K(s,\infty)$	bar free on one generator in $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$
$\beta_0(V)$	vector space of cokernels of maps $V_{a\leq b}$
$\mathrm{rank}(V)$	$\dim(\beta_0(V))$, rank invariant of V
$\mathscr{M}$	space of Lebesgue measurable functions $f\colon \mathbf{R} \to \mathbf{R}$
$\mathscr{M}_2$	space of Lebesgue measurable functions $f\colon \mathbf{R}^2 \to \mathbf{R}$
$\widehat{I}_d(X)$	stabilization of a discrete invariant of X
L_p	L_p-metric on $\mathscr{M}$ and $\mathscr{M}_2$
$d_{\bowtie}$	interleaving metric on $\mathscr{M}$ and $\mathscr{M}_2$
$\widehat{L_p}$	normalized L_p-metric on $\mathscr{M}_2$
$C(v,\varepsilon)$	contour, function $\mathbf{R}_\infty \times \mathbf{R} \to \mathbf{R}_\infty$
$C_{/\alpha}(v,\varepsilon)$	truncation of $C(v,\varepsilon)$ at α
$C_{//\alpha}(v,\varepsilon)$	translational truncation of $C(v,\varepsilon)$ by α
d_C	pseudometric on $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ with respect to $C(v,\varepsilon)$
$\widehat{\mathrm{rank}_C V}$	stable rank invariant of V with respect to $C(v,\varepsilon)$
$V_C[\delta]$	δ-shift of V with respect to $C(v,\varepsilon)$
$\mathrm{life}_C K(s,e)$	life span of $K(s,e)$ with respect to $C(v,\varepsilon)$

Contents

Chapter 1

Introduction

1.1 Persistence of data

The amount of data and the pace with which it is gathered will not, with high probability, decrease in the foreseeable future. Data is after all raw material of current information age. Collecting vast amounts of data is however a futile endeavour if we lack capabilities to make sense of it. The aim of data analysis then is to turn raw data into useful and interpretable information. Modern data comes in large amounts but also in increasing complexity and dimension as well as in varying forms. The amount of obtainable information increases but it also undermines the possibility of there being a canonical approach to data analysis. Methods in modern data scientist's toolbox should range from statistics to machine learning and to the subject of the work at hand, topological data analysis or TDA. Applications of TDA are already too long to list, to complement references mentioned in this work see for example [10], [36] and [24].
Topological data analysis has its origin in employing methods from algebraic topology to understand qualitatively geometric structure and features in data. Topology is the study of shapes or spaces in a very general setting. Topological properties are those which remain invariant under continuous deformations of the space. For example different stretchings of a closed loop would be topologically the same since the one distinguishing feature, the hole enclosed by the loop remains invariant in the stretchings. Algebraic topology translates questions in topology to questions in algebra by attaching some algebraic structure to spaces. This algebraic structure carries with it fundamental information about the underlying topological space. The question of e.g determining whether two

spaces are topologically the same translates to a question of determining algebraic isomorphism between their algebraic counterparts.

Now a given data set is nothing more than a space from the point of view of topology. The data analysis question then is to understand this space with tools of algebraic topology. Popular saying is that this is the study of the shape of the given data set. Compare this to statistics which is the study of the models which are most likely to produce the observed data. The question is the same, understand the given data, but point of views and tools are different.

There are many issues in modern data that TDA provides solutions to [10, 9]. Data often is very complex to understand in full detail so qualitative descriptions and summaries are more useful for data analysis. This is exactly the field of topology which retains from spaces only their essential features. For example, a map of subway lines is a concrete example of everyday topology. It is not necessary to depict the rails in full geographic detail, only how they connect different stations and lines is crucial. Qualitative global summaries are often also more informative than analysis at some specific paramater value. Choosing an optimal distance parameter in clustering is rather difficult, whereas a dendrogram provides information how the points cluster and how the clusters change over all parameter values. This corresponds to H_0 persistent homology introduced shortly and gives a starting motivation to use algebraic topology in data analysis. Returning to the map of subway lines, it enables simplifying the information compared to geographically precise map. These simplifications provided by summaries of data are crucial for managing and storing large amounts of data.

Different metrics and coordinates on data points are usually not a priori theoretically justified or there is only some coarse preliminary understanding of them. Consider for example a large set of people represented in the form of long vectors whose elements are different attibutes attached to a person. What is, then, the distance to use when measuring the similarity between two persons? And what is the actual relevance of and an appropriate coordinate system for all the different attributes? Data analysis method should not give wildly varying outputs with reasonably slight changes of metrics and coordinates. This is the domain of topological investigations.

From different approaches to TDA, persistent homology, or persistence, is the current main tool, tracing back to 1990s to the works of Frosini [21] with size functions and Robins [41] which already laid out fundamental

ideas of future field of persistence. Persistence offers some appealing characteristics for data analysis. It defines precisely the qualitative features studied. For practical data analysis it is easily computable through linear algebra. Data almost certainly comes with some inherent noise. Noise is a continuous deformation of the data set and by design topological tools for analyzing data should be robust with respect to noise. One distinguishing aspect of persistence theory is indeed that it is provably stable against small perturbations in input data.
Input to persistent homology pipeline is most often a finite metric space or a point cloud data X. Distance matrix of the space contains all the structural information but is complicated for understanding the space globally. First step in persistence is to use the metric d to construct a simplicial complex over the point cloud by sequentially scanning larger neighbourhoods of points. Concretely, for some ε, $k+1$ points create a k-simplex if all the points are pairwise closer than ε. Resulting simplicial complex is called Vietoris-Rips complex of X at scale ε and denoted by $VR_\varepsilon(X)$. Increasing the scale one obtains a filtration, a sequence of Vietoris-Rips complexes ordered by inclusions:

$$\cdots \subseteq VR_a(X) \subseteq VR_b(X) \subseteq VR_c(X) \subseteq \cdots .$$

To simplify the geometric information in the filtration, homology at degree $n \in \mathbf{N}$, H_n, is applied to each complex. Homology computed over a field K maps complexes to K-vector spaces and the inclusions to linear maps. This results in a sequence of vector spaces and linear maps:

$$\cdots \to H_n(VR_a(X)) \to H_n(VR_b(X)) \to H_n(VR_c(X)) \to \cdots .$$

We call this object a persistent vector space. This name is used for example in [10] even though persistence module is a more standard name. Chapter 3 studies the theoretical steps behind persistent vector spaces. It is this algebraic step that allows to extract structural information out of the filtration. Dimensions of homology vector spaces are known as Betti numbers and they count the number of n-dimensional topological features in the complexes such as connected components in H_0, holes and loops in H_1 and higher dimensional voids with larger n. An efficient algorithm for computing persistent Betti numbers of a filtration was presented in 2002 in [19]. The idea of indexing and visualizing homological features by their birth and death values in the filtration also appeared in [19]. This paper can be regarded as the start of computational persistence. Few years

later [49] gave a more general algorithm for computing homology of a persistent vector space, or persistent homology. Perhaps more importantly, [49] gave explanations to some results of [19] by defining the algebraic structure behind persistence. This marks the beginning of algebraic theory of persistence.

Homology makes it possible to obtain global connectivity information of the point cloud in different dimensions. When the filtration is parameterized by one parameter ε, persistent vector space decomposes into interval indecomposables, so called barcode, which was shown in [49] and used as a vizualization of persistent homology. From representation theory point of view, the sequence of vector spaces and linear maps is a case of a representation of a linear quiver and Gabriel's theorem from the theory of quiver representations provides the interval decomposition [36]. The use of representation theory in persistence was brought forth more extensively in [11]. There are other ways of proving the interval decomposition and we will use one approach in Chapter 3. The barcode can be represented by a set of intervals of $\mathbf{R}$, or bars, of the form $[a, b)$, where it's possible that $b = \infty$. As a sequence of vector spaces, a bar is the sequence

$$\cdots 0 \xrightarrow{0} K_a \xrightarrow{\mathbb{1}} K_b \xrightarrow{0} 0 \cdots ,$$

where the vector spaces are one dimensional vector spaces K for all t, $a \leq t < b$, and zero otherwise. The maps between non-zero values are isomorphisms.

Barcode is an invariant for persistent vector spaces. Moreover, the space of barcodes is a metric space with so called bottleneck distance. Crucial theorem in persistence theory is that the barcode is a stable invariant: in essence, if two data sets X and Y are close as metric spaces, then the corresponding barcodes are close in bottleneck distance. Original stability result was proved in [17]. This was soon extended into algebraic stability theorem in [16] by introducing so called interleaving distance between persistent vector spaces. More recently the stability theorem has been extended to multiparameter persistence (see below) in [28], to zigzag persistence in [5] and to generalized persistence in [7].

The amount of activity around stability shows its prominent role in persistence theory. The foundational issue is to define stable invariants. One of the main results of this thesis is the introduction of a, general framework for stabilizing invariants and proving its stability in Chapter 4. From data analysis point of view methods giving very different results for slightly different inputs are not very meaningful.

The actual data analysis step in standard practice of persistence is to interpret the interval decomposition. For a bar $[a, b)$, some topological feature detected by homology is understood to have appeared in the filtered complex at value a. It is then present in the subsequent complexes until filtration value b. For example, points in the data might connect to create a 1-dimensional loop. This loop persists until at some larger value of the distance parameter the points connect further to higher dimensional simplices and the loop vanishes. It is this lifespan of topological features where the name persistent homology derives from. The aim of data analysis is then to attach meaning to these persistent features. Note that the simplicial complexes are constructed using the metric and as such can be regarded as geometric objects. It can be argued that homology then detects different geometric constructions within the point set. The barcode, however, only gives global description of the amount and lifetimes of these constructions. Also coordinates do not play a role, only the relative proximities of data points. Thus in this work we will always talk about topological features.
The pipeline described above only works when the filtration is parameterized by a single parameter, often the distance scale of the metric. Major challenge in the theory of persistence is the extension to multiparameter persistence. Often we would like to filter our data X with more than one way. For example, we might use the distance scale filtration accompanied with a k-nearest neighbours codensity estimator ρ defined by

$$\rho(x) = \sqrt{\frac{1}{k}\sum_{j=1}^{k} d(x, x_j)^2},\ x, x_j \in X$$

For a positive real number α we can then filter the data by density, taking only points x for which $\rho(x) \leq \alpha$. For small values of α the filtering only considers densest regions in the data, effectively discarding outliers. As a result we get simplicial complexes and homology vector spaces parameterized by two parameters. It is well known in representation theory that these representations do not permit a barcode-like decompositions fully characterizing them. In [12] it was pointed out how complicated the invariant for multiparameter persistence is.
One workaround to this in persistence theory is to develop meaningful invariants for multiparameter persistence. These invariants, even though not complete descriptions of multiparameter persistent vector spaces, should be computable, visualizable, capture some relevant persistent

topological aspects from data and be stable. Stable rank invariant was introduced in [14] to attain these goals.

This thesis continues research around stable rank, complementing the previous work with an algorithmic and applied point of view. Stable rank is derived by defining a pseudometric d in the space of persistent vector spaces. Let F denote a persistent vector space. Rank is a natural invariant associated to F but it's highly unstable. Stabilization of rank of F is achieved by taking the minimum rank in larger and larger balls around F with respect to d. The approach of using metrics to stabilize discrete invariants is the key idea of this thesis. In the case of 1-persistence the needed pseudometrics can be produced with so called persistence contours [23], or simply contours, in a very computer implementable way. The study and use of metrics coming from contours of Chapter 5 has been in the center of the work in this thesis.

The rank of F is equal to its minimal number of generators as we will see in Chapter 3. Using fundamental properties of the rank we can prove the bar decomposition of F. Above it was explained that barcode arises from a decomposition theorem for quiver representations. Conceptually stable rank arises from a chosen pseudometric and does not rely on any decomposition theorem. As the barcode also arises from the properties of rank, as a main conclusion from Chapter 3 we propose that rank is something more fundamental in persistence. In one-dimensional case the bar decomposition only provides an efficient way for algorithmically computing stable rank with respect to a chosen contour. The derivation of this algorithm is the essence of Chapter 6.

We develop the theory behind metric stabilization of rank, contours and stable rank. This culminates in an embedding theorem in Chapter 7 showing that persistent vector spaces embed, through stable rank, in the space of Lebesgue measurable functions. This connects to the current development of persistence where the aim is to combine it with more traditional approaches to data analysis, i.e. statistics and machine learning. As a collection of intervals the barcode is difficult object to work with statistically. Machine learning algorithms also do not take inputs like the barcode. There has been many recent attempts to map the barcode into some function space in which it is more natural to develop statistics and which facilitate the connection to machine learning algorithms: persistence landscape [6], persistence image [20], persistence terrace [33] and accumulated persistence function [3]. Various kernel methods have also been proposed, see for example [27], [39] and [13]. Although these are important tools

and developments in the fied of persistence, the issue is that they are all based on the barcode and therefore do not directly generalize to multi-dimensional case. Main result of Chapter 7 shows how stable rank embeds persistent vector spaces into the space of Lebesgue measurable functions and therefore facilitates statistics and machine learning.

The traditional paradigm in persistence analysis has been that long bars indicate meaningful topology in the data and smaller bars are taken to be topological noise. This view is dictated by the idea of some simple geometry underlying the data which persistence analysis tries to recognize. The point of view taken in this work is that we make no assumption of some smooth geometry underlying the data. The length of a bar measures the robustness of a feature to perturbations of the data. Long bars after all represent features present on multiple distance scales and the distance scale can be interpreted as a radius of randomization around a data point. This view, however, is challenged by many recent studies showing that smaller features carry important information: study of brain artery trees in [2], functional networks of [46], relation of fullerene curvature energy with persistence in [48] and analysis of protein structures in [47]. Also [25] found that small loops in atomic configurations of amorphous silica glass provided explanation to observed diffraction peaks. Further examples that we also study in this work are point processes on a unit square. Here no large scale structure is to be expected and differences are found in the small scale clustering and looping of points.

It thus depends on the analysis at hand what is to be taken as topological noise. Data also always comes with some randomness and in persistence this manifests itself as bars in the barcode that carry no relevant information. It is therefore advantageous to be able to take averages to smooth out instabilities in the persistence output. The stable rank makes it possible to study the mean invariant of a group of data sets. With contours we can also emphasize features on different parts of filtration to explore what are the relevant features in the persistence output. Since contours give metrics and each metric gives a stable rank invariant, we can flexibly produce a whole range of stable ranks capturing different aspects of data topology. Large part of the work presented in this thesis was to implement and use the pipeline in various data analysis tasks. Chapter 8 is devoted to present these results. Material in Chapter 8 should be considered with equal importance with the mathematical development in preceding chapters. Using different contours can also be seen as a form of feature selection on bars and we demonstrate how it can be used to enhance classification

accuracy in supervised learning.

This thesis uses the language of category theory and Chapter 2 therefore introduces basics for non-acquainted reader. Besides of discussing the rank, Chapter 3 also gives an outline of persistence data analysis. Although some of this material is standard we will develop it for self-containment.

As a final remark before the subject matter we want the address the question of metrics which play a prominent role in this work. An expert in persistence theory may arguably ask why introduce new metrics in Chapter 5 when interleaving, bottleneck and Wasserstein distances have gained a standard status in the field. We know that the computation of interleaving in general multiparameter persistence is an NP-hard problem [4]. In 1-parameter persistence the isometry theorem [28] states that the interleaving and bottleneck distances are equivalent. But in the case of 1-persistence bottleneck and Wasserstein distances are defined on barcode decompositions and our aim is to shift focus away from barcodes to consider other stabilized invariants. Interleavings work on the algebraic level of persistent vector spaces, whereas for us metrics provide means to move toward analysis through the stabilization and then work in the obtained function spaces. From data analysis point of view, we want to produce a rich space of metrics which can be used for computing a variety of invariants, each capturing different aspects from the data. Simply put, we wish to optimize over the space of metrics for a given data analysis task. The above consideration leads one to ponder why interleaving and bottleneck distances would be so special for persistence analysis. This thesis tries to answer this. The aim of this thesis is also to work as a manual for a possible future user of the presented methodology. Therefore we map the whole path from input data over theoretical details to analysis end results.

1.2 Author's contribution

This thesis is a self-contained expansion of a preprint [40] and a manuscript [15] under review at SIAM Journal on Applied Algebra and Geometry jointly written by the author and W. Chachólski. Point of view of [40] was the use of our pipeline in applications which is the content of Chapter 8. Paper [15] complements this with the mathematical background and especially proves the embedding theorem of Chapter 7. This thesis puts everything together in a single coherent work.

Particularly from the main contributions, Chapter 3 gives much more details for the claim that the rank is a fundamental invariant and for the proof of barcode decomposition than [15]. Chapter 4 studies metric stabilization in more detail than [14] and [23] did before, particularly from analysis point of view. Also in Chapters 4, 5, 6 and 7 the author gives all the details for the theoretical development, much of which were omitted in the more condensed papers [40] and [15]. A new visualization tool for barcodes and contours is introduced in Chapter 8 that helps a practitioner to understand different choices of contours based on the data.
Applications to cloud data in Chapter 8 is a collaboration with the author and J. Licón-Saláiz. Part of this work has been published in [29]. Section 8.4 on homological density estimation and analysis of cloud fields is original for this thesis.
The author wrote an initial implementation to compute stable rank invariant. O. Gäfvert implemented a library to compute contours which was later modified by the author. All further data analysis pipelines used in Chapter 8 were implemented by the author. The author made all the data analyses presented in Chapter 8. The current work is a continuation of the research project presented in [14] and [23].

Chapter 2

Basic category theory

Category theory was initiated due to algebraic topology to bring clarity about the constructions involved in the field. This clarity comes about by leaving the intricacies of a specific mathematical subject behind and focusing on the higher level similarities between different mathematical constructions. Concretely, consider groups and their homomorphisms, vector spaces and their linear transformations, topological spaces and their homeomorphisms and so on. We see that mathematics occupies itself with different constructions with associated mappings between them. Many of these mappings share the same properties and this observation is a key ingredient for category theory.
After building your desired category you might want to study it through some other category by structure-transferring mapping. This is the so called functor which we will define soon and use extensively later in this work. More leisurely terms, category theory can be thought of as a 'social graph' of mathematical objects. A certain category is a group of individuals and associated friendship relations. Further, different categories are related how well they get along, the concept of a functor. To study how different functors are related we will define natural transformations, mappings between fucntors. These will also play a role in later theoretical developments. For more introductory material on category theory see [45] and in the context of persistence theory see [8]. But, without further ado, let us start building concrete understanding of category theory, starting with the axioms for a category.

► **Definition 2.1.** A **category C** consists of the following data:

- A class Ob(**C**) of objects $A, B, C, ...$

- A set $\mathrm{Hom}_{\mathbf{C}}(A, B)$ of morphisms (or arrows or mappings) between each pair of objects A and B in $\mathrm{Ob}(\mathbf{C})$. A morphism $f \in \mathrm{Hom}_{\mathbf{C}}(A, B)$ is thus assigned with a domain A and a codomain B and we write $f\colon A \to B$, or with simple arrow notation $A \to B$. $\mathrm{Hom}_{\mathbf{C}}(A, B) = \emptyset$ is a valid Hom-set. The collection of all Hom-sets in $\mathbf{C}$ is a class $\mathrm{Hom}(\mathbf{C})$.

- There is a composition of morphisms that is associative,

$$f \circ (g \circ h) = (f \circ g) \circ h.$$

  Domains and codomains have to agree such that

$$h\colon A \to B,\ g\colon B \to C,\ f\colon C \to D.$$

- For each object A an identity morphism $\mathbb{1}_A \in \mathrm{Hom}_{\mathbf{C}}(A, A)$ exists such that for any $f \in \mathrm{Hom}_{\mathbf{C}}(A, B)$ and $g \in \mathrm{Hom}_{\mathbf{C}}(B, A)$,

$$f \circ \mathbb{1}_A = f \quad \text{and} \quad \mathbb{1}_A \circ g = g.$$

◀

Looking at the previous definition it becomes clear that the objects themselves play only a minor role. The focus is shifted on the morphisms and their (very minimal) computational requirements. Category can therefore be viewed as defining an abstract algebra of mappings in general. In this respect categories 'try to do without elements'.

► **Definition 2.2.** A category $\mathbf{C}$ is called **small** if both $\mathrm{Ob}(\mathbf{C})$ and $\mathrm{Hom}(\mathbf{C})$ are sets. Otherwise $\mathbf{C}$ is called **large**.

◀

Let us see few examples of standard categories.

▼ **Example 2.3.**
Topological spaces. The category **Top** has the following content:

- $\mathrm{Ob}(\mathbf{Top})$ contains all topological spaces.

- $\mathrm{Hom}_{\mathbf{Top}}(A, B)$ contains all continuous maps from A to B.

- Composition is the usual composition of maps and identity is the identity map for each topological space.

Set. Continuing in similar fashion, we can define category **Set** of sets and functions between them.
Grp. Taking groups and homomorphisms between them as ingredients, we can construct the category **Grp**.
Vec$_K$. This category consists of finite-dimensional vector spaces over field K and linear mappings between them.
Met. This category consists of metric spaces and Lipschitz mappings between them with Lipschitz constant 1.
All the categories in this example are large. ▲

▼ **Example 2.4.**
Posets. A partially ordered set or poset is a pair $(A, \leq)$, where A is a set and $\leq$ is a binary relation satisfying for all $a, b, c \in A$
reflexivity: $a \leq a$,
transitivity: if $a \leq b$ and $b \leq c$, then $a \leq c$,
antisymmetry: if $a \leq b$ and $b \leq a$, then $a = b$.
Mappings that go together with posets are called monotone. A function

$$m\colon A \to B$$

between posets A and B is monotone, when for all $a, a' \in A$,

$$a \leq a' \quad \text{implies} \quad m(a) \leq m(a').$$

Now, for each poset A, the identity function $\mathbb{1}_A$ is clearly monotone:

$$a \leq a' \quad \text{implies} \quad \mathbb{1}_A(a) = a \leq a' = \mathbb{1}_A(a').$$

The composition of monotone functions is also monotone:
Let $f\colon$ A $\to$ B and $g\colon$ B $\to$ C be monotone functions. Then, for $a, a' \in A$, $a \leq a' \quad \text{implies} \quad f(a) \leq f(a')$. This again implies $g(f(a)) \leq g(f(a'))$, so $(g \circ f)(a) \leq (g \circ f)(a')$.
As a function composition the composition of monotone functions is associative. We have shown that posets as objects and monotone functions as morphisms form the large category **Posets**. ▲

The following example will make clear what we said a moment ago, that categories try to do without elements. The two previous examples had structured sets as objects and structure-preserving functions as morphisms, effectively taking elements from one set to another. In the following we have to expand our understanding of morphisms, away from functions acting on elements of sets.

▼ **Example 2.5.**
$(\mathbf{N}^n, \leq)$. Let us now fix the space $\mathbf{N}^n$ of n-tuples of natural numbers and give it the structure of partial order with the following definition:

$$(v_1, v_2, ..., v_n) \leq (w_1, w_2, ..., w_n) \iff v_i \leq w_i \, \forall \, i \in \{1, ..., n\}.$$

The objects in this category are all the elements $v \in \mathbf{N}^n$. Between each pair of objects there is at most one morphism as is evident from the definition of Hom-sets:

$$\mathrm{Hom}(v, w) = \{(v, w) \in \mathbf{N}^n \times \mathbf{N}^n | v \leq w\}.$$

From now on we use the arrow notation $v \to w$ instead of the ordered pair (v, w). Due to reflexivity of $\leq$, identity morphism exists for all objects:

$$v \leq v \Rightarrow \; v \to v \in \mathrm{Hom}(v, v).$$

The composition is defined due to transitivity:

$$\begin{aligned} & v \to w \in \mathrm{Hom}(v, w), w \to z \in \mathrm{Hom}(w, z) \\ \Longrightarrow \; & (v \leq w) \wedge (w \leq z) \\ \Longrightarrow \; & v \leq z \\ \Longrightarrow \; & v \to z \in \mathrm{Hom}(v, z). \end{aligned}$$

We can thus write $(w \to z) \circ (v \to w) = v \to z$. Composition with identity behaves then as it should: $(v \to w) \circ (v \to v) = v \to w$ and $(v \to v) \circ (w \to v) = w \to v$. Associativity is clear also due to transitivity and by resorting to the associativity of conjunction:

$$\begin{aligned} & (x \to z) \circ ((w \to x) \circ (v \to w)) \\ \Longrightarrow \; & (x \leq z) \wedge ((w \leq x) \wedge (v \leq w)) \\ \Longrightarrow \; & ((x \leq z) \wedge (w \leq x)) \wedge (v \leq w) \\ \Longrightarrow \; & ((x \to z) \circ (w \to x)) \circ (v \to w). \end{aligned}$$

In exactly similar manner we define poset categories $(\mathbf{Z}^n, \leq)$, $(\mathbf{Q}^n, \leq)$ and $(\mathbf{R}^n, \leq)$ of n-tuples of integers, rationals and reals. All of these are small categories.

▲

Often we are interested in substructures like subsets, subgroups, subspaces etc., so let us carry this over to categories.

► **Definition 2.6.** A **subcategory C** of a category **D** consists of the following data:

- Ob(**C**) is a subcollection of Ob(**D**)
- Hom(**C**) is a subcollection of Hom(**D**)
- If $f\colon X \to Y$ is in $\mathrm{Hom}_{\mathbf{C}}(X, Y)$, then X and Y are objects in **C**
- If $f\colon X \to Y$ and $g\colon Y \to Z$ are in $\mathrm{Hom}_{\mathbf{C}}(X, Y)$ and $\mathrm{Hom}_{\mathbf{C}}(Y, Z)$, then $g \circ f$ is in $\mathrm{Hom}_{\mathbf{C}}(X, Z)$
- If X is an object in **C**, then $\mathbb{1}_X$ is a morphism in **C**.

◄

▼ **Example 2.7.**
Ab, the category of abelian groups is a subcategory of **Grp**.
FinSet of finite sets is a subcategory of **Set**.

▲

We will use the notion of exactness. In a category where it is possbile to discuss kernels of morphisms, as in $\mathbf{Vec}_K$, a sequence of objects and morphisms

$$\cdots \longrightarrow V_{n+1} \xrightarrow{f_{n+1}} V_n \xrightarrow{f_n} V_{n-1} \longrightarrow \cdots$$

is called **exact** if $\mathrm{im}\, f_{n+1} = \ker f_n$ for all $n \in \mathbf{Z}$.
Now we can enlarge our algebra to include mappings between categories.

► **Definition 2.8.** Let **C** and **D** be categories. A mapping $F\colon \mathbf{C} \to \mathbf{D}$ is called a **functor**, when the following properties hold:

- Objects are sent to objects:

$$A \in \mathrm{Ob}(\mathbf{C}) \implies F(A) \in \mathrm{Ob}(\mathbf{D}),$$

- Morphisms are sent to morphisms:

$$F(f\colon A \to B) = F(f)\colon F(A) \to F(B),$$

- $F(\mathbb{1}_A) = \mathbb{1}_{F(A)}$,
- $F(g \circ f) = F(g) \circ F(f)$.

◄

▼ **Example 2.9.**
There is a floor functor $F\colon (\mathbf{Q}, \leq) \to (\mathbf{Z}, \leq)$ given on objects by $F(q) = \lfloor q \rfloor$, i.e. for a rational q it maps to the largest integer that is less than or equal to q. For a morphism $q \to r$ in $\mathbf{Q}$ we have $F(f\colon q \to r) = \lfloor f \rfloor\colon \lfloor q \rfloor \to \lfloor r \rfloor$. So morphisms are sent to morphisms since $q \leq r$ implies $\lfloor q \rfloor \leq \lfloor r \rfloor$. Identity $q \leq q$ is mapped to the identity $\lfloor q \rfloor \leq \lfloor q \rfloor$. For compositions

$$\begin{aligned} F((r \to s) \circ (q \to r)) &= F(q \to s) \\ &= \lfloor q \rfloor \to \lfloor s \rfloor \\ &= (\lfloor r \rfloor \to \lfloor s \rfloor) \circ (\lfloor q \rfloor \to \lfloor r \rfloor) \\ &= F(r \to s) \circ F(q \to r). \end{aligned}$$

▲

► **Definition 2.10.** In a category $\mathbf{C}$, a morphism $f\colon A \to B$ is an **isomorphism** if there exists morphism $g\colon B \to A$ such that $f \circ g\colon = \mathbb{1}_B$ and $g \circ f\colon = \mathbb{1}_A$.

◄

■ **Proposition 2.11.** *If $F\colon \boldsymbol{C} \to \boldsymbol{D}$ is a functor between categories $\boldsymbol{C}$ and $\boldsymbol{D}$ and f an isomorphism in $\boldsymbol{C}$, then $F(f)$ is an isomorphism in $\boldsymbol{D}$.*

Proof. If $f \circ g\colon = \mathbb{1}_B$, then by definition 2.8

$$F(f \circ g) = F(f) \circ F(g) = F(\mathbb{1}_B) = \mathbb{1}_{F(B)}$$

and

$$F(g \circ f) = F(g) \circ F(f) = F(\mathbb{1}_A) = \mathbb{1}_{F(A)}.$$

■

Functors thus preserve the structure of a category, so we can transfer our category to the realm of another category, or get a "picture" of category $\mathbf{C}$ in a category $\mathbf{D}$. It's evident that functors also compose correctly and that every category $\mathbf{C}$ has an identity functor $\mathbb{1}_{\mathbf{C}}\colon \mathbf{C} \to \mathbf{C}$. So we get another example of a category, the large category **Cat** with small categories as objects and functors as morphisms.

► **Definition 2.12.** Let $\mathbf{C}$ and $\mathbf{D}$ be categories and let $\Phi\colon \mathbf{C} \to \mathbf{D}$ and $\Psi\colon \mathbf{C} \to \mathbf{D}$ be functors. Then the family of morphisms $\tau = \{\tau_X\}_{X \in \mathrm{Ob}(\mathbf{C})}$ is called a **natural transformation** of Φ into Ψ, denoted $\tau\colon \Phi \to \Psi$, if

- for each X in $\mathrm{Ob}(\mathbf{C})$, a morphism τ_X in $\mathrm{Hom}_{\mathbf{D}}(\Phi(X), \Psi(X))$ is given,

- for each pair of objects X and Y in $\mathbf{C}$ and each f in $\mathrm{Hom}_{\mathbf{C}}(X, Y)$, the following diagram commutes:

$$\begin{array}{ccc} \Phi(X) & \xrightarrow{\Phi(f)} & \Phi(Y) \\ {\scriptstyle \tau_X}\downarrow & & \downarrow{\scriptstyle \tau_Y} \\ \Psi(X) & \xrightarrow[\Psi(f)]{} & \Psi(Y) \end{array}$$

The morphism τ_X is called the component of τ at X. ◀

▼ **Example 2.13.**
Consider the floor functor $F\colon (\mathbf{Q}, \leq) \to (\mathbf{Z}, \leq)$, $F(q) = \lfloor q \rfloor$ of Example 2.9. As another functor take the shift of floor of $\lfloor q \rfloor$ by 2: $T\colon (\mathbf{Q}, \leq) \to (\mathbf{Z}, \leq)$, $T(q) = \lfloor q \rfloor + 2$. For any q and r in $(\mathbf{Q}, \leq)$ there is either one unique or zero morphisms between them. When $\mathrm{Hom}_{(\mathbf{Q},\leq)}(q, r)$ is non-empty we can make a commutative diagram in $(\mathbf{Z}, \leq)$ which gives a natural transformation between F and T:

$$\begin{array}{ccc} \lfloor q \rfloor & \xrightarrow{\leq} & \lfloor r \rfloor \\ {\scriptstyle \leq}\downarrow & & \downarrow{\scriptstyle \leq} \\ \lfloor q \rfloor + 2 & \xrightarrow{\leq} & \lfloor r \rfloor + 2 \end{array}$$

▲

Let us take a brief outlook on things to come. Let J be a poset. Persistence can be formalized as a functor $P\colon J \to \mathbf{Vec}_K$ from some poset to vector spaces. In the following chapter we will look into this construction in detail. For now we note that we can make a category $\mathrm{Fun}(\mathbf{C}, \mathbf{D})$ of functors $F : \mathbf{C} \to \mathbf{D}$ and natural transformations between them. The goal of persistence theory is then to study properties of a certain subcategory of the functor category $\mathrm{Fun}(J, \mathbf{Vec}_K)$ and objects therein.
Natural transformations compose vertically. If $\tau\colon \Phi \to \Psi$ and $\eta\colon \Psi \to \Theta$ are natural transformations, their composition is a natural transformation $\eta\tau\colon \Phi \to \Theta$ whose components are given objectwise: $(\eta\tau)_X = \eta_X \tau_X$.

► **Definition 2.14.** Natural transformation τ is a **natural isomorphism** if every component τ_X of τ is an isomorphism. The functors Φ and Ψ are then said to be (naturally) isomorphic. ◀

Chapter 3

Formal persistence pipeline

Different approaches to data analysis have different views on data. Statistics considers data as a realization of sampling from an unknown probalistic model. Aim is then to infer this model from observed data. In machine learning one often wants to learn optimal model for classifying observed data point to a correct class. Learning is performed through different hypothesis models by minimizing empirical error between observed data and hypothesis output on the same data. Persistent homology is a method of turning point set of data into a collection of algebraic objects through geometric constructions. These algebraic objects provide simplification and enable understanding the data by its algebraic fingerprint. In this chapter we outline the theoretical machinery behind standard persistence. We call this formal persistence pipeline since it is simply a formal algebraic topological process of transforming a complicated mathematical object into one that we hope to be able to analyse. Analysis of this object is the main goal of the remaining chapters.

3.1 From data to geometry

Persistence analysis aims to quantify the various-dimensional relations in a dataset. Understanding these relations and the geometry and topology they span is achieved by turning data into a geometric model, which is the aim of this section. Concretely we want to build a functor $\mathbf{F} : \mathbf{R} \to \mathbf{Spaces}$, where **Spaces** denotes the category of simplicial complexes defined in this section. As such topological data analysis does not impose any assumption on the underlying model of the data. Different hypotheses are put in the

form of different relations and metrics used to construct the simplicial complexes on the data.

Data and finite metric spaces

Let us start by making precise how data is viewed in persistence analysis. Recall that a relation R over a set X is a subset of the Cartesian product, $R \subset X \times X$.

► **Definition 3.1.** A **positively parameterized relation** D over set X is given by a sequence of relations $\{D_\varepsilon\}_{\varepsilon=0}^{\infty}$ over X such that $D_\varepsilon \subseteq D_{\varepsilon'}$ whenever $\varepsilon \leq \varepsilon'$.

◄

▼ **Example 3.2.**
Let X be a group of people. We impose on this set an ε-parameterized relation S_ε where $xS_\varepsilon y$ if persons x and y have known each other at most ε years for $\varepsilon \in [0, \infty)$.

▲

When X is a finite set the relations at different values of t can be represented by binary matrices. For parameter value ε, the elements of the matrix $M(D_\varepsilon)$ are $m_{ij} = 1$ if $i, j \in X$ are related in D_ε, otherwise $m_{ij} = 0$. In practical data analysis one always observes only finite amount of data and in the rest of this work we will also consider our data always to be a finite set. Most often in persistence analysis the relation is a metric and data is a finite metric space.

► **Definition 3.3. Metric** on a set X is a map $d : X \times X \to \mathbf{R}$ such that for all $x, y, z \in X$ the following conditions hold:

1. $d(x, y) \geq 0$
2. $d(x, y) = 0 \Leftrightarrow x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, z) \leq d(x, y) + d(y, z)$.

The pair (X, d) is called a **metric space**. When X is a finite set the pair is called a **finite metric space**.

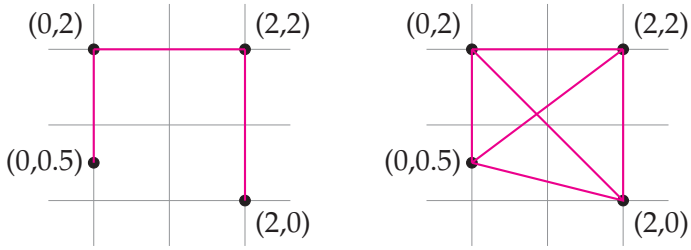
◄

If we discard the second condition and allow $d(x,y) = 0$ for $x \neq y$ and $d(x,x) = 0$ for all x, we obtain a **pseudometric** which will appear frequently in this work. When the metric is allowed to take ∞ as a value, it is called **extended**.
Metric spaces are convenient to describe as parameterized relations. We say that $x, y \in X$ are metrically related at ε if $d(x,y) \leq \varepsilon$. Metric structure of X can be described as a sequence of binary matrices in the same way as relations above. For parameter value ε, the elements of the matrix $M(X_\varepsilon)$ are $m_{ij} = 1$ if $d(i,j) \leq \varepsilon$ for $i, j \in X$, otherwise $m_{ij} = 0$. The structure of a finite metric space is usually given as a single distance matrix, where the element $m_{ij} = d(i,j)$ for $i, j \in X$. Note that the relation in Example 3.2 is not a metric since it is not a real-valued mapping; moreover triangle inequality (condition 4 in Definition 3.3) does not make sense.
Distance matrix contains all the information about the structure of the metric space but it is difficult to obtain global picture of the connectivity of the data based solely on the distance matrix. First step in persistence is to build a family of geometric objects representing the relational or metric structure of the data. These completely determine the resulting geometric objects. To test different models of the data is to impose different relations or metrics on the data and analyze the differences in resulting geometric objects.

▼ **Example 3.4.**
Mr. Euclid and Mr. Chebyshev wanted to investigate the difference between their respective metrics, denoted by d_E and d_C and defined as follows for points in $\mathbf{R}^n$:

$$d_E(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}\,, \qquad d_C(x,y) = \max_i |x_i - y_i|.$$

They decided to place four points on $\mathbf{R}^2$. Both then connected points that they saw were at most distance 2 apart according to their metric. The connections they made were as shown in the figures:



Euclid with d_E Chebyshev with d_C

This investigation of the two gentlemen demonstrates how different metrics can produce drastically different connectivity structures.

▲

Geometry constructions

Let us now take our data set X and focus on turning it into a sequence of geometric objects. For this we first need a way to construct simplicial complexes on X.

► **Definition 3.5.** Let X set. An **abstract simplicial complex** is a pair (X, Σ) where Σ is a collection of non-empty, finite subsets of X, closed under taking subsets: for each $\sigma \in \Sigma$ and for all $\tau \subset \sigma$, $\tau \in \Sigma$. Element $\sigma \in \Sigma$ is called ***k*-simplex**, where $k = |\sigma| - 1$ is the **dimension** of σ. The ***n*-face** of a k-simplex σ, with $n < k$, is a non-empty subset of σ with cardinality $n + 1$. If $\sigma, \tau \in \Sigma$, then $\sigma \cap \tau$ is either empty or face of both σ and τ. The **boundary** of σ is the union of all proper faces and denoted as $\partial\sigma$. When X is a finite set, (X, Σ) is a **finite** simplicial complex. Its dimension is defined to be the dimension of its highest dimensional simplex.

◄

In the following we will write simplicial complex of X as $C(X)$ or simply C, this being understood to be comprised of the collection of subsets Σ. We will denote as $\sigma \in C$ that σ is a simplex in C. Special names are often used for lowest dimensional simplices, particularly when we soon define their geometric realizations: vertex for 0-simplex, edge for 1-simplex, triangle for 2-simplex and tetrahedron for 3-simplex. Vertices are identified with elements of X. We let Vert(C) denote the set of all 0-simplices of a simplicial complex C. When σ is a k-simplex, Vert(σ) denotes its vertex set: Vert(σ) = $\{v_0, ..., v_k\} \subseteq X$. A k-simplex σ with vertex set $\{v_0, ..., v_k\}$ is denoted by $[v_0, ..., v_k]$.

▼ **Example 3.6.**
The finite simplicial complex given by all the subsets of $\{0, 1, \ldots, n\}$ is called the standard n-simplex and is denoted by Δ^n.
▲

► **Definition 3.7.** Let C and C' be simplicial complexes. A **simplicial map** $\phi : C \to C'$ is a function $\phi : \text{Vert}(C) \to \text{Vert}(C')$ such that if $[v_0, ..., v_k]$ is a simplex in C then $[\phi(v_0), ..., \phi(v_k)]$ is a simplex in C'. The set of vertices $\{\phi(v_0), ..., \phi(v_k)\}$ is allowed to have repetitions in which case it spans a simplex of dimension less than k.
◄

Simplicial map is called injective, surjective or an isomorphism if the induced map on the sets of simplices is respectively injective, surjective or bijective.

▼ **Example 3.8.**
Let U be a collection of subsets of X. The k-simplices of a **nerve complex**, $\mathcal{N}(U)$, correspond to the nonempty intersections of $k+1$ distinct elements from U.
▲

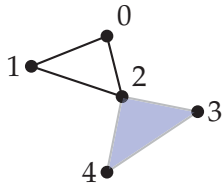
► **Definition 3.9.** A **subcomplex** C' of C is a simplicial complex such that $\text{Vert}(C') \subseteq \text{Vert}(C)$ and if $\sigma \in C'$ then $\sigma \in C$. We then denote $C' \subseteq C$.
◄

Equivalently we can say that C' is a subcomplex of C if the simplicial map $\phi : C' \to C$ is injective.
Abstract simplicial complexes are useful for modelling relations and connectivities in X. Consider Example 3.2. Using graphs it is possible to model only the pairwise social relations between individuals. It is more natural to consider relation between more than two people as a higher-dimensional component they span. For example, three people socially related form a 2-simplex. The pairwise relations are obtained as 1-faces.
For practical computations we need to realize simplices geometrically. Let $p_0, p_1, ..., p_k$ be points in $\mathbf{R}^n$. These $k+1$ points are affinely independent if and only if the k vectors $p_i - p_0$, for $1 \leq i \leq k$, are linearly independent. An $\mathbf{R}$-linear combination $s = \sum \lambda_i p_i$ is a convex combination if $\sum \lambda_i = 1$ and $\lambda_i \geq 0$ for all i. The convex hull is the set of convex combinations. Then a k-simplex is defined to be the convex hull of $k+1$ affinely independent points. Finite simplicial complexes can always be realized geometrically in $\mathbf{R}^n$ for some n. For more on simplicial complexes beyond the details presented in this chapter, we refer to [18], [35], [36] and [43].

▼ **Example 3.10.**
Consider simplicial complex

$$C = \{[0], [1], [2], [3], [4], [0,1], [0,2], [1,2], [2,3], [2,4], [3,4], [2,3,4]\}.$$

Its geometric realization can be visualized in $\mathbf{R}^2$ as below with 5 vertices, 6 edges and 1 triangle (filled with purple).



▲

We will now use the metric to construct simplicial complexes from point sets. We denote the closed metric ball with center x and radius r by $B(x, r)$.

► **Definition 3.11.** Let (X, d) be a finite metric space and $\varepsilon > 0$. **Čech complex** of X at radius ε, $\check{C}_\varepsilon(X)$, has as a k-simplex $k+1$ distinct elements $p \in X$ such that $\cap_{p \in X} B(p, \varepsilon) \neq \emptyset$.

◄

The Čech complex is thus the nerve of the collection $\{B(p, \varepsilon)\}_{p \in X}$. The fact that Čech complex can capture the topology of a point set stems from the following Nerve Theorem [18, 36].

■ **Theorem 3.12.** *Let $\mathcal{U} = \{U_\alpha\}$ be a finite collection of open, contractible subsets of a topological space. If all non-empty intersections of subcollections of $\mathcal{U}$ are contractible, then $\mathcal{N}(\mathcal{U})$ is homotopy equivalent to the union $\cup_\alpha U_\alpha$.*

Since it might be difficult in real computations to check the non-emptiness of intersection of metric balls, a pairwise check for the presence of 1-dimensional simplices can be used as an approximation.

► **Definition 3.13.** Let (X, d) be a finite metric space and $\varepsilon > 0$. **Vietoris-Rips complex** of X at radius ε, $VR_\varepsilon(X)$, has as a k-simplex $k+1$ distinct elements $p \in X$, all whose pairwise distance is $\leq \varepsilon$.

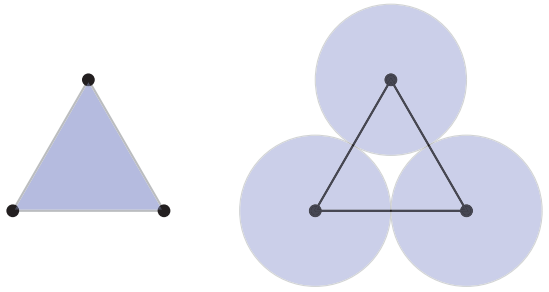
◄

Higher dimensional simplices are thus added when all their 1-faces are present and Vietoris-Rips complex is fully determined by its 1-simplices. Although easy to construct, Vietoris-Rips complex can contain topological artifacts of much larger dimension than the actual underlying space of the point set. Other actual topological features in Vietoris-Rips construction can vanish prematurely when higher dimensional simplex gets filled in immediately after its edges. Obvious from the construction is the inclusion of complexes, $\check{C}_\varepsilon \subseteq VR_\varepsilon$. There is in fact a sequence of inclusions [18].

■ **Theorem 3.14.** *For any $\varepsilon > 0$, there is a chain of inclusions*

$$VR_\varepsilon \subseteq \check{C}_{\sqrt{2}\varepsilon} \subseteq VR_{\sqrt{2}\varepsilon}.$$

▼ **Example 3.15.**
Consider the three vertex points of an equilateral triangle. Since all the vertex points are pairwise equal distance apart, they span a 2-simplex in the Vietoris-Rips construction when the 1-simplices are created. This is illustrated below left. In the Čech construction illustrated below right, the 1-simplices are created when the balls centered on the vertices intersect pairwise. The 2-simplex, however, is not spanned since the intersection of all the three balls is empty.



▲

For its ease of computation, most persistence software use Vietoris-Rips complexes. For completeness, let us also introduce other complex constructions in use in pratical TDA applications.

► **Definition 3.16.** Let (M, d) be a metric space and $X \subset M$ finite. **Voronoi cell** of a point $p \in X$ is the set

$$V_p = \{x \in M \mid d(p, x) \leq d(u, x), u \in X\}.$$

The corresponding **Voronoi diagram** of X is the collection $\{V_p\}_{p\in X}$. The **Delaunay complex** of X is the nerve of the Voronoi diagram:

$$\mathrm{Del}(X) = \{\sigma \subset X | \bigcap_{p\in\sigma} V_p \neq \emptyset\}$$

◀

► **Definition 3.17.** Let (M, d) be a metric space, $X \subset M$ finite and $\varepsilon > 0$. The union of closed balls on X, $\{B(p,\varepsilon)\}_{p\in X}$, is covered by the collection $\{B(p,\varepsilon) \cap V_p\}_{p\in X}$, where V_p is the Voronoi cell on point p. The **alpha complex** at ε is the nerve of this cover:

$$\mathrm{Alpha}_\varepsilon(X) = \{\sigma \subset X | \bigcap_{p\in\sigma} (B(p,\varepsilon) \cap V_p) \neq \emptyset\}$$

◀

Since $B(p,\varepsilon) \cap V_p \subseteq B(p,\varepsilon)$, we have that $\mathrm{Alpha}_\varepsilon(X) \subseteq \check{\mathrm{C}}_\varepsilon(X)$. Even though Čech complex is homotopy equivalent to the union of balls, it can also contain simplices of high dimension at subsets dense enough and with large radius ε. Alpha complex has the benefit that intersecting with Voronoi cells limits the dimension of simplices formed.

► **Definition 3.18.** Let X be a finite metric space and $L \subset X$ a set of **landmarks**. The **weak witness complex**, W_L, has L as a vertex set. A subset of points $S \subset L$ is a simplex in the witness complex if and only if for each subset $T \subset S$ there is a weak witness $x_T \in X$ such that

$$d(t, x_T) \leq d(t, x),$$

for all $t \in T$ and any $x \in L/T$.

◀

■ **Proposition 3.19.** *Finite simplicial complexes and simplicial maps form a category, denoted by* ***Spaces****.*

Proof. For any simplicial complex C there is a map $\mathbb{1}_C : C \to C$ sending each vertex to itself which is clearly simplicial. Then for maps $f : C \to D$ and $g : D \to C$,

$$(f \circ \mathbb{1}_C)([v_0, ..., v_k]) = f([\mathbb{1}_C(v_0), ..., \mathbb{1}_C(v_k)]) = f([v_0, ..., v_k]) \text{ and}$$
$$(\mathbb{1}_C \circ g)([v_0, ..., v_k]) = \mathbb{1}_C([g(v_0), ..., g(v_k)]) = g([v_0, ..., v_k]).$$

For simplicial maps $h : C \to D$, $g : D \to E$ and $f : E \to F$,

$$\begin{aligned}((f \circ g) \circ h)([v_0, ..., v_k]) &= (f \circ g)([h(v_0), ..., h(v_k)]) \\ &= [(f \circ g)(h(v_0)), ..., (f \circ g)(h(v_k))] \\ &= [f((g \circ h)(v_0)), ..., f((g \circ h)(v_k))] \\ &= (f \circ (g \circ h))([v_0, ..., v_k]),\end{aligned}$$

due to f, g and h being associative maps of sets.

■

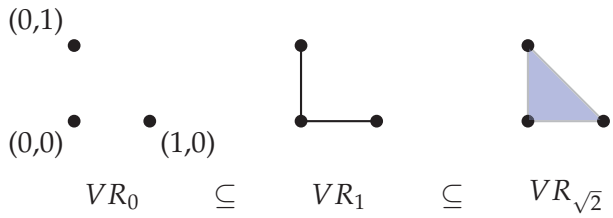
Filtrations

Let X be a finite set with a parameterized relation D. For parameter value ε we can represent the relation D_ε as an abstract simplicial complex C_ε by taking a k-simplex to be a subset of $k+1$ elements that are pairwise ε-related. If X is a finite metric space we can choose one of the metric based simplicial complex constructions of section 3.1 to build simplicial complexes C_ε at different values of ε. Due to construction and by Definition 3.1, the complexes defined above share the property that $C_\varepsilon \subseteq C_{\varepsilon'}$ whenever $\varepsilon \leq \varepsilon'$.

► **Definition 3.20.** Let $(\mathbf{R}, \leq)$ be the set of real numbers as a poset category. A **simplicial filtration** is a sequence of simplicial complexes C_i over $\mathbf{R}$ such that $C_i \subseteq C_j$ whenever $i \leq j \in \mathbf{R}$.

◄

▼ **Example 3.21.**
Illustration of a Vietoris-Rips simplicial filtration of three points.



▲

▼ **Example 3.22.**
Let us continue with Example 3.2. For the Vietoris-Rips construction at scale t, we can form a k-simplex spanned by $k+1$ persons if all the persons have known each other pairwise at most t years. Now at some value of t_1

four people might form a 3-simplex σ denoting a subgroup or a cluster of people in our social relation S_t. At some later value t_2 this cluster might connect to some other cluster ending the independent existence of σ. Even though the lifetime $t_2 - t_1$ of σ could be small, it might still indicate important information in the social relations of our data. ▲

■ **Proposition 3.23.** *Let $(\mathbf{R}, \leq)$ be the set of real numbers as a poset category and **Spaces** the category of finite simplicial complexes. Simplicial filtration is a functor $F : \mathbf{R} \to$ **Spaces**.*

Proof. For each v in $\mathbf{R}$ assign $F(v) = C_v$, where C_v is a simplicial complex at radius v. For each morphism $v \leq w$ we assign $C_v \subseteq C_w$, where $\subseteq$ denotes the simplicial inclusion map. The identity map $v \leq v$ then maps to the identity inclusion $C_v \subseteq C_v$. As simplicial maps,

$$C_u \subseteq C_w = (C_v \subseteq C_w) \circ (C_u \subseteq C_v)$$

for $u \leq v \leq w = (v \leq w) \circ (u \leq v)$. Thus

$$\begin{aligned} F((v \leq w) \circ (u \leq v)) = F(u \leq w) = (C_u \subseteq C_w) &= (C_v \subseteq C_w) \circ (C_u \subseteq C_v) \\ &= F(v \leq w) \circ F(u \leq w). \end{aligned}$$

■

If A is the distance matrix of a metric space X we can attach a simplicial filtration $F(X)$ to X using the information in A together with some simplicial complex construction of defined above. Specifically, $F(\varepsilon)(X) = F_\varepsilon(X) = C_\varepsilon$, where, for example, C_ε is the Vietoris-Rips complex of X at radius ε (see Definition 3.13). This was illustrated in Example 3.21. We conclude this section with the following theorem which essentially says that two finite metric spaces are isometric if and only if their simplicial Vietoris-Rips filtrations are isomorphic. Thus, Vietoris-Rips constructions embed finite metric spaces into sequences of simplicial complexes and we can therefore think about these simplicial complexes as geometric models of finite metric spaces. In this work we always assume that the simplicial complex $F_0(X)$ consists of vertices which are all the points of X.

■ **Theorem 3.24.** *Let $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ be $k \times k$ distance matrices of finite metric spaces X_A and X_B. Then the simplicial Vietoris-Rips filtrations $VR(X_A)$ and $VR(X_B)$ are isomorphic if and only if there is a permutation $\alpha \colon \{1, \ldots k\} \to \{1, \ldots k\}$ such that $a_{ij} = b_{\alpha(i)\alpha(j)}$ for any i and j.*

Proof. $\Rightarrow$: We assume that for any $s \leq t$ we have the following naturally isomorphic commutative square:

$$\begin{array}{ccc} VR(X_A)_s & \subseteq & VR(X_A)_t \\ {\scriptstyle f_s}\downarrow & & \downarrow{\scriptstyle f_t} \\ VR(X_B)_s & \subseteq & VR(X_B)_t \end{array}$$

The vertices of simplicial complexes $VR(X_A)_i$ and $VR(X_B)_i$ are labeled from 1 to k for any i. Since the simplicial maps f_i are isomorphisms, and thus bijections between $\text{Vert}(VR(X_A)_i)$ and $\text{Vert}(VR(X_B)_i)$ for any i, there is a permutation α of $\{1, \dots k\}$.
Let us assume that there are some i, j in $\{1, \dots k\}$ such that $a_{ij} \neq b_{\alpha(i)\alpha(j)}$ and let $[i, j]$ be a simplex in $VR(X_A)_{a_{ij}}$. If $a_{ij} < b_{\alpha(i)\alpha(j)}$, then

$$f_{a_{ij}}([i, j]) = [f_{a_{ij}}(i), f_{a_{ij}}(j)] = [\alpha(i), \alpha(j)] \quad (*)$$

since the isomorphisms f_i define the permutation α. Now $(*)$ contradicts $f_{a_{ij}}$ being an isomorphic simplicial map since $[\alpha(i), \alpha(j)]$ is not a simplex in $VR(X_B)_{a_{ij}}$ since $d(\alpha(i), \alpha(j)) = b_{\alpha(i)\alpha(j)} > a_{ij}$.
Let us then assume that $a_{ij} > b_{\alpha(i)\alpha(j)}$. Then there is a simplex $[\alpha(i), \alpha(j)]$ in $VR(X_B)_{b_{\alpha(i)\alpha(j)}}$ but its preimage $[i, j]$ is not a simplex in $VR(X_A)_{b_{\alpha(i)\alpha(j)}}$, again contradicting $f^{-1}_{b_{\alpha(i)\alpha(j)}}$ being an isomorphic simplicial map.

$\Leftarrow$: We can arrange the entries above diagonal of matrix A into an increasing sequence $S_A = \{a_{qr} \leq a_{st} \leq \dots\}$. The permutation α induces ordering of the above diagonal entries of B, $S_B = \{b_{\alpha(q)\alpha(r)} \leq b_{\alpha(s)\alpha(t)} \leq \dots\}$, and by assumption these sequences are the same. We therefore have simplicial filtrations

$$VR(X_A)_{a_{qr}} \subseteq VR(X_A)_{a_{st}} \subseteq \dots,$$
$$VR(X_B)_{b_{\alpha(q)\alpha(r)}} \subseteq VR(X_B)_{b_{\alpha(s)\alpha(t)}} \subseteq \dots.$$

Also for any t between two consecutive elements x, y of the sequences S_A and S_B, $x < t < y$, we clearly have that $VR(X_\bullet)_x \subseteq VR(X_\bullet)_t \subseteq VR(X_\bullet)_y$, with $\bullet$ equal to either A or B.
Since permutation α is a bijection on $\{1, \dots k\}$, it induces isomorphisms between $\text{Vert}(VR(X_A)_i)$ and $\text{Vert}(VR(X_B)_i)$ for any i. Let $[i, j]$ be a simplex in $VR(X_A)_{a_{ij}}$. Then $[\alpha(i), \alpha(j)]$ is simplex in $VR(X_B)_{b_{\alpha(i)\alpha(j)}}$ since $d(\alpha(i), \alpha(j)) \leq b_{\alpha(i)\alpha(j)} = a_{ij}$. Similarly for any simplex in $VR(X_B)_{b_{\alpha(i)\alpha(j)}}$

its preimage is a simplex in $VR(X_A)_{a_{ij}}$. The above also applies to any higher dimensional simplex since they are determined by their 1-simplex faces. We therefore have isomorphic simplicial maps between $VR(X_A)_t$ and $VR(X_B)_t$ for any t. This makes any square with $s \leq t$ commute:

$$\begin{array}{ccc} VR(X_A)_s \ni \{i,j\} & \hookrightarrow & \{i,j\} \in VR(X_A)_t \\ \downarrow & & \downarrow \\ VR(X_B)_s \ni \{\alpha(i),\alpha(j)\} & \hookrightarrow & \{\alpha(i),\alpha(j)\} \in VR(X_B)_t \end{array}$$

■

Even though Theorem 3.24 validates the use of Vietoris-Rips complexes as models for finite metric spaces, it only applies when the spaces contain the same number of points. Often in data analysis this is not the case and we therefore turn to an algebraic description of the global structure of simplicial complexes. Enter homology.

3.2 From geometry to algebra

In the previous section we translated finite metric spaces into their geometric models. However, these models contain all the information and are thus as complicated as the original spaces. Trying to understand data in its full complexity is often not very fruitful from the point of view of data analysis as well. One seeks simplifications that capture some relevant information from data. As an example consider clustering. Precise geometry of the underlying point set is not interesting, only grouping and corresponding labeling of points. To perform simplification on a simplicial filtration we employ homology, the core concept of this section. Machinery of algebraic topology works by associating spaces with some algebraic structures, vector spaces in the case of homology. In the process full information of the space is lost but one obtains some description of the possibly very complicated space. Homology tells us what kind of topological features the space contains and how many of them. In terms of data, clustering corresponds to degree 0 homology. Analogously, higher degrees of homology measure the higher connectivity of the data. Our aim is to show that homology is a functor from **Spaces** to $\mathbf{Vec}_K$.

Simplicial homology

Homology theory is standard material in almost any book on algebraic topology, see for example [43]. Here we will introduce homology algorithmically in the setting of finite simplicial complexes, explaining how it can be explicitly computed. For more about computational topology see [18] and particularly about computing persistent homology see [35].
Let C be a finite simplicial complex of dimension m. For a natural number k, the symbol C_k denotes the set of k-simplices in C. A **chain complex** is a sequence of vector spaces and linear maps

$$\cdots \longrightarrow V_{n+1} \xrightarrow{f_{n+1}} V_n \xrightarrow{f_n} V_{n-1} \longrightarrow \cdots, \; n \in \mathbf{Z},$$

such that $f_n \circ f_{n+1} = 0$ for each n. Here is an algorithm of how to define and calculate homology with coefficients in a field K for a simplicial complex C:

1. Choose an ordering on C_0 and use it to order elements in any simplex.
2. For natural numbers k and $0 \leq i \leq k$ and a simplex σ in C_k, define $d_i \colon C_k \to C_{k-1}$ to be the simplex in C_{k-1} formed by removing from σ its i-th element. The ordering on C_0 is needed in order to be able to specify which element in a simplex is its i-th element.
3. For any natural number k, let $\Delta(C)_k := \oplus_{\sigma \in C_k} K\sigma$ be the K-vector space with a base given by all simplices in C_k. An element τ in $\Delta(C)_k$ is then given by $\tau = \sum_{\sigma \in C_k} t_\sigma \sigma, \; t \in K$.
4. Define $\partial_k \colon \Delta(C)_k \to \Delta(C)_{k-1}$ to be the linear map assigning to a base element given by a simplex σ in C_k the linear combination $\sum_{i=0}^{k} (-1)^i d_i(\sigma)$ of $(k-1)$-simplices. Let π be a permutation of $\{0, \ldots, k\}$ and define $[v_0, \ldots, v_k] = (\operatorname{sgn} \pi)[v_{\pi(0)}, \ldots, v_{\pi(k)}]$ where $\operatorname{sgn} \pi = \pm 1$ depending on parity of π. This defines the multiplication by -1. The map ∂_k is called the **boundary operator**. Define $\Delta(C)_{-1} = 0$ and $\Delta(C)_k = 0$ for $k > m$.
5. The maps ∂_k form a chain complex (see below for a proof) and homology on degree k over a field K is defined as a quotient vector space:

$$H_k(C, K) = \frac{\ker \partial_k \colon \Delta(C)_k \to \Delta(C)_{k-1}}{\operatorname{im} \partial_{k+1} \colon \Delta(C)_{k+1} \to \Delta(C)_k}, \quad \text{for } k \geq 0.$$

We usually omit the indication of the field K and simply write $H_k(C)$ or even H_k when K and the simplicial complex C are clear.

■ **Theorem 3.25.** *For all $k \geq 0$, $\partial_k \circ \partial_{k+1} = 0$.*

Proof. Let σ be a basis simplex in $\Delta(C)_{k+1}$, $\sigma = [v_0, \ldots, v_{k+1}]$. Then

$$
\begin{aligned}
&\partial_k \circ \partial_{k+1}(\sigma) \\
&\quad = \partial_k \left(\sum_{i=0}^{k+1} (-1)^i d_i(\sigma) \right) \\
&\quad = \partial_k([v_1, \ldots, v_k] - [v_0, v_2, \ldots, v_k] + \cdots + (-1)^{k+1}[v_0, \ldots, v_k]) \\
&\quad = \partial_k([v_1, \ldots, v_k]) - \partial_k([v_0, v_2, \ldots, v_k]) + \cdots + (-1)^{k+1}\partial_k([v_0, \ldots, v_k]) \\
&\quad = [v_2, \ldots, v_k] - [v_1, v_3, \ldots, v_k] + \cdots + (-1)^k[v_1, \ldots, v_{k-1}] \\
&\qquad - [v_2, \ldots, v_k] + [v_0, v_3, \ldots, v_k] - \cdots - (-1)^k[v_0, v_2, \ldots, v_{k-1}] \\
&\qquad + \ldots \\
&\qquad + (-1)^{k+1}[v_1, \ldots, v_k] - (-1)^{k+1}[v_0, v_2, \ldots, v_k] + \ldots \\
&\qquad + (-1)^{2k+1}[v_0, \ldots, v_{k-1}].
\end{aligned}
$$

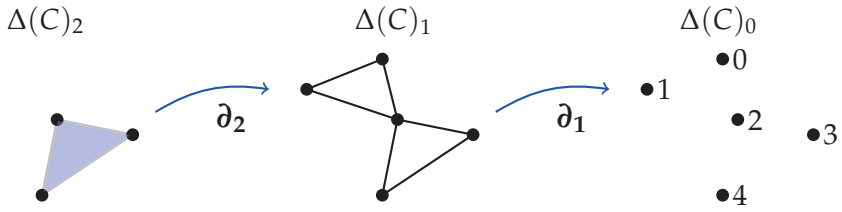
The last sum is arranged into $k+2$ rows each having $k+1$ terms, altogether $k^2 + 3k + 2$ terms which is always even regardless of parity of k. Let us number rows from 0 to $k+1$ and elements in rows from 0 to k. On row n on position m there is σ with nth and mth elements removed with coefficient $(-1)^{m+n}$ if $m < n$. On row m but now on position $n-1$ there is the same simplex with the opposite coefficient $(-1)^{m+n-1}$.
If $m \geq n$, then on row n on position m there is σ with nth and $(m+1)$th elements removed with coefficient $(-1)^{n+m}$. This same simplex with the opposite coefficient $(-1)^{n+m+1}$ occurs on row $m+1$ on position n. Since there are even number of terms, they cancel in pairs and we get the desired claim.

■

Theorem 3.25 establishes the fact that $\operatorname{im} \partial_{k+1} \subseteq \ker \partial_k$ and we can thus take the quotient to define homology in step 4 of the above algorithm.

▼ **Example 3.26.**
Let us compute the homology of simplicial complex in Example 3.10 using the algorithm above. The basis elements of vector spaces of simplices are illustrated below along with the boundary operators ∂_1 and ∂_2. All the

vector spaces and corresponding maps to the left and to the right of the illustration are zero.



Let us choose the field of coefficients K to be the field $\mathbf{F}_2$ with two elements 0 and 1. Particularly $-1 = 1$. We thus have a chain complex

$$\cdots \xrightarrow{0} 0 \xrightarrow{\partial_3=0} \mathbf{F}_2 \xrightarrow{\partial_2} \mathbf{F}_2^6 \xrightarrow{\partial_1} \mathbf{F}_2^5 \xrightarrow{\partial_0=0} 0.$$

We will order the basis simplices with lexicographic ordering. The bases for the relevant vector spaces are then

$$\begin{aligned}
&\Delta(C)_2 : \{[2,3,4]\} \\
&\Delta(C)_1 : \{[0,1],[0,2],[1,2],[2,3],[2,4],[3,4]\}, \\
&\Delta(C)_0 : \{[0],[1],[2],[3],[4]\}.
\end{aligned}$$

Next we compute the boundary operator $\partial_2 : \Delta(C)_2 \to \Delta(C)_1$:

$$\partial_2([2,3,4]) = 1 \cdot [2,3] + 1 \cdot [2,4] + 1 \cdot [3,4].$$

We then get a matrix presentation

$$\partial_2 = [0,0,0,1,1,1]^T.$$

Similarly for ∂_1 using the ordered bases we get a matrix of ∂_1:

$$\partial_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

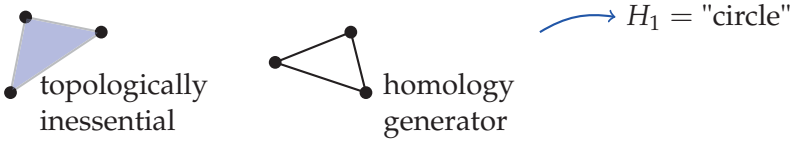
To compute homologies we find the bases of images and kernels of the various boundary operators by row reduction. This gives us

$$\begin{aligned} \text{im } \partial_3 = 0, \ker \partial_2 = 0, \\ \text{im } \partial_2 \simeq \mathbf{F}_2, \ker \partial_1 \simeq \mathbf{F}_2^2, \\ \text{im } \partial_1 \simeq \mathbf{F}_2^4, \ker \partial_0 \simeq \mathbf{F}_2^5. \end{aligned}$$

Homology vector spaces of our complex are thus

$$H_2 = 0, \ H_1 \simeq \mathbf{F}_2, \ H_0 \simeq \mathbf{F}_2.$$

Let us gain some intuition from this example. Linear combinations of simplices are called *chains*. Closed chains are called *cycles* and cycles ought not to have a boundary so they belong to $\ker \partial_k$. In this example $\partial_1([0,1] + [0,2] + [1,2]) = 0$ and $\partial_1([2,3] + [2,4] + [3,4]) = 0$.
Now some cycles in $\Delta(C)_k$ can be *boundaries* of higher-dimensional simplices and so they belong to $\text{im } \partial_{k+1}$. Above we computed that $[2,3] + [2,4] + [3,4]$ is a boundary of $[2,3,4]$. Homology $H_k = \frac{\ker \partial_k}{\text{im } \partial_{k+1}} = \text{"}\frac{\text{cycles}}{\text{boundaries}}\text{"}$ then quotients out k-cycles that are boundaries and we are left with those that represent k-dimensional holes in the complex. For $k = 0$, H_0 measures the number of linearly independent points that make up boundaries of 1-simplices, effectively the number of connected components.
Topologically the one 2-simplex in this example is contractible to a point and is topologically inessential. Homologies H_1 and H_0 indicate that there is one 1-dimensional hole and one connected component. Topologically this complex is then equivalent to a circle as illustrated below.



▲

Matrices of boundary operators store the global connectivity information. If we name columns by the basis elements of $\Delta(C)_p$ and rows by $\Delta(C)_{p-1}$, we see that ones in the matrices correspond exactly to the boundary relations of the simplices (matrices contain only zeros and ones since we are working over $\mathbf{F}_2$). The purpose of the lengthy example above was to gain understanding why we are interested in homology in persistence. Homology gives us exactly the global connectivity information of data that we seek.

Functoriality of homology

Fundamental property of homology is its functoriality and this is where usefulness and effectiveness of homology lies. When we have a simplicial map between simplicial complexes, functoriality makes sure that the information in this map is consistently carried into the linear map between the associated homology vector spaces. For more about functoriality in TDA see [9], and in statistics see [31]. Functoriality is the final piece we need to get to persistent homology and we therefore go through the construction. The proof is standard, see for example [43] or [34].

■ **Theorem 3.27.** *H_k: **Spaces** $\to$ **Vec**$_K$ is a functor for each $k \geq 0$.*

Proof. We have already defined H_k for objects in **Spaces**. Let $\phi\colon C \to C'$ be a simplicial map. This induces a linear map $\phi_*\colon \Delta(C)_k \to \Delta(C')_k$, for any $k \geq 0$, defined on basis simplices $\sigma = [v_0, \ldots, v_k]$ by

$$\phi_*(\sigma) = \begin{cases} \phi(\sigma), & \text{if } \phi(\sigma) \in \Delta(C')_k \\ 0, & \text{otherwise.} \end{cases}$$

Next we need to prove commutativity of diagram

$$\begin{array}{ccc} \Delta(C)_k & \xrightarrow{\partial_k} & \Delta(C)_{k-1} \\ \downarrow{\scriptstyle \phi_*} & & \downarrow{\scriptstyle \phi_*} \\ \Delta(C')_k & \xrightarrow{\partial_k} & \Delta(C')_{k-1} \end{array}$$

Maps ∂_k of course depend on C and C' and the various ϕ_* depend on k. In the following we use standard practice of dropping the indices denoting these dependencies to keep notation less cumbersome.
It is enough to check commutativity on basis simplices σ. We have

$$\partial_k\phi_*(\sigma) = \sum_{i=0}^{k}(-1)^i d_i(\phi(\sigma)) \text{ and } \phi_*\partial_k(\sigma) = \sum_{i=0}^{k}(-1)^i\phi(d_i(\sigma)). \qquad (1)$$

For injective ϕ between Vert(C) and Vert(C') the equality $\partial_k\phi_*(\sigma) = \phi_*\partial_k(\sigma)$ follows immediately from the definitions of ∂_k and ϕ_*.
If ϕ is such that $\phi(\sigma) \in \Delta(C')_l$, $l \leq k-2$, then left of (1) is zero by definition. Right of (1) is zero because for all i, there are at least two repeated vertices in $d_i(\sigma)$ and thus $\phi(d_i(\sigma)) \in \Delta(C')_{l+1}$.

If ϕ is such that $\phi(\sigma) \in \Delta(C')_{k-1}$ then left of (1) is again zero by definition. On the right, order the vertices of σ such that vertices v_s and v_t, $\phi(v_s) = \phi(v_t)$, occupy first two positions. Then the sum has only two non-zero terms and equals

$$[\phi(v_t), \phi(v_j), \ldots, \phi(v_k)] - [\phi(v_s), \phi(v_j), \ldots, \phi(v_k)]$$

which is zero as desired.
The commutativity allows us to show the following:

$$\begin{aligned} \sigma \in \ker \partial_k \implies \partial_k \phi_*(\sigma) = \phi_* \partial_k(\sigma) = 0 \implies \phi_*(\sigma) \in \ker \partial_k, \\ \sigma \in \operatorname{im} \partial_{k+1} \text{ or } \sigma = \partial_{k+1}(\tau) \implies \phi_*(\sigma) = \phi_* \partial_{k+1}(\tau) = \partial_{k+1} \phi_*(\tau) \\ \implies \phi_*(\sigma) \in \operatorname{im} \partial_{k+1}. \end{aligned}$$

The induced map ϕ_* thus sends cycles to cycles and boundaries to boundaries and we can define linear map between homologies:

$$\begin{aligned} H_k(\phi) \colon H_k(C) &\to H_k(C') \\ z + \operatorname{im} \partial_{k+1} &\mapsto \phi_*(z) + \phi_*(\operatorname{im} \partial_{k+1}). \end{aligned}$$

If $i : C \to C$ is an identity simplicial map then i_* is an identity and by definition $H_k(i)$ is an identity on $H_k(C)$.
For simplicial maps $\phi \colon C \to C'$ and $\gamma \colon C' \to C''$, $(\gamma \circ \phi)_* = \gamma_* \circ \phi_*$ as is easily seen from the definition. Then $H_k(\gamma \circ \phi) = H_k(\gamma) \circ H_k(\phi)$ since

$$\begin{aligned} &\gamma_*(\phi_*(z) + \phi_*(\operatorname{im} \partial_{k+1})) \\ &\qquad = \gamma_*(\phi_*(z)) + \gamma_*(\phi_*(\operatorname{im} \partial_{k+1})) \\ &\qquad = (\gamma \circ \phi)_*(z) + (\gamma \circ \phi)_*((\operatorname{im} \partial_{k+1}). \end{aligned}$$

■

Technical question still left to answer is whether homology depends on the orderings of the simplices in a complex. It turns out that it is independent of this choice, more precisely different orderings give isomorphic homology vector spaces. For details see for example [43].

Rank of homology

For a collection of objects T from some category of interest, a **discrete invariant** is a function $I \colon T \to \mathbf{N}$ with values in the set of natural numbers, such that $T(X) = T(Y)$ for isomorphic objects X and Y. For example, take

T to be a collection of vector spaces and I mapping each to its dimension. If T consists of finite metric spaces, invariant I might assign to a metric space the number of clusters obtained by applying some clustering algorithm. Often proving whether two objects are isomorphic from first principles is extremely difficult. To disprove it, it is enough to find an invariant that differs on the objects.
Being a functor, homology vector spaces $H_n(X)$ and $H_n(Y)$, $n \geq 0$ are isomorphic for isomorphic simplicial complexes X and Y. Vector spaces are fully characterized by their dimension so $\dim(H_n(X)) = \dim(H_n(Y))$ for $X \cong Y$. This dimension is called the **rank**. Rank is also known as the nth Betti number of X, $\beta_n(X)$. In Example 3.26,

$$\beta_1(C) = \dim(\ker \partial_1) - \dim(\operatorname{im} \partial_2) = 1 \text{ and } \beta_0(C) = 1.$$

► **Definition 3.28.** Let T be collection of objects in **Spaces**. The mapping

$$\operatorname{rank}_n \colon T \to \mathbf{N}, \quad X \in T \mapsto \dim(H_n(X))$$

is a discrete invariant.

◄

3.3 Persistent vector spaces

Functoriality of homology makes it possible to construct our main objects of study. Let us denote by $(C_t, \subseteq_t)_{t\in\mathbf{R}}$ a simplicial filtration. Applying homology on degree k on all C_t and $\subseteq_t$ we get a sequence of vector spaces and linear maps

$$(H_n(C_t), \to_t)_{t\in\mathbf{R}} = \cdots \to H_n(C_a) \to H_n(C_b) \to H_n(C_c) \to \cdots .$$

We call this a **persistent vector space**. It is an object in the functor category $\operatorname{Fun}(\mathbf{R}, \mathbf{Vec}_K)$. There is a famous correspondence theorem [49] which says we can put persistent vector space into a single module structure. Persistence literature often uses the term persistence module. This is the reason why this section has some resemblance to module theory and we use material for example from [44].
For notational purposes we will also denote persistent vector spaces simply by V and the linear maps $V_{a\leq b} \colon V_a \to V_b$, for $a \leq b$ in $\mathbf{R}$, are called **transition maps**. The transition maps have the property that $V_{b\leq c} \circ V_{a\leq b} = V_{a\leq c}$. Persistent vector space V is a **subspace** of W if $V_t \subseteq W_t$ for any $t \in \mathbf{R}$ and $V_{a\leq b} = W_{a\leq b}|_{V_a}$.

Since the input data for building simplicial filtration is finite, there are only finitely many values in $\mathbf{R}$ when new simplices are added. Therefore there are also finitely many values when the resulting homology vector spces are not isomorphic. The persistent vector spaces we are interested in are thus not arbitrary but satisfy properties of the following definition.

► **Definition 3.29.** A persistent vector space is **tame** if:

- V_t is finite dimensional for all t in $\mathbf{R}$,
- there is a finite sequence $\{t_i\}_{i\in\mathbf{N}}$ such that for any $a < b \in \mathbf{R}$, the transition maps $V_{a\leq b}\colon V_a \to V_b$ are not isomorphisms only if $a < t_i \leq b$.

◄

The subcategory of $\mathrm{Fun}(\mathbf{R}, \mathbf{Vec}_K)$ we will be working with consists of tame persistent vector spaces and is denoted by $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$. It contains a zero object 0 whose value at any t in $\mathbf{R}$ is 0. It also contains direct sums.

► **Definition 3.30.** Direct sum between persistent vector space V and V' is defined

$$V \oplus V' = ((V_t \oplus V'_t)_{t\in\mathbf{R}}, V_{a\leq b} \oplus V'_{a\leq b} = \left[\begin{smallmatrix} V_{a\leq b} & 0 \\ 0 & V'_{a\leq b} \end{smallmatrix}\right])_{a\leq b\in\mathbf{R}}.$$

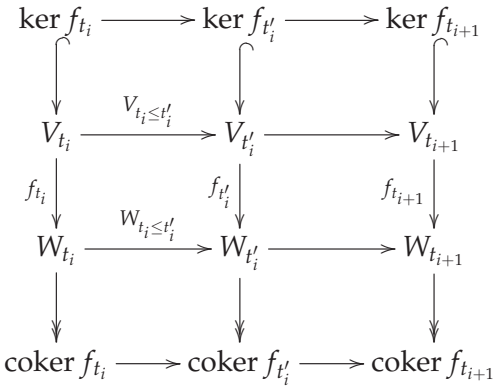
◄

■ **Proposition 3.31.** *Direct sum of tame persistent vector spaces is tame.*

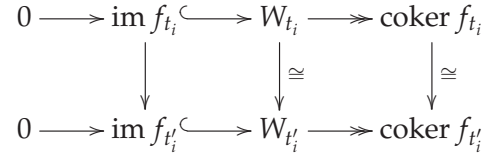
Proof. Let $V, V' \in \mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$. Since V_t and V'_t are finite dimensional for any t, the direct sums $V_t \oplus V'_t$ are finite dimensional.
Let $t_0 < \cdots < t_k$ to be a sequence in $\mathbf{R}$ such that $V_{a\leq b}\colon V_a \to V_b$ is not an isomorphism only if $a < t_i \leq b$. Similarly let $t'_0 < \cdots < t'_l$ to be a sequence in $\mathbf{R}$ such that $V'_{a\leq b}\colon V'_a \to V'_b$ is not an isomorphism only if $a < t'_i \leq b$. Take the union of the sequences to be $s_0 < \cdots < s_{k+l+2}$. This sequence is finite and due to the tameness of V and V' the map $\left[\begin{smallmatrix} V_{a\leq b} & 0 \\ 0 & V'_{a\leq b} \end{smallmatrix}\right]$ is not an isomorphism only if $a < s_i \leq b$. The second condition of tameness is thus satisfied. ■

■ **Proposition 3.32.** *Let $f\colon V \to W$ be a morphism in Tame$(\mathbf{R}, \mathbf{Vec}_K)$. Then $\ker f$, coker f and im f are tame.*

Proof. Let $t_0 < \cdots < t_k$ be a sequence in $\mathbf{R}$ such that $V_{a\leq b}\colon V_a \to V_b$ is not an isomorphism only if $a < t_i \leq b$. Then for any $t_i < t_i' < t_{i+1}$ we have a commutative diagram with $V_{t_i \leq t_i'}$ and $W_{t_i \leq t_i'}$ isomorphisms:



The maps f_{t_i} and $f_{t_i'}$ are then the same and it follows that $\ker f_{t_i} \to \ker f_{t_i'}$ and $\operatorname{coker} f_{t_i} \to \operatorname{coker} f_{t_i'}$ are isomorphisms.
Tameness of $\operatorname{im} f$ follows from the diagram with exact rows:



Proposition 2.71 [44] then gives the unique isomorphism $\operatorname{im} f_t \to \operatorname{im} f_s$. ∎

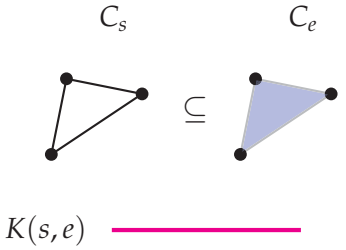
Recall that K is a field and every finite dimensional vector space is isomorphic to K^n for some n. For $s < e$ in $\mathbf{R}$, object $K(s,e)$ in $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ is called a **bar** with start s and end e and defined by

$$K(s,e)_a = \begin{cases} K, & \text{if } s \leq a < e \\ 0, & \text{otherwise,} \end{cases}$$

$$K(s,e)_{a\leq b}\colon K(s,e)_a \to K(s,e)_b = \begin{cases} \mathbb{1}_K, & \text{if } s \leq a \leq b < e \\ 0, & \text{otherwise.} \end{cases}$$

The vector spaces in $K(s,e)$ are finite dimensional for any s and e and there are only s and e in $\mathbf{R}$ such that $K(s,e)_{a\leq b}$ is not an isomorphism if

$a < s \leq b$ and $a < e \leq b$. Therefore every bar functor is tame. Bar functor represents a homological feature that appears in the filtration in simplicial complex C_s and vanishes in complex C_e. A finite bar can be illustrated as a line segment between its endpoints. In the illustration below, a hole appears in C_s and gets filled in by a 2-simplex in C_e and we illustrate the corresponding bar.



If $e < \infty$, then the bar $K(s,e)$ is called **finite**. It is possible that $e = \infty$. In persistence computations, there is always one component that persists indefinitely when all data points are interconnected. In H_0 persistent vector space there is always at least one bar $K(s,\infty)$ representing this component.
Morphisms in $\text{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ between V and W are natural transformations $f\colon V \to W$ or sequences $\{f_a\colon V_a \to W_a\}_{a\in\mathbf{R}}$ of linear maps making the following diagram commute for any $a \leq b$ in $\mathbf{R}$:

$$\begin{array}{ccc} V_a & \xrightarrow{V_{a\leq b}} & V_b \\ {\scriptstyle f_a}\downarrow & & \downarrow{\scriptstyle f_b} \\ W_a & \xrightarrow[W_{a\leq b}]{} & W_b \end{array}$$

$\text{Nat}(V,W)$ denotes the collection of natural transformations between V and W. An element in $\text{Nat}(V,W)$ is surjective/injective/isomorphism if the all the components f_a are surjections/injections/isomorphisms between V_a and W_a. If $f \in \text{Nat}(V,W)$ then $\ker f = \{\ker f_a\}_{a\in\mathbf{R}}$ is a vector subspace of V and $\text{im } f = \{\text{im } f_a\}_{a\in\mathbf{R}}$ is a vector subspace of W.

■ **Theorem 3.33.** *Let $s \in \mathbf{R}$ and V be a tame persistent vector space. The function*

$$y\colon Nat(K(s,\infty), V) \to V_s,\ f \mapsto f_s(1) \in V_s$$

is a bijection.

Proof. For natural transformations $f, g \in \mathrm{Nat}(K(s,\infty), V)$ let us assume that $y(f) = y(g)$ so $f_s(1) = g_s(1) = x$. For all $a < s$ the components f_a and g_a are zero maps. For any $t \geq s$ and for both f and g there is only one choice of components f_t and g_t to make the following diagram commute:

$$\begin{array}{ccc} 1 & \longmapsto & 1 \\ \downarrow & & \downarrow \\ x & \longmapsto & V_{s\leq t}(x) \end{array}$$

Consequently $f = g$.
Consider element $x \in V_s$. Define component $f_s \colon K \to V_s$ by $f_s(1) = x$. As in the diagram above there are unigue choices f_t, $t \geq s$, to assure commutativity. Thus $f \in \mathrm{Nat}(K(s,\infty), V)$ and y is surjective. ■

Theorem 3.33 is the reason why we say $K(s,\infty)$ to be **free on one generator**. As in the proof of Theorem 3.33, any element x in V_s yields a unique map denoted by the same symbol $x \colon K(s,\infty) \to V$ for which $x_s(1) = x$. Whenever the map $x \colon K(s,\infty) \to V$ is a surjection we say that x **generate** V. Since every map surjects onto its image, the image of x is generated by x.
Let V be in $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$. Denote by $\bar{n}$ the set $\{1, 2, \ldots, n\}$ and define a function $s \colon \bar{n} \to \mathbf{R}$. Consider the set of elements $\{g_i \in V_{s(i)}\}_{i\in\bar{n}}$, where s tells in which vector space each g_i belongs to. Each g_i yields a unique map $g_i \colon K(s(i),\infty) \to V$. We can then define a unique map $(g_1, \ldots, g_n) \colon \oplus_{i=1}^n K(s(i),\infty) \to V$. We say that $\oplus_{i=1}^n K(s(i),\infty)$ is **free on *n* generators**. Being a direct sum of tame objects, $\oplus_{i=1}^n K(s(i),\infty)$ is tame. Since V is tame and particularly since every V_t is finite dimensional, there are only finitely many g_i in $\{g_i \in V_{s(i)}\}_{i\in\bar{n}}$, for some n, that yield unique maps $g_i \colon K(s(i),\infty) \to V$. Therefore every tame persistent vector space is **finitely generated**. It is also true that if V is finitely generated then V_t is finite dimensional for any t. Indeed, if we have a surjection $(g_1, \ldots, g_n) \colon \oplus_{i=1}^n K(s(i),\infty) \to V$ with finite set $\{g_1, \ldots, g_n\}$ it follows that V_t has to be finite dimensional for any t.
The image of a map $(g_1, \ldots, g_n) \colon \oplus_{i=1}^n K(s(i),\infty) \to V$ is denoted by $\langle g_1, \ldots, g_n \rangle$ and is a persistent vector subspace of V finitely generated by $\{g_i \in V_{s(i)}\}_{i\in\bar{n}}$. If $\langle g_1, \ldots, g_n \rangle = V$, then the set $\{g_i \in V_{s(i)}\}_{i\in\bar{n}}$ is a **set of generators** of V. By Propositions 3.31 and 3.32, $\langle g_1, \ldots, g_n \rangle$ is the image of a tame persistent vector space and hence it is tame as well. We conclude that

■ **Proposition 3.34.** *Finitely generated subspaces of tame persistent vector spaces are tame.*

Let $K(s,e)$ be a finite bar, $V \in \text{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ and consider the function mapping a natural transformation $f\colon K(s,e) \to V$ to the element $f_s(1) = x$ in V_s. As in the proof of Theorem 3.33, commutativity of

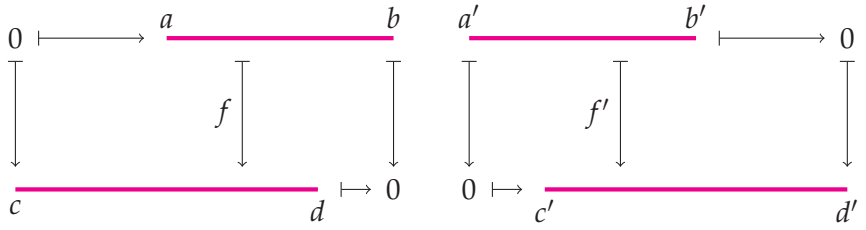
$$\begin{array}{ccc} 1 & \longmapsto & 0 \\ \downarrow & & \downarrow \\ x & \xmapsto[V_{s\leq e}]{} & 0 \end{array}$$

shows that $x \in \ker V_{s\leq e}$. Thus any element x in $\ker V_{s\leq e}$ gives a unique map, denoted by the same symbol, $x\colon K(s,e) \to V$ for which $x_s(1) = x$. The natural transformations $K(s,e) \to V$ are parametrized by elements in $\ker V_{s\leq e}$, in the same way as natural transformations $K(s,\infty) \to V$ are parametrized by elements in V_s.

▼ **Example 3.35.**
Consider the two natural transformations

$$f : K(a,b) \to K(c,d) \text{ and } f' : K(a',b') \to K(c',d')$$

illustrated below.

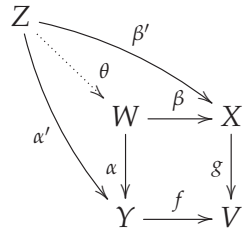


The mappings from and to zero are of course zero maps. To make the naturality squares commute, the components f can be chosen to be identity maps. However, on the right commutativity forces all the components f' to be zero maps. Generally there is a non-trivial map between bars $K(a,b)$ and $K(c,d)$ only if $c \leq a < d \leq b$.

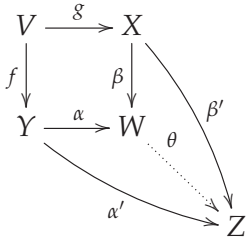
▲

We need to recall the notions of a pullback and a pushout [44]. In the diagram in $\text{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ below, the **pullback** of maps f and g consists

of a tame persistent vector space W together with maps α and β such that $f\alpha = g\beta$. Pullback has the universal property that for any Z and maps α' and β' such that $f\alpha' = g\beta'$, there exists a unique map θ making the diagram commute.



In Tame($\mathbf{R}$, $\mathbf{Vec}_K$) all pullbacks exist (this follows from being an abelian category). The direct sum $W = \ker f \oplus \ker g$ along with the respective projections p_Y and p_X constitute a pullback. We will particularly use the fact that if f is surjective/injective then β is surjective/injective and similarly for g.
In the diagram below, the **pushout** of maps f and g consists of a tame persistent vector space W together with maps α and β such that $\alpha f = \beta g$. Pushout has the universal property that for any Z and maps α' and β' such that $\alpha' f = \beta' g$, there exists a unique map θ making the diagram commute.



As with pullbacks, in Tame($\mathbf{R}$, $\mathbf{Vec}_K$) all pushouts exist. We will use the knowledge that the quotient $X \oplus Y / \mathrm{im}(g, -f)$ with $(g, -f)\colon V \to X \oplus Y$ is a pushout.

3.4 Rank of a persistent vector space

As with homology, we would like to assign a rank invariant for sequences of homologies. Similarly the rank should count homological features in a filtered simplicial complex. The rank is the most fundamental discrete invariant of homology and in this section we define and study properties of rank of objects in Tame($\mathbf{R}$, $\mathbf{Vec}_K$).

Let V be in Tame($\mathbf{R}, \mathbf{Vec}_K$). Let $\{t_i\}_{i\in\mathbf{N}} = t_0 < \cdots < t_k$ be a sequence in $\mathbf{R}$ such that $V_{a\leq b}$ is not an isomorphism only if $a < t_i \leq b$. The map $V_{t_i\leq b}$ is an isomorphism for $b \in (t_i, t_{i+1})$ and thus coker $V_{t_i\leq b} = 0$. Furthermore coker $V_{t_i\leq t_{i+1}} =$ coker $V_{b\leq t_{i+1}}$ (recall the property of transition maps that $V_{t_i\leq t_{i+1}} = V_{b\leq t_{i+1}} \circ V_{t_i\leq b}$ and $V_{t_i\leq b}$ is an isomorphism). Adding elements to the sequence $\{t_i\}_{i\in\mathbf{N}}$ does not change the cokernels and we can thus give definition:

► **Definition 3.36.** Let V be in Tame($\mathbf{R}, \mathbf{Vec}_K$) and $\{t_i\}_{i\in\mathbf{N}} = t_0 < \cdots < t_k$ be a sequence in $\mathbf{R}$ such that $V_{a\leq b}$ is not an isomorphism only if $a < t_i \leq b$. Define its $\boldsymbol{\beta_0}$ to be the vector space

$$\beta_0(V) = V_{t_0} \oplus \text{coker}(V_{t_0\leq t_1}) \oplus \text{coker}(V_{t_1\leq t_2}) \cdots \oplus \text{coker}(V_{t_{k-1}\leq t_k}).$$

Due to tameness of V the vector space $\beta_0(V)$ is finite dimensional. ◄

■ **Theorem 3.37.** *If V and W are isomorphic, then $\beta_0(V)$ and $\beta_0(W)$ are isomorphic.*

Proof. Let $\{t_i\}_{i\in\mathbf{N}} = t_0 < \cdots < t_k$ be a sequence in $\mathbf{R}$ such that $V_{a\leq b}$ or $W_{a\leq b}$ is not an isomorphism only if $a < t_i \leq b$. Let $f\colon V \to W$ be natural isomorphism. Consider the diagram for any i

$$\begin{array}{ccccccc}
V_{t_i} & \xrightarrow{V_{t_i\leq t_{i+1}}} & V_{t_{i+1}} & \xrightarrow{\phi} & \text{coker}(V_{t_i\leq t_{i+1}}) & \longrightarrow & 0 \\
\downarrow{\scriptstyle f_{t_i}} & & \downarrow{\scriptstyle f_{t_{i+1}}} & & \downarrow{\scriptstyle \overline{f_i}} & & \\
W_{t_i} & \xrightarrow{W_{t_i\leq t_{i+1}}} & W_{t_{i+1}} & \xrightarrow{\phi'} & \text{coker}(W_{t_i\leq t_{i+1}}) & \longrightarrow & 0
\end{array}$$

The quotient maps ϕ and ϕ' are surjective with kernels im $V_{t_i\leq t_{i+1}}$ and im $W_{t_i\leq t_{i+1}}$ respectively. The rows are thus exact. The components of f are isomorphisms and the left square commutes by naturality. Proposition 2.70 in [44] then gives the unique isomorphism $\overline{f_i}$ making the diagram commute. We can then define the isomorphism $f_{t_0} \oplus \bigoplus_{i=0}^{k-1} \overline{f_i}$ between $\beta_0(V)$ and $\beta_0(W)$. ■

■ **Proposition 3.38.** *$\beta_0(V \oplus W)$ is isomorphic to $\beta_0(V) \oplus \beta_0(W)$.*

Proof. Let the sequence $\{t_i\}_{i\in\mathbf{N}}$ be such that $V_{a\leq b}$ is not an isomorphism only if $a < t_i \leq b$ and $\{t'_i\}_{i\in\mathbf{N}}$ be such that $W_{a\leq b}$ is not an isomorphism only if $a < t'_i \leq b$. Above we noted that $\beta_0(V)$ and $\beta_0(W)$ do not change

if we add elements to the sequences. Take $\{s_i\}_{i\in\mathbf{N}}$ to be a sequence containing both $\{t_i\}_{i\in\mathbf{N}}$ and $\{t'_i\}_{i\in\mathbf{N}}$. With respect to $\{s_i\}_{i\in\mathbf{N}}$, $\beta_0^{s_1}(V) \cong \beta_0(V)$ and $\beta_0^{s_1}(W) \cong \beta_0(W)$. For $(V \oplus W)_{s_0}$ we can take $V_{s_0} \oplus W_{s_0}$ in $\beta_0(V \oplus W)$. Define transition maps in $V \oplus W$ by $V_{a\leq b} \oplus W_{a\leq b}$ which are not isomorphisms only if $a < s_i \leq b$. The claim follows from the isomorphism ([44])

$$\frac{V_{s_{i+1}} \oplus W_{s_{i+1}}}{\operatorname{im} V_{s_i\leq s_{i+1}} \oplus \operatorname{im} W_{s_i\leq s_{i+1}}} \cong \frac{V_{s_{i+1}}}{\operatorname{im} V_{s_i\leq s_{i+1}}} \oplus \frac{W_{s_{i+1}}}{\operatorname{im} W_{s_i\leq s_{i+1}}}.$$

■

Let us consider what information $\beta_0(V)$ carries. Since the maps $V_{t_i\leq t_{i+1}}$ are not isomorphisms the cokernels may not be zero. The quotient removes from the homology vector space $V_{t_{i+1}}$ the generators, or basis elements, which come from previous non-isomorphic homology vector space. β_0 is thus a vector space of the new homology generators that appear in the persistent vector space. In the context of filtrations of input data sets, this is a way of keeping track of how homology generators evolve in the filtered simplicial complex.

► **Definition 3.39.** For a V in $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$, its **rank** is defined to be

$$\mathrm{rank}(V) = \dim(\beta_0(V)).$$

◄

By Theorem 3.37 rank is a discrete invariant $\mathrm{rank}\colon \mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K) \to \mathbf{N}$ and is the number

$$\dim V_{t_0} + \dim \operatorname{coker}(V_{t_0\leq t_1}) + \dim \operatorname{coker}(V_{t_1\leq t_2}) \cdots + \dim \operatorname{coker}(V_{t_{k-1}\leq t_k}).$$

If V is such that all the transition maps are inclusions, then this number becomes

$$\dim V_{t_0} + \dim V_{t_1} - \dim V_{t_0} + \cdots - \dim V_{t_{k-1}} + \dim V_{t_k} = \dim V_{t_k},$$

the dimension of the last vector space after which all the maps are isomorphisms. Now if V and W are such that all the transition maps are inclusions and $V \subseteq W$, then it follows that $\mathrm{rank}(V) \leq \mathrm{rank}(W)$.

■ **Proposition 3.40.** *The rank has the following properties:*

1. $rank(V \oplus W) = rank(V) + rank(W)$,
2. $rank(V) = 0$ *if and only if* $V = 0$,

3. *rank*$(K(s,e)) = 1$ *for any* $s < e \in \mathbf{R}$,

4. *rank*$(\oplus_{i=1}^{n} K(s(i),e)) = n$ *for any* $s(i) < e \in \mathbf{R}$.

Proof. 1. This follows directly from Proposition 3.38 and since

$$\dim(\beta_0(V) \oplus \beta_0(W)) = \dim(\beta_0(V)) + \dim(\beta_0(W)).$$

2. Assume $\text{rank}(V) = 0$. Then $V_{t_0} = 0$ and all the transition maps are isomorphisms or all the other vector spaces are also 0. In either case we conclude that $V = 0$. The other direction is obvious.

3. This follows from $\dim K_s + \dim \text{coker}(K_s \to K_e) = 1 + 0$.

4. Follows directly from properties 1. and 3. ■

Together with the following crucial theorem the properties in Proposition 3.40 can be used to show that the rank of V equals the smallest number of its generators.

■ **Theorem 3.41.** *Let* $f\colon V \to W$ *be map between tame persistent vector spaces. Then* $f\colon V \to W$ *is surjective if and only if* $\beta_0^f\colon \beta_0(V) \to \beta_0(W)$ *is surjective.*

Proof. As in the proof of Theorem 3.37, there is a unique linear map $\overline{f_i}$ for any i in the following diagram:

$$\begin{array}{ccccccc}
V_{t_i} & \xrightarrow{V_{t_i \le t_{i+1}}} & V_{t_{i+1}} & \xrightarrow{\phi} & \text{coker}(V_{t_i \le t_{i+1}}) & \longrightarrow & 0 \\
\downarrow{\scriptstyle f_{t_i}} & & \downarrow{\scriptstyle f_{t_{i+1}}} & & \downarrow{\scriptstyle \overline{f_i}} & & \\
W_{t_i} & \xrightarrow{W_{t_i \le t_{i+1}}} & W_{t_{i+1}} & \xrightarrow{\phi'} & \text{coker}(W_{t_i \le t_{i+1}}) & \longrightarrow & 0
\end{array}$$

The maps ϕ ans ϕ' are surjections so together with the assumption we have that $\overline{f_i}$'s are surjections. Define $\beta_0^f\colon \beta_0(V) \to \beta_0(W)$ to be $f_0 \oplus \bigoplus_{i=1}^{k} \overline{f_i}$. With this definition the other implication follows from the weak four lemma [30]. ■

If $\text{rank}(\oplus_{i=1}^{n} K(s(i), \infty)) = \dim(\beta_0(\oplus_{i=1}^{n} K(s(i), \infty))) = \dim(\beta_0(V)) = \text{rank}(V)$ then by Theorem 3.41 the map $f\colon \oplus_{i=1}^{n} K(s(i), \infty) \to V$ is a surjection. By Proposition 3.40 $\text{rank}(\oplus_{i=1}^{n} K(s(i), \infty)) = n$. It follows that n is the smallest number such that f is surjective and hence the rank of V equals the minimal number of its generators. Importance of Theorem 3.41 stems from the fact that it applies in multi-dimensional persistence as well. This implies that the minimal number of generators is a fundamental invariant for $\text{Tame}(\mathbf{R}^d, \mathbf{Vec}_K)$ for any d (see [14] for details).

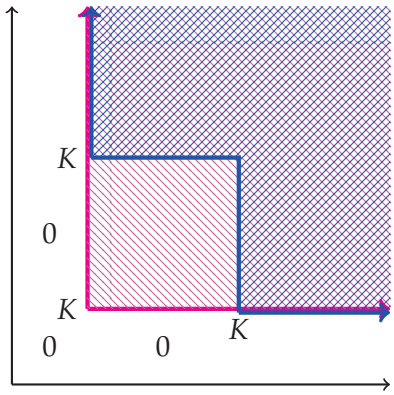
■ **Theorem 3.42.** *$\text{rank}(V) \leq \text{rank}(W)$ for $V \subseteq W$.*

Proof. From the assumption and since W is finitely generated we can construct a pullback square with vertical maps surjections and $m = \text{rank}(W)$:

$$\begin{array}{ccc} P & \hookrightarrow & \oplus_{i=1}^{m} K(s(i), \infty) \\ \downarrow & & \downarrow \\ V & \hookrightarrow & W \end{array}$$

All the transition maps of $\oplus_{i=1}^{m} K(s(i), \infty)$ are inclusions so it follows that all the transition maps of the subspace P are inclusions. We noted above that in this case $\text{rank}(P) \leq \text{rank}(W)$. By Theorem 3.41 $\text{rank}(V) \leq \text{rank}(P)$ and the claim follows. ■

We refer to Theorem 3.42 as the **monotonicity of the rank**. It is a decisive difference between one-dimensional and multi-dimensional persistence. As an example from two-dimensional persistence consider the figure below. The red region visualizes a functor free on one generator with vector spaces K on every point and identity maps between them. The functor marked with blue region is a subspace but free on two generators. The subspace therefore has larger rank. In the next section we see how the monotonicity of the rank leads to the bar decomposition which has been the main tool in persistence since [19] and [49].



3.5 Bar decomposition

We conclude this chapter by showing the fundamental structure theorem for tame persistent vector spaces. It states that any finitely generated object in $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ admits a set of bar generators. In particular, any tame persistent vector space is isomorphic to a direct sum of bars.
Let $V \in \mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$. Let s be the smallest number such that $V_s \neq 0$ and take $x \neq 0$ in V_s. Define $e(x) = \sup\{t \in [s, \infty) \mid V_{s \leq t}(x) \neq 0\}$ to be the end of x. It is possible that $e(x) = \infty$. If $e(x) < \infty$ tameness implies that all the transition maps $V_{s \leq a}$, $a < e(x)$, restricted to the subspace generated by x are isomorphisms and $V_{s \leq e(x)}(x) = 0$. The unique map $x \colon K(s, e(x)) \to V$ induced by x such that $x_s(1) = x$ is injective and hence an inclusion.
We define an element $x \neq 0$ in V_s to generate a bar in V if $x \colon K(s, e(x)) \to V$ has a retraction r that makes the following diagram commute:

$$\begin{array}{ccc} K(s,e(x)) & \xrightarrow{\ x\ } & V \\ & \searrow^{\mathbb{1}} & \downarrow r \\ & & K(s,e(x)) \end{array}$$

In this case $K(s, e(x))$ is a direct summand of V, i.e. $V = K(s, e(x)) \oplus W$ for some W [44]. For a set of generators $\{g_i \in V_{s(i)}\}_{i \in \bar{n}}$, above we defined the start of g_i by the function $s \colon \bar{n} \to \mathbf{R}$. Define the end function $e \colon \bar{n} \to \mathbf{R}$ by $e(i) = \sup\{t \in [s, \infty) \mid V_{s(i) \leq t}(g_i) \neq 0\}$. The set $\{g_i \in V_{s(i)}\}_{i \in \bar{n}}$ is called a **sequence of bar generators** for V if they induce an isomorphism $(g_1 \cdots g_n) \colon \oplus_{i=1}^{n} K(s(i), e(i)) \cong V$.
The structure theorem can be proved by induction on the rank. Let V be a tame persistent vector space. If $\mathrm{rank}(V) = 0$, then $V = 0$ and the empty

set is a sequence of bar generators. If $\text{rank}(V) = 1$, $\beta_0(V) = K$ and V is of the form

$$\cdots \to 0 \to K_s \to \cdots \to K_t \to 0 \cdots .$$

for some s and t. Take 1 in K_s and $e(1) = \sup\{t \in [s, \infty) \mid V_{s \leq t}(1) \neq 0\}$. Then there is a retraction $V \to K(s, e(1))$ making the above retraction diagram commute and $1 \in V_s$ generates a bar. The set $\{1 \in V_s\}$ is a sequence of bar generators for V.

Let $\text{rank}(V) = n > 1$. Assume that any W has a sequence of bar generators if $\text{rank}(W) < n$. Choose a set $\{g_i \in V_{s(i)}\}_{i \in \bar{n}}$ of generators of V. Let $l = \max\{e(1) - s(1), \ldots, e(n) - s(n)\}$, the largest lifetime of generators. From the set of generators with lifetime l, $\{g_i \mid e(i) - s(i) = l\}$, choose g_j with the largest end $e(j)$. That g_j generates a bar follows from the diagram, where $t < e(j)$ and $V = W \oplus \text{im } g_j$, where $W / \text{im } g_j$:

$$\begin{array}{ccccccc}
0 & \longrightarrow & K_{s(j)} & \longrightarrow & K_t & \longrightarrow & 0 \\
\downarrow & & \downarrow{\scriptstyle g_j} & & \downarrow & & \downarrow \\
W_q \oplus (\text{im } g_j)_q & \longrightarrow & W_{s(j)} \oplus (\text{im } g_j)_{s_{(j)}} & \longrightarrow & W_t \oplus (\text{im } g_j)_t & \longrightarrow & W_u \oplus (\text{im } g_j)_u \\
\downarrow & & \downarrow{\scriptstyle r} & & \downarrow & & \downarrow \\
0 & \longrightarrow & K_{s(j)} & \longrightarrow & K_t & \longrightarrow & 0
\end{array}$$

Since g_j is a generator the induced map surjects onto its image. Hence in the middle row $\text{im } g_j = 0$ for $q < s(j)$ and $t \geq e(j)$. The map $V_{q \leq s(j)}|_{W_q} \oplus 0$ makes the lower left square commute. The unique retract r is induced from $W_{s_{(j)}} \oplus (\text{im } g_j)_{s_{(j)}} \mapsto 0 \oplus 1$. This implies that V is isomorphic to $K(s(j), e(j)) \oplus W$. In fact, $W = \ker r$. Thus $\text{rank}(W) = n - 1$ and by induction W admits a sequence of bar generators. Note that the proof is exactly the same even if $e(j) = \infty$.

Since $V \cong \oplus_{i=1}^{n} K(s(i), e(i))$, $\text{rank}(V) = \text{rank}(\oplus_{i=1}^{n} K(s(i), e(i))) = n$ by 3.37 and 3.40. Our discrete invariant for V, the rank of V or the minimal number of generators, is therefore equal to the number of its bar generators in the case of one-dimensional persistence.

Important point to make here is the role played by the monotonicity of the rank, Proposition 3.40. The whole derivation of the bar decomposition rested on the inductive property that a subspace has smaller rank. Rank is a fundamental invariant in persistence which by Theorem 3.41 generalizes to persistent vector spaces parameterized by $\mathbf{R}^n$ for any n. But the bar decomposition, an edifice upon which so much theory and applications

of topological data analysis rests, is a technical consequence of the monotonicity of the rank which only holds in one-dimensional persistence as we saw at the end of previous section. Barcode and vast developments built on top of it are of course the cornerstone of persistence theory and its applications. But due to its restricted generalizability we propose an alternative avenue. Starting from next chapter we see how the rank leads to a stable invariant and we work through its theory to reach the final chapter showing that the stabilization of the rank is a valid invariant for concrete data analysis applications.

Chapter 4

Metric stabilization

In the previous chapter, the rank invariant of a tame persistent vector space was found to be the number of bar generators in its bar decomposition. Discrete invariants, however, are usually not stable. Consider a simplicial filtration of data set X. At some value t there might be a large number of short-lived simplices occurring. This makes dim coker$(V_{s\leq t})$ large for some s. Correspondingly the rank obtains a large value even though the simplices might be artifacts due to noise in X.
This chapter introduces **metric stabilization**, a general framework for converting discrete invariants into stable invariants. To obtain stable invariants is a key step central to persistence analysis. After all, distinguishing feature of persistence theory as a method for data analysis is its stability as discussed in the introduction.
Let us fix T to be a collection of interesting objects from some category and I the invariant of interest. For the purposes of the current work T is of course a set of finite metric spaces and I is the rank of associated tame persistent vector space. The key in converting a discrete invariant into a stable one is the choice of a pseudometric d on T. Once a pseudometric is chosen, we can take a t-radius ball around X, $B(X,t) = \{Y \mid d(X,Y) \leq t\}$, and look at the function $\widehat{I}_d(X)$ taking the minimum value of I on a ball around X with increasing radii t:

$$\widehat{I}_d(X)(t) = \min\{I(Y) \mid Y \in B(X,t)\}.$$

The function $\widehat{I}_d(X)$ is therefore non-increasing and piecewise constant with values in natural numbers $\mathbf{N}$, namely a simple function, and hence Lebesgue measurable. We will regard $\mathbf{N}$ as embedded in $\mathbf{R}$. Ultimately we will want to take e.g. pointwise means of the functions and the values

of the mean no more lie in $\mathbf{N}$. Let us therefore denote by $\mathscr{M}$ the space of Lebesgue measurable functions from non-negative reals to non-negative reals, $\mathscr{M} = \{f : \mathbf{R} \to \mathbf{R}\}$. In general functions $\widehat{I}_d(X)$ are not continuous so we cannot use the space of continuous functions $C(\mathbf{R})$. We will anyway want to work with some nice class of funtions and particularly for our purposes we need to be able to integrate. Since $\widehat{I}_d(X)$ is non-increasing with values in $\mathbf{R}$ there is some t such that for all $s \geq t$ in $\mathbf{R}$, $\widehat{I}_d(X)(s) = \widehat{I}_d(X)(t)$. The function $\widehat{I}_d(X)$ is thus eventually constant and has a **limit**, $\lim \widehat{I}_d(X)$.

► **Definition 4.1. Metric stabilization** of a discrete invariant $I : T \to \mathbf{N}$ with respect to a pseudometric d on T is the function

$$\widehat{I}_d : T \to \mathscr{M},\ T \ni X \mapsto (\widehat{I}_d(X) : \mathbf{R} \to \mathbf{R}).$$
$$\text{Explicitly, } \widehat{I}_d(X)(t) = \min\{I(Y) \mid Y \in T,\ d(X,Y) \leq t\}$$

◄

The importance of the range $\mathscr{M}$ of $\widehat{I}_d$ is that as a function space it has more structure than the range $\mathbf{N}$ of the original invariant I. With I we can only look at the difference in ranks. As we will see in Chapters 6 and 8, $\widehat{I}_d$ associated to an object in $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ gives much more refined information about the behaviour of the rank. In particular $\mathscr{M}$ has much richer geometry with metrics that are of interest to us. We have the standard ***L_p*-metrics** for $p \geq 1$ (we will denote the metric by L_p for notational purposes):

$$L_p(f,g) = \left(\int_0^\infty |f(t) - g(t)|^p dt \right)^{1/p}.$$

We can also define **interleaving** distance between functions $f, g \in \mathscr{M}$:

$$S = \{\varepsilon \in \mathbf{R} \mid f(t) \geq g(t+\varepsilon) \text{ and } g(t) \geq f(t+\varepsilon) \text{ for all } t \in \mathbf{R}\}$$
$$d_{\bowtie}(f,g) = \begin{cases} \inf(S) & \text{, if } S \text{ is non-empty} \\ \infty & \text{, otherwise.} \end{cases}$$

If $\inf(S) = \varepsilon$, we say f and g are ***ε*-interleaved**. In general S is a sequence whose limit is $d_{\bowtie}(f,g)$. Note that for non-increasing functions, if f and g are ε-interleaved, then for any $\tau > \varepsilon$

$$f(t) \geq g(t+\tau) \text{ and } g(t) \geq f(t+\tau) \text{ for all } t \in \mathbf{R}.$$

We remark that interleaving between functions $\gamma : X \to \mathbf{R}^n$, where X is a topological space, was considered in [28]. There the shift of γ was

defined in the values of the function so that ε-shift is $\gamma(x) - \varepsilon$, whereas in our definition the shifts are on the argument domain. In X one cannot naturally talk about the shifts of points. Our definition also allows to prove the continuity of metric stabilization of Theorem 4.3. As the proof shows, our definition of interleaving of functions is inherently related to the stabilization with respect to enlarging metric balls.

■ **Proposition 4.2.** *The interleaving $d_{\bowtie}$ is a pseudometric on $\mathscr{M}$.*

Proof. Since $d_{\bowtie}$ is defined as an infimum of a set of non-negative real numbers and for every function $f(t) \geq f(t)$, $d_{\bowtie}(f,g) \geq 0$ and $d_{\bowtie}(f,f) = 0$ by definition. That interleaving is symmetric follows straight from the definition as well.
For triangle inequality let $d_{\bowtie}(f,h) = \alpha$ and $d_{\bowtie}(h,g) = \beta$. Then there are sequences $\{\tau_i\}$ with limit α and $\{\tau_i'\}$ with limit β. Thus

$$\begin{aligned} f(t) \geq h(t+\tau_i) \text{ and } h(t) \geq f(t+\tau_i), \\ h(t) \geq g(t+\tau_i') \text{ and } g(t) \geq h(t+\tau_i') \end{aligned}$$

for all i and any $t \in \mathbf{R}$. We then have

$$f(t) \geq g(t+\tau_i+\tau_i') \text{ and } g(t) \geq f(t+\tau_i+\tau_i'),$$

which implies in the limit that $d_{\bowtie}(f,g) \leq \alpha + \beta$. ■

The next Theorem 4.3 establishes Lipschitz and Hölder-like continuity of metric stabilization. Since the constant c in the second point depends on the functions, this is not strictly Hölder continuity. The result is anyway a stability theorem of metric stabilization. This is the crucial property and states that it is a process of converting a discrete invariant $I\colon T \to \mathbf{N}$ into a stable invariant $\widehat{I}_d\colon T \to \mathscr{M}$. From data analysis point of view the importance of metric stabilization in the light of this theorem is that we can map persistent vector space from our TDA pipeline in a continuous way into a function space where more statistical and machine learning methods are available. This has been an area of extensive research as we discussed in the introduction.
For the proof we use the fact $(b-a)^p \leq b^p - a^p$ for $a \leq b$ in $\mathbf{R}$ and $p \geq 1$. Let us show that this holds. Since $a \leq b$ there is some $t \geq 1$ such that $ta = b$. Now the inequality is equivalent to $(ta-a)^p \leq t^p a^p - a^p$ which simplifies to $(t-1)^p \leq t^p - 1$. Consider the function $f(t) = t^p - (t-1)^p - 1$. We need to show $f(t) \geq 0$ for all $t \geq 1$. At $t = 1$ this holds. The derivative

$f'(t) = p(t^{p-1} - (t-1)^{p-1})$ is positive so $f(t)$ is an increasing function and the claim follows.

■ **Theorem 4.3.** *Let d be a pseudometric on T, $I\colon T \to \mathbf{N}$ a discrete invariant and $p \geq 1$ a real number. Then for any X and Y in T:*

1. $d(X,Y) \geq d_{\bowtie}\left(\widehat{I}_d(X), \widehat{I}_d(Y)\right)$,
2. $c\, d(X,Y)^{1/p} \geq L_p\left(\widehat{I}_d(X), \widehat{I}_d(Y)\right)$, *where* $c = max\{\widehat{I}_d(X)(0), \widehat{I}_d(Y)(0)\}$.

Proof. (1): If $d(X,Y) = \infty$ we are done. Assume $\varepsilon = d(X,Y) < \infty$. For any t in $\mathbf{R}$, let $Z \in B(Y,t)$ so $d(Y,Z) \leq t$. By triangle inequality $d(X,Z) \leq d(X,Y) + d(Y,Z) \leq \varepsilon + t$ which implies $B(Y,t) \subset B(X,t+\varepsilon)$. Similarly there is an inclusion $B(X,t) \subset B(Y,t+\varepsilon)$. By the definition of metric stabilization these imply that $\widehat{I}_d(Y)(t) \geq \widehat{I}_d(X)(t+\varepsilon)$ and $\widehat{I}_d(X)(t) \geq \widehat{I}_d(Y)(t+\varepsilon)$. As this happens for any $t \in \mathbf{R}$ the claim follows.
(2): From (1) the claim can be formulated as

$$c\, d(X,Y)^{1/p} \geq c\, d_{\bowtie}\left(\widehat{I}_d(X), \widehat{I}_d(Y)\right)^{1/p} \geq L_p\left(\widehat{I}_d(X), \widehat{I}_d(Y)\right),$$

so it is enough to prove that for non-increasing functions f and g:

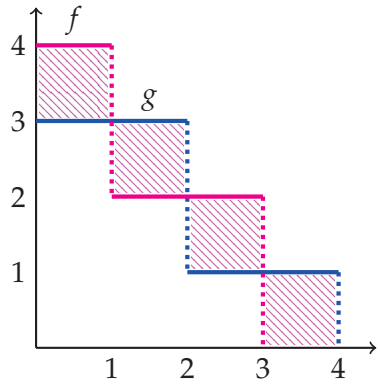
$$\max\{f(0), g(0)\} d_{\bowtie}(f,g)^{1/p} \geq L_p(f,g).$$

This is clear if $d_{\bowtie}(f,g) = \infty$ so assume that $d_{\bowtie}(f,g) = \varepsilon < \infty$. Define the upper envelope of f and g to be the function $h(x) = \max\{f(x), g(x)\}$ and its shift the function $h_\varepsilon(t) = h(t+\varepsilon)$. Since f and g are non-increasing we have that $h \geq f \geq h_\varepsilon$ and $h \geq g \geq h_\varepsilon$. Then $h - h_\varepsilon \geq f - g$ and therefore $L_p(h, h_\varepsilon) \geq L_p(f,g)$.
Now $h_\varepsilon \leq h$ so the inequality derived above gives $(h - h_\varepsilon)^p \leq h^p - h_\varepsilon^p$. The claim now follows from

$$\begin{aligned}
L_p(h, h_\varepsilon) &= \left(\int_0^\infty (h(t) - h_\varepsilon(t))^p dt\right)^{1/p} \leq \left(\int_0^\infty h(t)^p - h_\varepsilon(t)^p dt\right)^{1/p} \\
&= \left(\int_0^\varepsilon h(t)^p dt + \int_\varepsilon^\infty h(t)^p dt - \int_0^\infty h(t+\varepsilon)^p dt\right)^{1/p} \\
&= \left(\int_0^\varepsilon h(t)^p dt + \int_\varepsilon^\infty h(t)^p dt - \int_\varepsilon^\infty h(t)^p dt\right)^{1/p} \\
&\leq (h(0)^p \varepsilon)^{1/p} = \max\{f(0), g(0)\}\varepsilon^{1/p}.
\end{aligned}$$

■

▼ **Example 4.4.**
Consider functions f and g illustrated below. The shaded area represents the value of the integral $\int_0^\infty |f(x) - g(x)|dx$ so $L_1(f,g) = 4$. Interleaving distance between f and g is 1 and $\max\{f(0), g(0)\} = 4$ so the inequality $\max\{f(0), g(0)\}d_{\bowtie}(f,g) \geq L_1(f,g)$ from the proof of Theorem 4.3 is sharp for this example.



▲

Observe that the stability results for metric stabilization in Theorem 4.3 followed from the choice of a pseudometric d on T. Different pseudometrics on T lead to different invariants. The point of view here is that persistence is about identifying pseudometrics on T. The associated stabilized invariants then reflect structural properties of T relevant to a given data analysis task. The expectation is that these properties should be reflected by the geometry of the image of $\widehat{I}_d$ described by metrics in $\mathscr{M}$ if an appropriate pseudometric d on T is chosen. We explore this in Chapter 8.

Chapter 5

Metrics

The key parameter of metric stabilisation in the previous chapter was the pseudometric d on $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$. In this chapter we look how we can produce a rich space of these metrics. We briefly summarize the theory of so called noise systems. This leads to our main source of metrics called persistence contours or simply contours.

5.1 Noise systems

Noise systems were introduced in [14] as a general machinery to define metrics in the category $\mathrm{Tame}(\mathbf{Q}^n, \mathbf{Vec}_K)$ of multidimensional persistent vector spaces. We give a brief account of this theory as a background of our developments. Instead of $\mathrm{Tame}(\mathbf{Q}^n, \mathbf{Vec}_K)$ in this section we frame the main points of the theory in terms of $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ as this is our category of interest. The original source [14] used $\mathbf{Q}^r$ as the indexing category for minor technical details.
Idea behind a **noise system** is to say that some tame functors (persistent vector spaces) are ε-small for some number $\varepsilon \geq 0$. The collection of ε-small functors is called an ε-component of noise. The formal definition is

► **Definition 5.1.** A noise system in $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ is a sequence $\{S_\varepsilon\}_{\varepsilon \in \mathbf{R}}$ of collections of tame functors such that:

1. the zero functor belongs to S_ε for all ε,
2. if $0 \leq \tau < \varepsilon$, then $S_\tau \subseteq S_\varepsilon$,
3. if $0 \to F \to G \to H \to 0$ is an exact sequence in $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$, then

a) if $G \in S_\varepsilon$, then $F \in S_\varepsilon$ and $H \in S_\varepsilon$,

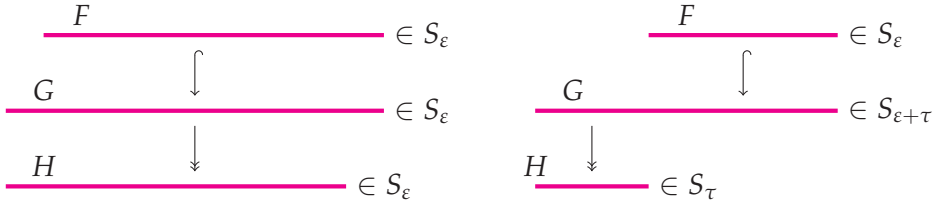
b) if $F \in S_\varepsilon$ and $H \in S_\tau$, then $G \in S_{\varepsilon+\tau}$.

◀

An example of a noise system is **standard noise**. Its components are defined by

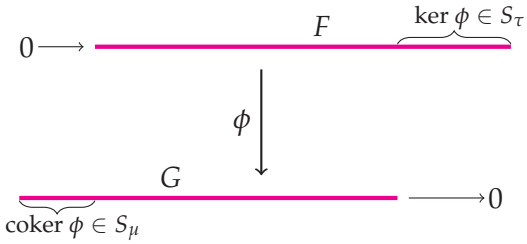
$$S_\varepsilon^{\text{stand}} = \{V \in \text{Tame}(\mathbf{R}, \mathbf{Vec}_K) \mid V_{t \le t+\varepsilon} = 0 \text{ for all } t \in \mathbf{R}\}.$$

A bar functor $K(s,e)$ would then be an element in the S_ε-component of standard noise if $e - s \le \varepsilon$. The property (2) of noise, called component monotonicity, says that $K(s,e)$ is also an element in S_τ-component for $\tau > \varepsilon$. The property (3a) is the functor inclusion (left in the illustration below for bars) and property (3b) is called additivity (right in the illustration).

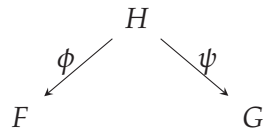


A noise system $\{S_\varepsilon\}_{\varepsilon \in \mathbf{R}}$ is **closed under direct sums** if, for any ε, the component S_ε is closed under direct sums: F and G belong to S_ε their direct sum $F \oplus G$ belongs to S_ε.

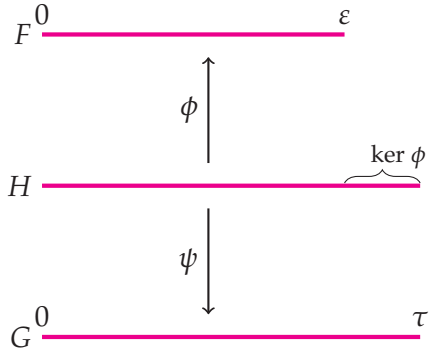
The component S_ε is an abstraction of a closed disc of radius ε around the zero functor and the members of S_ε are functors which are ε-small. The notion of smallness was extended to a notion of proximity among finitely generated tame functors F and G in [14]. A natural transformation ϕ between functors F and G in $\text{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ is called an ***ε*-equivalence** if $\ker \phi$ is in S_τ, $\text{coker } \phi$ is in S_μ and $\tau + \mu \le \varepsilon$. This is illustrated below for bars.



Then define F and G to be **ε-close** if there exists H in Tame($\mathbf{R}$, $\mathbf{Vec}_K$) and natural transformations ϕ and ψ in the diagram below such that ϕ is a τ-equivalence, ψ is a μ-equivalence and $\tau + \mu \leq \varepsilon$.



The reason to introduce intermediate functor H is that there might not be non-trivial natural transformations between all functors. Consider the illustration below with bars F and G. Recall from Example 3.35 that there is no non-trivial map between them. They are, however, both in S_τ (with respect to standard noise). We would like then to have a notion that they are close. We therefore introduce functor H as shown. Then $\phi\colon H \to F$ is a $(\tau - \varepsilon)$-equivalence and $\psi\colon H \to G$ is a 0-equivalence. Since $\tau - \varepsilon \leq \tau$, F and G are τ-close, in fact $(\tau - \varepsilon)$-close. Of course being in a component S_τ does not imply that two functors are τ-close, similarly as two points in a ball of radius τ can be more than τ apart.



A key result of [14] is that a pseudometric d on Tame($\mathbf{R}$, $\mathbf{Vec}_K$) can be defined by

$$d(F,G) = \begin{cases} \inf\{\varepsilon \mid F \text{ and } G \text{ are } \varepsilon\text{-close}\} & \text{, if } F \text{ and } G \text{ are close for some } \varepsilon, \\ \infty & \text{, otherwise.} \end{cases}$$

Noise systems were developed as an abstract way of defining metrics in Tame($\mathbf{R}$, $\mathbf{Vec}_K$) (or generally in Tame($\mathbf{R}^n$, $\mathbf{Vec}_K$)) but for implementation

purposes they are not easy to work with. To this end so called simple noise systems were introduced in [23]. We briefly recall only the definition:

► **Definition 5.2.** A noise system $\{S_\varepsilon\}_{\varepsilon\in\mathbf{R}}$ is called **simple** if:

1. it is closed under direct sums,
2. for any finitely generated tame functor $F\colon \mathbf{R} \to \mathbf{Vec}_K$ and any τ in $\mathbf{R}$, there is a minimal element $F[\tau]$ in the set

$$B_{\subseteq}(F,\tau) = \{G \mid G \text{ is tame, } G \subseteq F \text{ and } F/G \text{ belongs to } S_\tau\}$$

with respect to inclusion $F[\tau] \subseteq G \subseteq F$.

◄

Due to monotonicity of rank, $\operatorname{rank}(F[\tau]) \leq \operatorname{rank}(F)$ for any τ in $\mathbf{R}$. In $\operatorname{Tame}(\mathbf{R}^n, \mathbf{Vec}_K)$ this does not hold and needs to be included as a requirement for a simple noise system. What makes simple noise systems interesting is that they are parametrized by persistence contours or simply **contours** [23]:

■ **Theorem 5.3.** *There is a bijection between the set of persistence contours and the set of simple noise systems.*

The current work continues the investigation of contours and their use in concrete persistence data analysis so we now turn to study them.

5.2 Contours

► **Definition 5.4.** A **contour** is a function $C : \mathbf{R}_\infty \times \mathbf{R} \to \mathbf{R}_\infty$ satisfying the following inequalities for all v, w in $\mathbf{R}_\infty$ and ε, τ in $\mathbf{R}$:

1. if $v \leq w$ and $\varepsilon \leq \tau$, then $C(v,\varepsilon) \leq C(w,\tau)$,
2. $v \leq C(v,\varepsilon)$,
3. $C(C(v,\varepsilon),\tau) \leq C(v,\varepsilon+\tau)$.

◄

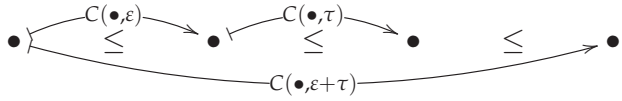
► **Definition 5.5.** A **translation** on a poset $(J,\leq)$ is a map $T : J \to J$ such that $v \leq T(v) \leq T(w)$ for every $v \leq w \in J$.

◄

As noted in [7], the set of translations on J, $\mathbf{Trans}_J$,

- is a monoid with respect to composition of translations, i.e. composition is associative and has an identity,
- is a poset with the relation $\Gamma \leq \Lambda$ iff $\Gamma(v) \leq \Lambda(v)$ for all $v \in J$, Γ and Λ being translations.

One way to characterize contour is by noting that it is an idexed set of translations of $\mathbf{R}_\infty$. The first and second property make $C(-, \varepsilon)$ a translation for a fixed ε. They also induce a poset structure on $\{C(-, \varepsilon)\}_{\varepsilon \in \mathbf{R}}$. The first condition of 5.4 also makes sure that a contour preserves the poset structure of $\mathbf{R}_\infty$. The third property makes contour to have a sub-additivity property which can be depicted in the following diagram.



Note that the constant function $C(v, \varepsilon) = k$ for any finite k cannot be a contour.

▶ **Definition 5.6.** Let C be a contour.
C is an **action** of monoid $(\mathbf{R}, +)$ on $(\mathbf{R}, +)$ if $C(v, 0) = v$ and $C(C(v, \tau), \varepsilon) = C(v, \tau + \varepsilon)$.
C is **closed** if $\{\varepsilon \in \mathbf{R} \mid C(v, \varepsilon) \geq w\}$ is closed for any $v < w$ in $\mathbf{R}_\infty$.
C is **regular** if the following conditions are satisfied:

- for any ε in $\mathbf{R}$, the function $C(-, \varepsilon) \colon \mathbf{R}_\infty \to \mathbf{R}_\infty$ is an injection,
- for any v in $\mathbf{R}$, the function $C(v, -) \colon \mathbf{R} \to [v, \infty)$ is injective on the image $[v, \infty)$.

◀

Being closed implies that the set $\{\varepsilon \in \mathbf{R} \mid C(v, \varepsilon) \geq w\}$ has a minimal element since it is a ray $[\varepsilon, \infty)$. For if $\varepsilon < \tau$, then $w \leq C(v, \varepsilon) \leq C(v, \tau)$ by contour definition 5.4. Note also that if $C(v, \varepsilon) = \infty$ for some $v < \infty$, then by the properties of contour $C(w, \tau) = \infty$ for any $w \geq v$ and $\tau \geq \varepsilon$.
Regularity of C says that for any v and ε there are unique preimages $C(v, -)^{-1}$ and $C(-, \varepsilon)^{-1}$. Regularity then implies that for any $v < w$ in $\mathbf{R}_\infty$:

$$\{\varepsilon \in \mathbf{R} \mid C(v, \varepsilon) \geq w\} = \begin{cases} \{\varepsilon \in \mathbf{R} \mid \varepsilon \geq C(v, -)^{-1}(w)\} & \text{, if } v < w < \infty \\ \emptyset & \text{, if } v < w = \infty. \end{cases}$$

All the sets on the right above are closed so regular contours are therefore closed.
For contours to be useful for producing metrics in $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ and as tools in data analysis we need methods to produce them. Definition 5.4 gives three functional inequalities implicitly characterizing contours. The last inequality however makes it difficult to find explicit functions in general. We can nonetheless make initial guesses for the functional form of a contour and then try to find a formula satisfying the requirements of Definition 5.4.

■ **Definition-Proposition 5.7.** *For (v, ε) in $\mathbf{R}_\infty \times \mathbf{R}$, consider a function of the form $C(v, \varepsilon) = f(\varepsilon)v$. If $f : \mathbf{R} \to \mathbf{R}$ is non-decreasing and $f(0) \geq 1$, then C satisfies the first two properties of Definition 5.4. The third property is equivalent to:*

$$C(C(v, \varepsilon), \tau) = C(f(\varepsilon)v, \tau) = f(\tau)f(\varepsilon)v \leq f(\varepsilon + \tau)v = C(v, \varepsilon + \tau)$$

*For instance, since $e^\tau e^\varepsilon = e^{\varepsilon+\tau}$, the function $C(v, \varepsilon) = e^\varepsilon v$ is a contour. In fact we could choose any positive base number r other than e. Such contours are called **exponential**. Exponential contours are actions.*

■ **Definition-Proposition 5.8.** *For (v, ε) in $\mathbf{R}_\infty \times \mathbf{R}$, consider a function of the form $C(v, \varepsilon) = v + f(\varepsilon)$. If f is non-decreasing, then C satisfies the first two properties of Definition 5.4. The third property gives*

$$v + f(\varepsilon) + f(\tau) \leq v + f(\varepsilon + \tau).$$

*Thus for C to be a contour, $f : \mathbf{R} \to \mathbf{R}$ should be superlinear i.e. a function satisfying $f(\varepsilon) + f(\tau) \leq f(\varepsilon + \tau)$. Contours like this are called **additive**. Additive contours can be of different types. For example $C(v, \varepsilon) = v + \varepsilon$ is called **standard contour**. Standard contours are regular actions. Another example is the **parabolic contour** $C(v, \varepsilon) = v + \varepsilon^2$. Its not an action but it is regular. In fact all additive contours are regular since superlinearity implies that f is strictly increasing.*

Contours can also be described very generally by certain integral equations. Let $f : \mathbf{R} \to \mathbf{R}_{/0}$ be a measurable function with strictly positive values referred to as **density**.

■ **Definition-Proposition 5.9.** *Let $f : \mathbf{R} \to \mathbf{R}_{/0}$ be a density function. For any v and ε in $\mathbf{R}$ there is a unique solution $C(v, \varepsilon)$ to*

$$\varepsilon = \int_v^{C(v,\varepsilon)} f(x)dx.$$

*Then $C(v,\varepsilon)$ is a contour which is an action, if we set $C(\infty,\varepsilon)=\infty$. This contour is said to be of **distance type** as it describes the distance needed to move from v to the right in order for the area under the graph of f to reach ε.*

Proof. Since f has strictly positive values, for any v and ε in $\mathbf{R}$, there is a unique number $C(v,\varepsilon)\geq v$ for which

$$\varepsilon=\int_v^{C(v,\varepsilon)} f(x)dx.$$

$v\leq C(v,\varepsilon)$: Since $0\leq\varepsilon$ we have the inequality

$$\int_v^v f(x)dx=0\leq\varepsilon=\int_v^{C(v,\varepsilon)} f(x)dx$$

Because f is strictly positive then by the properties of integration the upper limits satisfy $v\leq C(v,\varepsilon)$.

$v\leq w \implies C(v,\varepsilon)\leq C(w,\varepsilon)$:

$$\int_v^{C(v,\varepsilon)} f(x)dx=\varepsilon=\int_w^{C(w,\varepsilon)} f(x)\leq\int_v^{C(w,\varepsilon)} f(x)dx$$

and it follows that $C(v,\varepsilon)\leq C(w,\varepsilon)$.

$\varepsilon\leq\tau \implies C(v,\varepsilon)\leq C(v,\tau)$: Follows immediately from

$$\int_v^{C(v,\varepsilon)} f(x)dx=\varepsilon\leq\tau=\int_v^{C(v,\tau)} f(x)dx.$$

$C(C(v,\varepsilon),\tau)=C(v,\varepsilon+\tau)$: We have

$$\varepsilon=\int_v^{C(v,\varepsilon)} f(x)dx \text{ and } \tau=\int_{C(v,\varepsilon)}^{C(C(v,\varepsilon),\tau)} f(x)dx.$$

Then

$$\varepsilon+\tau=\int_v^{C(v,\varepsilon+\tau)} f(x)dx=\int_v^{C(v,\varepsilon)} f(x)dx+\int_{C(v,\varepsilon)}^{C(C(v,\varepsilon),\tau)} f(x)dx.$$

Thus

$$\int_v^{C(v,\varepsilon+\tau)} f(x)dx=\int_v^{C(C(v,\varepsilon),\tau)} f(x)dx$$

and the claim follows.

Moreover,

$$0=\int_v^{C(v,0)} f(x)dx$$

implies $C(v,0)=v$. ∎

■ **Definition-Proposition 5.10.** *For v in $\mathbf{R}_\infty$, choose the unique y in $\mathbf{R}_\infty$ such that $v = \int_0^y f(x)dx$. The function $C(v,\varepsilon)$ given by*

$$C(v,\varepsilon) = v + \int_y^{y+\varepsilon} f(x)dx$$

*is a contour which is an action. It is called **shift type** since it is a shift of v by the ε-step integral of the density.*

Proof. $v \leq C(v,\varepsilon)$: Since density function f is a positive function, the inequality

$$v \leq v + \int_y^{y+\varepsilon} f(x)dx$$

is true.

$v \leq w \implies C(v,\varepsilon) \leq C(w,\varepsilon)$: Now $y \leq z$ from

$$\int_0^y f(x)dx = v \leq w = \int_0^z f(x)dx.$$

Then

$$\int_0^y f(x)dx + \int_y^{y+\varepsilon} f(x)dx \leq \int_0^z f(x)dx + \int_z^{z+\varepsilon} f(x)dx.$$

$\varepsilon \leq \tau \implies C(v,\varepsilon) \leq C(v,\tau)$: The claim is evident from

$$v + \int_y^{y+\varepsilon} f(x)dx \leq v + \int_y^{y+\tau} f(x)dx.$$

$C(C(v,\varepsilon),\tau) = C(v,\varepsilon+\tau)$: The claim follows from the sequence of equalities

$$\begin{aligned} C(C(v,\varepsilon),\tau) &= C(\int_0^{y+\varepsilon} f(x)dx, \tau) = \int_0^{y+\varepsilon} f(x)dx + \int_{y+\varepsilon}^{y+\varepsilon+\tau} f(x)dx \\ &= \int_0^y f(x)dx + \int_y^{y+\varepsilon+\tau} f(x)dx = C(v,\varepsilon+\tau). \end{aligned}$$

Clearly

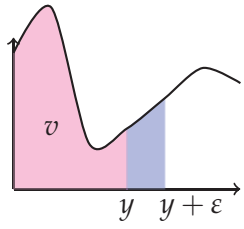
$$C(v,0) = v + \int_y^y f(x)dx = v.$$

■

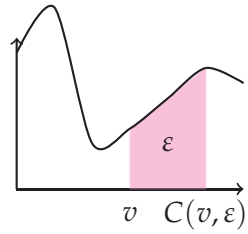
If density is chosen to be constant function $f(x) = 1$, then distance and shift contours retrieve standard contour:

$$\varepsilon = \int_v^{C(v,\varepsilon)} dx = C(v,\varepsilon) - v \implies C(v,\varepsilon) = v + \varepsilon,$$
$$v = \int_0^v dx \implies C(v,\varepsilon) = \int_0^{v+\varepsilon} dx = v + \varepsilon.$$

Note that both distance and shift contours are regular. In the figure below we illustrate how shift type (left) and distance type (right) contours are obtained by integrating the density function:



shift type



distance type

Exploration and visualization of contours is deferred to Chapter 8 where we use them in actual data analysis. We conclude this section by describing a way to cut, or truncate, a contour at some specified value.

■ **Definition-Proposition 5.11.** *Let $C : \mathbf{R}_\infty \times \mathbf{R} \to \mathbf{R}_\infty$ be a contour. Choose an element α in $\mathbf{R}_\infty$. For (v,ε) in $\mathbf{R}_\infty \times \mathbf{R}$ define:*

$$C_{/\alpha}(v,\varepsilon) = \begin{cases} C(v,\varepsilon) & \text{, if } C(v,\varepsilon) < \alpha, \\ \infty & \text{, if } C(v,\varepsilon) \geq \alpha; \end{cases}$$

$$C_{//\alpha}(v,\varepsilon) = \begin{cases} \infty & \text{, if } v = \infty, \\ C(v,\varepsilon) & \text{, if } v < \infty \text{ and } v + \alpha > C(v,\varepsilon), \\ \infty & \text{, if } v < \infty \text{ and } v + \alpha \leq C(v,\varepsilon). \end{cases}$$

*$C_{/\alpha}$ is called the **truncation** of C at α and $C_{//\alpha}$ is called the **translational truncation** of C by α. They are both contours.*

Proof. Fix $\alpha < \infty$. Let us see $C_{/\alpha}$ first. If $C(v,\varepsilon) < \alpha$, then for $w \geq v$ and $\tau \geq \varepsilon$, $C_{/\alpha}(v,\varepsilon) = C(v,\varepsilon) \leq C(w,\tau) = C_{/\alpha}(w,\tau)$ or $C_{/\alpha}(w,\tau) = \infty$. In either case the first property of Definition 5.4 is true. If $C(v,\varepsilon) \geq \alpha$ its

also true since all subsequent values of $C_{/\alpha}$ are ∞. The second property of Definition 5.4 is obvious by definition.

For the third property we need to show $C_{/\alpha}\big(C_{/\alpha}(v,\varepsilon),\tau\big) \le C_{/\alpha}(v,\varepsilon+\tau)$. This is clear if $C(v,\varepsilon+\tau) \ge \alpha$. Assume $C(v,\varepsilon+\tau) < \alpha$. Then also $C(v,\varepsilon) < \alpha$ and $C(C(v,\varepsilon),\tau) < \alpha$. Consequently $C_{/\alpha}\big(C_{/\alpha}(v,\varepsilon),\tau\big) = C(C(v,\varepsilon),\tau)$ and $C_{/\alpha}(v,\varepsilon+\tau) = C(v,\varepsilon+\tau)$ and the desired inequality follows from C being a contour.

Similarly as above, the first and second properties of contour are true for $C_{//\alpha}$. The inequality $C_{//\alpha}\big(C_{//\alpha}(v,\varepsilon),\tau\big) \le C_{//\alpha}(v,\varepsilon+\tau)$ is clear if $v=\infty$ or $v<\infty$ and $v+\alpha \le C(v,\varepsilon+\tau)$. Assume $v<\infty$ and $v+\alpha > C(v,\varepsilon+\tau)$. This implies

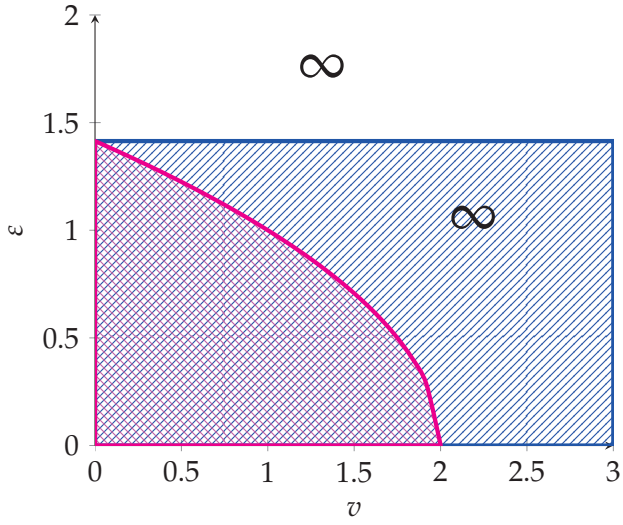
$$\begin{aligned} C(v,\varepsilon) &\le C(v,\varepsilon+\tau) < v+\alpha, \\ C(C(v,\varepsilon),\tau) &\le C(v,\varepsilon+\tau) < v+\alpha \le C(v,\varepsilon)+\alpha. \end{aligned}$$

Therefore we have $C_{//\alpha}\big(C_{//\alpha}(v,\varepsilon),\tau\big) = C_{//\alpha}\big(C(v,\varepsilon),\tau\big) = C(C(v,\varepsilon),\tau)$ and $C_{//\alpha}(v,\varepsilon+\tau) = C(v,\varepsilon+\tau)$ and the desired inequality follows again from C being a contour. ■

Neither of the truncated contours is regular. If C is closed, then so are both of its truncations for any α in $\mathbf{R}_\infty$. Observe that $C_{/0} = C_{//0} = \infty$ and $C_{/\infty} = C_{//\infty} = C$. Truncation reverses the ordering between contours. If $\alpha \le \beta$ in $\mathbf{R}_\infty$, then for any (v,ε) in $\mathbf{R}_\infty \times \mathbf{R}$:

$$\begin{aligned} \infty = C_{/0}(v,\varepsilon) \ge C_{/\alpha}(v,\varepsilon) \ge C_{/\beta}(v,\varepsilon) \ge C_{/\infty}(v,\varepsilon) = C(v,\varepsilon), \\ \infty = C_{//0}(v,\varepsilon) \ge C_{//\alpha}(v,\varepsilon) \ge C_{//\beta}(v,\varepsilon) \ge C_{//\infty}(v,\varepsilon) = C(v,\varepsilon). \end{aligned}$$

▼ **Example 5.12.**
Choose α to be 2. Let us study truncations of the parabolic contour $C(v,\varepsilon) = v+\varepsilon^2$.
For $C_{/2}$ we have that its equal to C if $v+\varepsilon^2 < 2$ or $v < \sqrt{2-\varepsilon^2}$. Outside this region $C_{/2} = \infty$. This is illustrated below in magenta.
For $C_{//2}$ we have that its equal to C if $v+2 > v+\varepsilon^2$ or $\varepsilon < \sqrt{2}$. Outside this region $C_{//2} = \infty$. This is illustrated below in blue.



▲

Contours appeared independently in [7] under the name superlinear families. Let J be a poset and $\mathbf{Trans}_J$ be its set of translations. A superlinear family is a function $\Omega : \mathbf{R} \to \mathbf{Trans}_J$, satisfying $\Omega(\varepsilon_2) \circ \Omega(\varepsilon_1) \leq \Omega(\varepsilon_1 + \varepsilon_2)$, whenever $\varepsilon_1, \varepsilon_2 \geq 0$. In [7] persistence theory was extended to diagrams indexed by J and constructed by possibly other tools than homology, thus taking values in any appropriate category $\mathbf{D}$. Resulting persistence modules were called generalized persistence modules and superlinear families were then used to define abstract interleavings between them. Since then generalized persistence modules have gathered some interest, see [38] and [32]. In our view, contours lead to a rich space of pseudometrics on $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$. We have also provided general means to construct contours, greatly enlarging [7], in which the authors only give the standard contour as a concrete example of a superlinear family.

5.3 Contours and pseudometrics

We conclude this chapter by showing how contours define pseudometrics on $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$. Note the similarities to noise systems in Section 5.1. With general noise systems, however, the notion of ε-equivalence and ε-equivalence with respect to a contour C are different concepts. Simple noise systems, though, are parameterized by contours. Initial definition and results in Definition 5.13, Proposition 5.14 and Proposition 5.15 are acknowledged to O. Gäfvert and reproduced here, see also [23].

► **Definition 5.13.** Let $C\colon \mathbf{R}_\infty \times \mathbf{R} \to \mathbf{R}_\infty$ be a contour, V and W be tame persistent vector spaces and $\varepsilon \in \mathbf{R}$.

1. A natural transformation $f\colon V \to W$ is called an ***ε*-equivalence with respect to *C*** if, for any t in $\mathbf{R}$ such that $C(t,\varepsilon) < \infty$, there exists a **lift**, the indicated dotted linear map making the following diagram commutative:

$$\begin{array}{ccc} V_t & \xrightarrow{V_{t\le C(t,\varepsilon)}} & V_{C(t,\varepsilon)} \\ {\scriptstyle f_t}\downarrow & \nearrow & \downarrow{\scriptstyle f_{C(t,\varepsilon)}} \\ W_t & \xrightarrow[W_{t\le C(t,\varepsilon)}]{} & W_{C(t,\varepsilon)} \end{array}$$

2. V and W are ***ε*-equivalent with respect to *C*** if there is $X \in \mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ and maps

$$V \xrightarrow{f} X \xleftarrow{g} W$$

such that f is a τ-equivalence with respect to C, g is a μ-equivalence with respect to C and $\tau + \mu \le \varepsilon$.

◄

For more fluent text we shorten ε-equivalence with respect to C and ε-equivalent with respect to C simply to ε-equivalence and ε-equivalent, the choice of contour C always being implicitly included.
Note that if f is ε-equivalence then it is τ-equivalence for any $\tau > \varepsilon$ such that $C(t,\tau) < \infty$. The required lift is the composition of the ε-equivalence lift and $V_{C(t,\varepsilon)\le C(t,\tau)}$ in the following diagram:

$$\begin{array}{ccccc} V_t & \xrightarrow{V_{t\le C(t,\varepsilon)}} & V_{C(t,\varepsilon)} & \xrightarrow{V_{C(t,\varepsilon)\le C(t,\tau)}} & V_{C(t,\tau)} \\ {\scriptstyle f_t}\downarrow & \nearrow & \downarrow{\scriptstyle f_{C(t,\varepsilon)}} & & \downarrow{\scriptstyle f_{C(t,\tau)}} \\ W_t & \xrightarrow[W_{t\le C(t,\varepsilon)}]{} & W_{C(t,\varepsilon)} & \xrightarrow[W_{C(t,\varepsilon)\le C(t,\tau)}]{} & W_{C(t,\tau)} \end{array}$$

The following Proposition 5.14 states some basic properties of equivalences.

■ **Proposition 5.14.** *Let C be a contour and U, V and W in Tame($\mathbf{R}$, $\mathbf{Vec}_K$).*

1. *If $g\colon U \to V$ is a ε-equivalence and $f\colon V \to W$ is an τ-equivalence, then the composition $fg\colon U \to W$ is a $(\tau+\varepsilon)$-equivalence.*

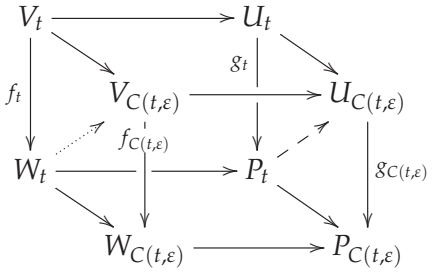
2. *In the following pushout square P is tame and if f is an ε-equivalence, then so is g:*

$$\begin{array}{ccc} V & \longrightarrow & U \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\ W & \longrightarrow & P \end{array}$$

Proof. (1): If $C(t, \tau+\varepsilon) < \infty$, then $C(t,\tau) \leq C(C(t,\tau),\varepsilon) \leq C(t,\tau+\varepsilon) < \infty$. Since f and g are τ-equivalence and ε-equivalence, respectively, we have the following commutative diagram whose diagonal morphism $W_t \to U_{C(t,\tau+\varepsilon)}$ is a lift required for fg to be a $(\tau+\varepsilon)$-equivalence:

$$\begin{array}{ccccccc} U_t & \longrightarrow & U_{C(t,\tau)} & \longrightarrow & U_{C(C(t,\tau),\varepsilon)} & \longrightarrow & U_{C(t,\tau+\varepsilon)} \\ {\scriptstyle g_t}\downarrow & & \downarrow & \nearrow & \downarrow & & \downarrow \\ V_t & \longrightarrow & V_{C(t,\tau)} & \longrightarrow & V_{C(C(t,\tau),\varepsilon)} & \longrightarrow & V_{C(t,\tau+\varepsilon)} \\ {\scriptstyle f_t}\downarrow & \nearrow & \downarrow & & \downarrow & & \downarrow \\ W_t & \longrightarrow & W_{C(t,\tau)} & \longrightarrow & W_{C(C(t,\tau),\varepsilon)} & \longrightarrow & W_{C(t,\tau+\varepsilon)} \end{array}$$

(2): Tameness of P follows from being a pushout, see Section 3.3. If $C(t,\varepsilon) < \infty$, the following commutative diagram can be formed where the dotted morphism is given by f being an ε-equivalence and the dashed arrow comes from the universal property of the back square being a pushout:



■

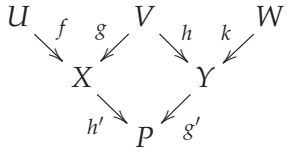
The properties stated in Proposition 5.14 allow us to prove:

■ **Proposition 5.15.** *Let $V, W \in Tame(\mathbf{R}, \boldsymbol{Vec}_K)$ and $S = \{\varepsilon \in \mathbf{R} \mid V$ and W are ε-equivalent with respect to $C\}$. Define*

$$d_C(V,W) = \begin{cases} \infty & , \text{if } S = \emptyset, \\ inf(S) & , \text{if } S \neq \emptyset. \end{cases}$$

Then d_C is a pseudometric on $Tame(\mathbf{R}, \boldsymbol{Vec}_K)$.

Proof. $d_C(V,V) = 0$ and symmetry is also clear. For the triangle inequality let U and V be ε-equivalent and V and W be τ-equivalent. Then in the diagram



f is an ε_1-equivalence, g is an ε_2-equivalence, $\varepsilon_1 + \varepsilon_2 \leq \varepsilon$ and h is a τ_1-equivalence, k is a τ_2-equivalence, $\tau_1 + \tau_2 \leq \tau$.

The central square in this diagram is a pushout. Then by Proposition 5.14, g' is an ε_2-equivalence, h' is a τ_1-equivalence, $h'f$ is a $(\tau_1 + \varepsilon_1)$-equivalence and $g'k$ is a $(\tau_2 + \varepsilon_2)$-equivalence.

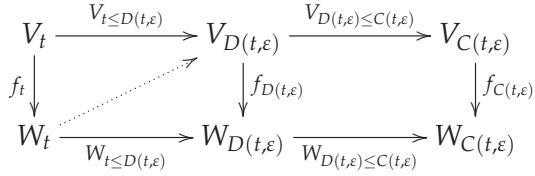
Since $\varepsilon + \tau \geq \tau_1 + \varepsilon_1 + \tau_2 + \varepsilon_2$, we can conclude that U and W are $(\varepsilon + \tau)$-equivalent and the triangle inequality $d_C(U,W) \leq d_C(U,V) + d_C(V,W)$ follows. ■

In Chapter 7 we need sequences of pseudometrics on $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$, particularly non-decreasing sequences. The following proposition gives us means to do that.

■ **Proposition 5.16.** *Assume C and D are contours such that $C \geq D$. Then*

1. *An ε-equivalence with respect to D implies ε-equivalence with respect to C.*
2. $d_C(V,W) \leq d_D(V,W)$.

Proof. (1): Let $f\colon V \to W$ be an ε-equivalence with respect to D. If $C(t,\varepsilon) < \infty$, then also $D(t,\varepsilon) < \infty$ and we can form the following commutative diagram whose diagonal is the lift assuring f is an ε-equivalence with respect to C:



(2): In the diagram below let f and g be τ- and μ-equivalences with respect to D, respectively, and $\tau + \mu \leq \varepsilon$:

$$V \xrightarrow{f} X \xleftarrow{g} W$$

By (1) this implies that V and W are ε-equivalent with respect to C and the claim follows. ■

We finish by gathering some properties of ε-equivalence and d_C in the next proposition.

■ **Proposition 5.17.** *Let V and W be in Tame($\mathbf{R}$, $\boldsymbol{Vec}_K$) and let C be a contour.*

1. *If $C = \infty$, then any map $f\colon V \to W$ is a 0-equivalence and $d_C(V, W) = 0$.*

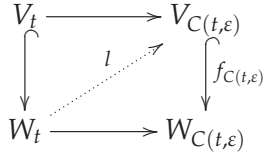
2. *An injective $f\colon V \to W$ is an ε-equivalence with respect to C if and only if $im(W_{t\leq C(t,\varepsilon)}\colon W_t \to W_{C(t,\varepsilon)}) \subseteq im(f_{C(t,\varepsilon)}\colon V_{C(t,\varepsilon)} \to W_{C(t,\varepsilon)})$ for any t in $\mathbf{R}$ such that $C(t,\varepsilon) < \infty$.*

3. *$0 \to W$ is an ε-equivalence if and only if $W_{t\leq C(t,\varepsilon)}\colon W_t \to W_{C(t,\varepsilon)}$ is the zero map for any t in $\mathbf{R}$ such that $C(t,\varepsilon) < \infty$.*

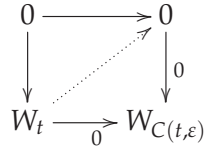
4. *W is ε-equivalent to 0 if and only if $0 \to W$ is an ε-equivalence.*

5. *$d_C(0, W) < \varepsilon$ if and only if $W_{t\leq C(t,\varepsilon)}\colon W_t \to W_{C(t,\varepsilon)}$ is the zero map for any t in $\mathbf{R}$ such that $C(t,\varepsilon) < \infty$.*

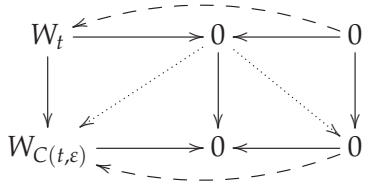
Proof. (1): If $C = \infty$, by Definition 5.13 there is no t such that $C(t,\varepsilon) < \infty$ and any map can be said to be 0-equivalence.
(2): Since $f\colon V \to W$ is an ε-equivalence we have the commuting diagram



The first implication follows from $\operatorname{im} W_{t\leq C(t,\varepsilon)} = \operatorname{im} f_{C(t,\varepsilon)} \circ l \subseteq \operatorname{im} f_{C(t,\varepsilon)}$. In the other direction, for any x in W_t we have a unique $f^{-1}_{C(t,\varepsilon)}(W_{t\leq C(t,\varepsilon)}(x))$ since $f_{C(t,\varepsilon)}$ is injective and $\operatorname{im} W_{t\leq C(t,\varepsilon)} \subseteq \operatorname{im} f_{C(t,\varepsilon)}$. We can then define a lift $x \mapsto f^{-1}_{C(t,\varepsilon)}(W_{t\leq C(t,\varepsilon)}(x))$.
(3): The claim follows straightforwardly from the diagram



(4): There is an ε-equivalence $0 \to 0$ for any $\varepsilon \geq 0$. We can thus construct the inside squares of following diagram where W is ε-equivalent to 0:



By the commutativity of the triangles in the left square, $W_t \to W_{C(t,\varepsilon)}$ is the zero map and $0 \to W$ is an ε-equivalence by (3). In the other direction we have a map $0 \to W$ given by the dashed arrows, which is an ε-equivalence with a lift given by the composition of $W_t \to 0$ with the dotted arrow $0 \to 0$.

(5): Since $d_C(0, W) = \inf\{\varepsilon \,|\, 0 \text{ and } W \text{ are } \varepsilon\text{-equivalent}\} < 0$ the claim follows immediately from (4) and (3). ■

It is not true in general that if $\varepsilon = d_C(0, W) < \infty$, then $W_t \to W_{C(t,\varepsilon)}$ is the zero map for any t in $\mathbf{R}$ such that $C(t, \varepsilon) < \infty$. Consider C that is an action, i.e. $C(t, 0) = t$ and assume $d_C(0, W) = 0$. Then the composition with the lift $W_t \to 0 \to W_{C(t,0)}$ is the zero map but $W_t \to W_{C(t,0)}$ is the identity.

Chapter 6

Stable rank invariant

After defining a class of metrics in $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ we are now ready to stabilize the rank invariant of a tame persistent vector space. The fundamental computational property of stable rank in one dimensional persistence turns out to be that it is the number of bar generators satisfying a relation imposed by the metric coming from the chosen contour. This gives an efficient algorithmic way of computing the stabilization of rank.

▶ **Definition 6.1.** Let C be a contour and d_C be the associated pseudometric on $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$. **Stable rank** is the metric stabilization of the rank function $\mathrm{rank}\colon \mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K) \to \mathbf{N}$ with respect to d_C and is denoted by:

$$\widehat{\mathrm{rank}}_C\colon \mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K) \to \mathscr{M}.$$

Explicitly

$$\widehat{\mathrm{rank}}_C V(t) = \min\left\{\mathrm{rank}(W) \mid W \in \mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K) \text{ and } d_C(V, W) \leq t\right\}. \quad \blacktriangleleft$$

The function $\widehat{\mathrm{rank}}_C V$ is non-increasing with values in $\mathbf{N}$. There are thus finitely many elements $0 < t_0 < \cdots < t_n$ in the domain $\mathbf{R}$ such that $\widehat{\mathrm{rank}}_C V$ is constant on the open intervals $(0, t_0), \ldots, (t_i, t_{i+1}), \ldots, (t_n, \infty)$. In general, depending on the contour type, $\widehat{\mathrm{rank}}_C V$ may fail however to be right or left continuous, i.e. the intervals containing left or right endpoints. The method in this chapter follows one used in [23] and is based on the following definition:

▶ **Definition 6.2.** Let C be a contour, $\delta \in \mathbf{R}$ and $V \in \mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$. The **$\delta$-shift** of V with respect to C, denoted by $V_C[\delta]$, is the subspace of V

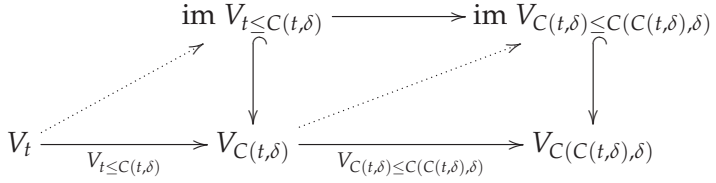
generated by all the elements in the images of the transition functions $V_{t\leq C(t,\delta)} \colon V_t \to V_{C(t,\delta)}$ for all t in $\mathbf{R}$ such that $C(t,\delta) < \infty$. ◀

Due to the composition property of transition maps, $V_{t\leq u} \circ V_{s\leq t} = V_{s\leq u}$, the image of $V_{s\leq u}$ is contained in the image of $V_{t\leq u}$. The shift operation has the following properties:

■ **Proposition 6.3.** *Let C be a contour, V and W be in Tame($\mathbf{R}$, $\boldsymbol{Vec}_K$) and $\delta \in \mathbf{R}$.*

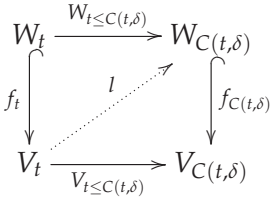
1. *If V is generated by $\{g_i \in V_{s(i)}\}_{i\in\bar{n}}$, then $V_C[\delta]$ is generated by:*
$$\{V_{s(i)\leq C(s(i),\delta)}(g_i) \mid i \in \bar{n} \text{ and } C(s(i),\delta) < \infty\}.$$
2. *$V_C[\delta]$ is tame.*
3. *The inclusion $V_C[\delta] \subset V$ is a δ-equivalence with respect to C.*
4. *An injective $f \colon W \subset V$ is a δ-equivalence if and only if $V_C[\delta] \subseteq \mathrm{im}\, f$.*
5. *The shift is linear: $(V \oplus W)_C[\delta]$ and $V_C[\delta] \oplus W_C[\delta]$ are isomorphic.*
6. *$\left(\bigoplus_{i\in\bar{n}} K(s(i),e(i))\right)_C[\delta] \cong \bigoplus_{\{i \mid C(s(i),\delta)<e(i)\}} K(C(s(i),\delta),e(i))$.*
7. *$\mathrm{rank}\left(\left(\bigoplus_{i\in\bar{n}} K(s(i),e(i))\right)_C[\delta]\right) = |\{i \mid C(s(i),\delta) < e(i)\}|$.*

Proof. (1): This is a direct consequence of the definition of δ-shift.
(2): Being a finitely generated subspace of tame persistent vector space, $V_C[\delta]$ is tame (Proposition 3.34).
(3): This follows from the diagram

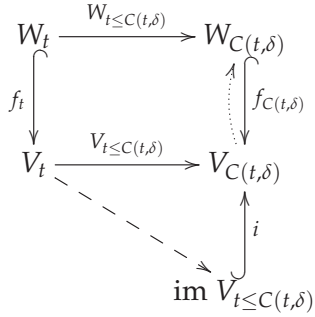


since the lifts are the surjections on the images of the maps on lower row.

(4): When f is a δ-equivalence we have the commutative diagram with lift l:



Hence $V_{t\leq C(t,\delta)} = f_{C(t,\delta)} \circ l$ and im $V_{t\leq C(t,\delta)} \subseteq$ im $f_{C(t,\delta)}$.
In the other direction consider diagram



Since f is injective we have the dotted inverse map from im $f_{C(t,\delta)}$ to $W_{t\leq C(t,\delta)}$. This composed with the dashed map and inclusion i provide the needed lift.

(5): Images of the transition maps $(V \oplus W)_{t\leq C(t,\delta)}$ generate $(V \oplus W)_C[\delta]$. By Definition 3.30 the transition maps in a direct sum are $V_{t\leq C(t,\delta)} \oplus W_{t\leq C(t,\delta)}$ and $\text{im}(V \oplus W)_{t\leq C(t,\delta)} \cong \text{im } V_{t\leq C(t,\delta)} \oplus \text{im } W_{t\leq C(t,\delta)}$.

(6): By (5), $\left(\bigoplus_{i\in\bar{n}} K(s(i),e(i))\right)_C[\delta] \cong \bigoplus_{i\in\bar{n}} K(s(i),e(i))_C[\delta]$. For a single $K(s,e)$, $K(s,e)_C[\delta]$ is generated by the images of $K(s,e)_{s\leq C(s,\delta)}$ and these are non-zero only if $C(s,\delta) < e$. It follows that $K(s,e)_C[\delta] = K(C(s,\delta),e)$.

(7): This follows directly from (6) and Proposition 3.40. ■

If $\delta \leq \delta'$ in $\mathbf{R}$, then $V_{C(t,\delta)\leq C(t,\delta')} \circ V_{t\leq C(t,\delta)} = V_{t\leq C(t,\delta')}$. Then im $V_{t\leq C(t,\delta')}$ is generated by the image of im $V_{t\leq C(t,\delta)}$ under $V_{C(t,\delta)\leq C(t,\delta')}$. Consequently $V_C[\delta'] \subset V_C[\delta]$ and it then follows from the monotonicity of rank that $\text{rank}(V_C[\delta']) \leq \text{rank}(V_C[\delta])$. Therefore the rank function of the shift defined by $\delta \mapsto \text{rank}(V_C[\delta])$ is non-increasing with values in $\mathbf{N}$. As with the stable rank $\widehat{\text{rank}}_C V$, there are finitely many elements $0 < \delta_0 < \cdots < \delta_n$ in $\mathbf{R}$ such that $\text{rank}(V_C[-])$ is constant on the open intervals $(0,\delta_0)$,..., (δ_i,δ_{i+1}),..., (δ_n,∞). The following proposition establishes when the intervals contain their left endpoints.

■ **Proposition 6.4.** *Let V in Tame$(\mathbf{R}, \mathbf{Vec}_K)$. If C is a closed contour, then rank$(V_C[-])$ is a right continuous function, i.e. rank$(V_C[-])$ is constant on the intervals $[0,\delta_0),\ldots,[\delta_i,\delta_{i+1}),\ldots,[\delta_n,\infty)$.*

Proof. Let $0 < t_0 < \cdots < t_k$ be such that, for any $a \leq b$ in $\mathbf{R}$, the map $V_{a\leq b}\colon V_a \to V_b$ is not an isomorphism only if $a < t_i \leq b$. Let

$\{g_i \in V_{s(i)}\}_{1\in\bar{n}}$ be a set of generators of V. Consider only those for which $C(s(i),\delta) < \infty$ for $\delta \in \mathbf{R}$.

Since C is closed, the set $\{\varepsilon \in \mathbf{R} \mid C(s(i),\varepsilon) \geq t_{i+1}\}$ for $s(i) \in [t_i, t_{i+1})$ is a ray $[\varepsilon, \infty)$. Hence for all $s(i)$ there is $\delta \leq \delta'$ such that both $C(s(i),\delta)$ and $C(s(i),\delta')$ are in one of the intervals $[0,t_0),\ldots,[t_n,\infty)$. Consequently the transition functions $V_{C(s(i),\delta)\leq C(s(i),\delta')}$ are isomorphisms and $V_C[\delta]$ and $V_C[\delta']$ have the same rank. ■

The following theorem is the key relation between the stable rank and its shift establishing that the values of stable rank are bounded between the ranks of the shifts.

■ **Theorem 6.5.** *Let V be in Tame($\mathbf{R}$, $\mathbf{Vec}_K$) and let C be a contour. Then for any $\delta < \delta'$ in $\mathbf{R}$:*

$$rank(V_C[\delta]) \geq \widehat{rank}_C V(\delta) \geq rank(V_C[\delta']).$$

Proof. By Proposition 6.3 $V_C[\delta] \subset V$ is a δ-equivalence. Then $V_C[\delta]$ and V are δ-equivalent and $d_C(V, V_C[\delta]) \leq \delta$. Since stable rank of V at δ is the minimum rank in its δ-neighbourhood, this gives the first inequality.

Let W be in Tame($\mathbf{R}$, $\mathbf{Vec}_K$) such that $d_C(V,W) \leq \delta$ and $\mathrm{rank}(W) = \widehat{\mathrm{rank}}_C V(\delta)$. Since V are W are δ-equivalent, they are δ'-equivalent and by definition there are τ_1-equivalence f and τ_2-equivalence g, $\tau_1 + \tau_2 \leq \delta'$, in the diagram

$$V \xrightarrow{f} X \xleftarrow{g} W$$

For any t in $\mathbf{R}$ such that $C(t,\tau_1) < \infty$ we have the following commutative diagram where the vertical arrows indicate the transition functions and β_t is the map that comes from f being a τ_1-equivalence:

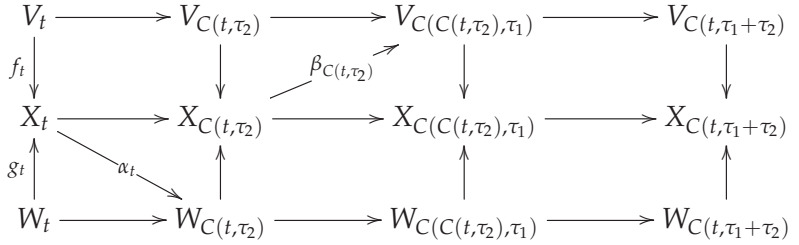
$$\begin{array}{ccccc} V_t & \xrightarrow{f} & X_t & \xleftarrow{g} & W_t \\ \downarrow & \overset{\beta_t}{\swarrow} & \downarrow & & \downarrow \\ V_{C(t,\tau_1)} & \xrightarrow[f]{} & X_{C(t,\tau_1)} & \xleftarrow[g]{} & W_{C(t,\tau_1)} \end{array}$$

Let $n = \mathrm{rank}(W)$ and $\{w_i \in W_{s(i)}\}_{i\in\bar{n}}$ be a smallest set of generators of W. Collect the indices i of those generators that the contour $C(-,\tau_1)$ does not

map to infinity, look at their images in $V_{C(s(i),\tau_1)}$ and define the subspace of V generated by these images:

$$\begin{aligned} I &= \{i \mid C(s(i), \tau_1) < \infty\}, \\ v_i &= \beta_{s(i)}(g(w_i)) \in V_{C(s(i),\tau_1)},\ i \in I, \\ V' &= \langle v_i \rangle \subset V. \end{aligned}$$

Since $\tau_1 + \tau_2 \leq \delta'$ we have that $\text{rank}(V_C[\tau_1 + \tau_2]) \geq \text{rank}(V_C[\delta'])$. We also have that $\widehat{\text{rank}}_C V(\delta) = \text{rank}(W) \geq \text{rank}(V')$ since V' is generated by a subset of generators of W (Proposition 3.40). Likewise we show that $\text{rank}(V') \geq \text{rank}(V_C[\tau_1 + \tau_2])$ by showing the inclusion $V_C[\tau_1 + \tau_2] \subset V'$. For any t in $\mathbf{R}$ such that $C(t, \tau_1 + \tau_2) < \infty$ we have the following commutative diagram where all the horizontal arrows are the transition functions, vertical arrows come from V and W being δ'-equivalent and α_t and $\beta_{C(t,\tau_2)}$ come from f being a τ_1-equivalence and g being a τ_2-equivalence, respectively:



Since the diagram commutes for any such t, the image of the transition function $V_{t \leq C(t,\tau_1+\tau_2)}$ belongs to V' and since $V_C[\tau_1 + \tau_2]$ is generated by these images, we have that $V_C[\tau_1 + \tau_2] \subset V'$. ■

The function $\text{rank}(V_C[-])$ is constant on some set of open intervals. Moreover from Theorem 6.5 we have for any $\varepsilon > 0$ that $t < t + \varepsilon$ and therefore

$$\begin{aligned} \text{rank}(V_C[t]) &\geq \widehat{\text{rank}}_C V(t + \varepsilon), \\ \widehat{\text{rank}}_C V(t) &\geq \text{rank}(V_C[t + \varepsilon]). \end{aligned}$$

Then, by recalling from the proof of Theorem 4.3 the relation between $d_{\bowtie}$ and L_p we have a direct consequence of Theorem 6.5:

■ **Corollary 6.6.** *Let C be a contour and V in Tame($\mathbf{R}$, $\mathbf{Vec}_K$). Then there are elements $0 < \tau_0 < \cdots < \tau_n$ in $\mathbf{R}$ such that the functions $\widehat{\text{rank}}_C V$ and*

rank$(V_C[-])$ *agree on the open intervals* $(0,\tau_0),\ldots,(\tau_i,\tau_{i+1}),\ldots,(\tau_n,\infty)$. *In particular, for any* $1 \leq p < \infty$*:*

$$d_{\bowtie}\left(\widehat{rank}_C V, rank(V_C[-])\right) = 0 = L_p\left(\widehat{rank}_C V, rank(V_C[-])\right).$$

Proposition 6.4 says that for a closed contour the intervals of Corollary 6.6 are left closed so together with 6.5 we get

■ **Corollary 6.7.** *For* V *in Tame*$(\mathbf{R}, \boldsymbol{Vec}_K)$ *and a closed contour* C

$$\widehat{rank}_C V = rank(V_C[-]).$$

We can now state our main theorem for this chapter. It gives the fundamental properties of stable rank and shows how it can be calculated in one-dimensional persistence from the chosen contour.

■ **Theorem 6.8.** *If* C *is a closed contour, then* $\widehat{rank}_C$: *Tame*$(\mathbf{R}, \boldsymbol{Vec}_K) \to \mathscr{M}$ *is a right continuous function satisfying the following properties:*

1. $\widehat{rank}_C$ *is linear: for any* V *and* W *in Tame*$(\mathbf{R}, \boldsymbol{Vec}_K)$,

$$\widehat{rank}_C(V \oplus W) = \widehat{rank}_C V + \widehat{rank}_C W.$$

2. $\widehat{rank}_C \bigoplus_{i\in\bar{n}} K(s(i),e(i))(t) = |\{i \mid C(s(i),t) < e(i)\}|$.

If also C *is such that* $C(t,0) = t$ *for any* t *in* $\mathbf{R}$*, then for* V *in Tame*$(\mathbf{R}, \boldsymbol{Vec}_K)$*:*

3. $\widehat{rank}_C V(0) = rank(V)$.
4. $\widehat{rank}_C V = 0$ *if and only if* $V = 0$.

Proof. (1) and (2): These are direct consequences of Corollary 6.7 and Proposition 6.3

(3): Again directly from Corollary 6.7 and the first property of Proposition 6.3 we have $\widehat{\text{rank}}_C V(0) = \text{rank}(V_C[0]) = \text{rank}(V)$.

(4): Since $\widehat{\text{rank}}_C V$ is a non-increasing function with values in $\mathbf{N}$, the equality $\widehat{\text{rank}}_C V = 0$ is equivalent to $\widehat{\text{rank}}_C V(0) = 0 = \text{rank}(V)$, which is equivalent to $V = 0$ by Proposition 3.40. ■

We have now developed necessary background for computing the stabilization of the rank invariant for V in Tame$(\mathbf{R}, \mathbf{Vec}_K)$. Theorem 6.8 gives that the stable rank of V is the number of its bar generators modified by the contour. Different contours produce different stable ranks and we explore this computationally in Chapter 8.

▼ **Example 6.9.**
Consider bars $K(2,6)$ and $K(3,7)$ in Tame($\mathbf{R}$, $\mathbf{Vec}_K$). Let us choose contour $C(v,t) = v + 2t$. By Theorem 6.8

$$\widehat{\text{rank}}_C K(2,6)(t) = 1, \text{ if } 2 + 2t < 6,$$
$$\widehat{\text{rank}}_C K(3,7)(t) = 1, \text{ if } 3 + 2t < 7.$$

Even though the two bars are not isomorphic, or don't even have a non-trivial map between them, they have the same stable rank function given by

$$\widehat{\text{rank}}_C(t) = \begin{cases} 1 & \text{, if } t < 2, \\ 0 & \text{, if } t \geq 2. \end{cases}$$

▲

Chapter 7

Embedding theorem

For the last theoretical development we return to metric stabilization by sequences of pseudometrics. In general the stabilization $\widehat{I}_d$ of a discrete invariant I may map objects that we do not intend to identify to the same function and there is a loss of information. This was illustrated for bar functors and stable rank invariant in Example 6.9 at the end of Chapter 6. To retain more information we consider sequences of pseudometrics on T and the induced stabilizations over this sequence. It turns that the metric stabilization is also stable with respect to this sequence. In the end we are able to prove that tame persistent vector spaces embed into the space $\mathscr{M}_2$ of Lebesgue measurable functions of the form $\mathbf{R}^2 \to \mathbf{R}$.

7.1 Metric stabilization by sequences of pseudometrics

▶ **Definition 7.1.** Let $I\colon T \to \mathbf{N}$ be a discrete invariant and $\{d_\alpha\}_{\alpha\in\mathbf{R}_\infty}$ a sequence of pseudometrics on T.

1. $\{d_\alpha\}_{\alpha\in\mathbf{R}_\infty}$ is called **non-decreasing** if for any $\alpha < \beta$ in $\mathbf{R}$ and X, Y in T:
$$d_\alpha(X,Y) \leq d_\beta(X,Y).$$
2. For X in T, $\overline{I}(X)\colon \mathbf{R}^2 \to \mathbf{R}$ is a function defined as follows:
$$\overline{I}(X)(\alpha,t) = \widehat{I}_{d_\alpha}(X)(t)$$
3. $\{d_\alpha\}_{\alpha\in\mathbf{R}_\infty}$ is **admissible** for I if the function $\overline{I}(X)\colon \mathbf{R}^2 \to \mathbf{R}$ is Lebesgue measurable for any X in T.

4. Let $\{d_\alpha\}_{\alpha\in\mathbf{R}_\infty}$ be a sequence of pseudometrics on T admissible for I. Then the function:

$$\overline{I}\colon T \to \mathscr{M}_2,$$
$$\overline{I}(X)(\alpha,t) = \min\{I(Y) \mid Y \in T,\ d_\alpha(X,Y) \leq t\}$$

is called the **metric stabilization of** I **along the sequence** $\{d_\alpha\}_{\alpha\in\mathbf{R}_\infty}$.

◀

Recall from Proposition 5.16 that an ordered sequence of contours gives an oppositely ordered sequence of pseudometrics on $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$. Also recall that for truncations (Definition-Proposition 5.11), if $\alpha \leq \beta$ in $\mathbf{R}_\infty$ then $C_{/\alpha} \geq C_{/\beta}$. This gives by Proposition 5.16 that $d_{C_{/\alpha}} \leq d_{C_{/\beta}}$. Truncations of contours are a way of constructing non-decresing sequences of pseudometrics.

■ **Proposition 7.2.** *Non-decreasing sequence* $\{d_\alpha\}_{\alpha\in\mathbf{R}_\infty}$ *of pseudometrics on* T *is universally admissible, i.e. it is admissible for any* $I\colon T \to \mathbf{N}$.

Proof. Let $\{d_\alpha\}_{\alpha\in\mathbf{R}_\infty}$ be a non-decreasing sequence of pseudometrics on T. Choose ε in $(0,\infty)$. For α in $(0,\infty)$, $\lfloor \alpha/\varepsilon \rfloor$ is the largest natural number not bigger than α/ε. Define metrics

$$d^\varepsilon_\alpha = d_{\lfloor \alpha/\varepsilon \rfloor \varepsilon} \text{ and } d^\varepsilon_\infty = d_\infty.$$

In this way any ε leads to a new non-decreasing sequence $\{d^\varepsilon_\alpha\}_{\alpha\in\mathbf{R}_\infty}$ of pseudometrics on T. Let $\overline{I}_\varepsilon(X)\colon \mathbf{R}^2 \to \mathbf{R}$ be the metric stabilization of I corresponding to this new sequence as defined in Definition 7.1.

The sequence $\{d^\varepsilon_\alpha\}_{\alpha\in\mathbf{R}_\infty}$ is constant on intervals of the form $[n\varepsilon, (n+1)\varepsilon)$ where n is a natural number. The function $\overline{I}_\varepsilon(X)\colon \mathbf{R}^2 \to \mathbf{R}$ is therefore Lebesgue measurable as it is constant on left closed rectangles that cover $\mathbf{R}^2$.

For any ε in $(0,\infty)$ we therefore get an admissible sequence for I. The limit of $\{d^\varepsilon_\alpha\}_{\alpha\in\mathbf{R}_\infty}$ as ε goes to 0 is $\{d_\alpha\}_{\alpha\in\mathbf{R}_\infty}$. Then $\overline{I}(X)$ is the limit of $\overline{I}_\varepsilon(X)$. As a limit of measurable functions, $\overline{I}(X)$ is also measurable and $\{d_\alpha\}_{\alpha\in\mathbf{R}_\infty}$ admissible for I. ■

Similarly to $\mathscr{M}$ we have metrics on $\mathscr{M}_2$, the normalized L_p- (for $p \geq 1$) and interleaving metrics:

$$\widehat{L_p}(f,g) = \lim_{a\to\infty} \frac{1}{a}\int_0^a \left(\int_0^\infty |f(\alpha,t)-g(\alpha,t)|^p dt\right)^{1/p} d\alpha;$$

$$S = \left\{\varepsilon \,\middle|\, \begin{array}{l} f(\alpha,t) \geq g(\alpha,t+\varepsilon) \\ g(\alpha,t) \geq f(\alpha,t+\varepsilon) \end{array} \text{ for } (\alpha,t) \in \mathbf{R}\times\mathbf{R}\right\},$$

$$d_{\bowtie}(f,g) = \begin{cases} \inf(S) & \text{, if } S \text{ is non-empty,} \\ \infty & \text{, otherwise.} \end{cases}$$

For two metrics $d_\alpha \leq d_\beta$ there is an inclusion of metric balls $B_{d_\beta}(x,t) \subset B_{d_\alpha}(x,t)$. Therefore the function $\overline{I}$ is non-decreasing with respect to α. To bound the L_p-distance as α goes to infinity we use the normalized form. As with a single pseudometric, the metric stabilization along a non-decreasing sequence of pseudometrics on T is a process of converting a discrete invariant $I\colon T \to \mathbf{N}$ into a stable invariant $\overline{I}\colon T \to \mathscr{M}_2$. This is manifested in the following theorem.

■ **Theorem 7.3.** *Let $I\colon T \to \mathbf{N}$ be a discrete invariant. Assume $\{d_\alpha\}_{\alpha\in\mathbf{R}_\infty}$ is a non-decreasing sequence of pseudometrics on T. Let $\overline{I}\colon T \to \mathscr{M}_2$ be the metric stabilization of I along this sequence. Then, for any X and Y in T:*

1. $d_\infty(X,Y) \geq d_{\bowtie}\left(\overline{I}(X),\overline{I}(Y)\right)$,
2. $c\, d_\infty(X,Y)^{1/p} \geq L_p\left(\overline{I}(X),\overline{I}(Y)\right)$, *where* $c = max\{\widehat{I}_{d_\infty}(X)(0), \widehat{I}_{d_\infty}(Y)(0)\}$.

Proof. The arguments are similar as used in the proof of Theorem 4.3.
(1): If $d_\infty(X,Y) = \infty$, the statement is clear. Assume $\varepsilon = d_\infty(X,Y) < \infty$. Since $\{d_\alpha\}_{\alpha\in\mathbf{R}_\infty}$ is non-decreasing, for any (α,t) in $\mathbf{R}\times\mathbf{R}$ we have inclusions of metric balls

$$B_{d_\alpha}(Y,t) \subset B_{d_\alpha}(X,t+\varepsilon),$$

and $\overline{I}(Y)(\alpha,t) \geq \overline{I}(X)(\alpha,t+\varepsilon)$. By symmetry $\overline{I}(X)(\alpha,t) \geq \overline{I}(Y)(\varepsilon,t+\varepsilon)$ and hence $d_{\bowtie}\left(\overline{I}(X),\overline{I}(Y)\right) \leq \varepsilon$.
(2): Using (1), it is enough to prove that for functions f and g increasing with respect to first argument and decreasing with respect to second argument

$$\max\{f(\infty,0), g(\infty,0)\} d_{\bowtie}(f,g)^{1/p} \geq L_p(f,g).$$

The inequality is clear if $d_{\bowtie}(f,g) = \infty$. For a fixed α, the inner integral in $\widehat{L_p}(f,g)$ is the L_p distance between f and g. So assume there is ε such that

$f(\alpha,t) \geq g(\alpha,t+\varepsilon)$ and $g(\alpha,t) \geq f(\alpha,t+\varepsilon)$ for any t. From the proof of Theorem4.3 we then have that

$$L_p(f(\alpha,t),g(\alpha,t)) \leq \max\{f(\alpha,0),g(\alpha,0)\}\varepsilon^{1/p}.$$

It then follows that for f and g increasing with respect to α,

$$\begin{aligned}
&\lim_{a\to\infty}\frac{1}{a}\int_0^a L_p(f(\alpha,t),g(\alpha,t))d\alpha \\
&\quad\leq \lim_{a\to\infty}\frac{1}{a}\int_0^a \max\{f(\alpha,0),g(\alpha,0)\}\varepsilon^{1/p}d\alpha \\
&\quad\leq \max\{f(\infty,0),g(\infty,0)\}\varepsilon^{1/p}.
\end{aligned}$$

■

7.2 Life span

To prepare for the embedding theorem, in this section we look at how contour modifies the lengths of bars. This also gives even more explicit computation of the stable rank from the bar decomposition. Let $s < e$ be in $\mathbf{R}_\infty$. Since $\text{rank}(K(s,e)) = 1$, then for any contour C, the value of $\widehat{\text{rank}}_C K(s,e)(t)$ is either 1 or 0 for any t in $\mathbf{R}$. As the function $\widehat{\text{rank}}_C K(s,e)$ is non-increasing, there is l in $\mathbf{R}_\infty$ such that

$$\widehat{\text{rank}}_C K(s,e)(t) = \begin{cases} 1 & \text{, if } t < l, \\ 0 & \text{, if } t > l. \end{cases}$$

We call l the **life span** of the bar $K(s,e)$ with respect to C we denote it by the symbol $\text{life}_C K(s,e)$. If $l = \text{life}_C K(s,e) < \infty$, then the value $\widehat{\text{rank}}_C K(s,e)(l)$ can be either 1 or 0, depending on the contour type. For example, for a closed contour the set of l's such that $C(s,l) \geq e$ is left closed. Then according to Theorem 6.8, $\widehat{\text{rank}}_C K(s,e)(l) = 0$. The following proposition describes how to calculate the life span of a bar with respect to a regular contour.

■ **Proposition 7.4.** *Let $s < e$ be in $\mathbf{R}_\infty$ and α be in $\mathbf{R}$. If C is a regular contour, then*

$$life_C K(s,e) = \begin{cases} \infty & , \text{if } e = \infty, \\ C(s,-)^{-1}(e) & , \text{if } e < \infty; \end{cases}$$

$$life_{C_{/\alpha}} K(s,e) = \begin{cases} 0 & , \text{if } \alpha \leq s, \\ C(s,-)^{-1}(\alpha) & , \text{if } s < \alpha \leq e, \\ C(s,-)^{-1}(e) & , \text{if } e < \alpha; \end{cases}$$

$$life_{C_{/\!/\alpha}} K(s,e) = \begin{cases} C(s,-)^{-1}(s+\alpha) & , \text{if } s+\alpha \leq e, \\ C(s,-)^{-1}(e) & , \text{if } s+\alpha > e. \end{cases}$$

Proof. Recall that according to Theorem 6.8

$$\widehat{\text{rank}}_C K(s,e)(t) = \begin{cases} 1 & , \text{if } C(s,t) < e, \\ 0 & , \text{if } C(s,t) \geq e. \end{cases}$$

Together with the injectivity of regular C this implies that we can find the life span by finding the inverse of $C(s,-) = e$. If $e = \infty$ then $C(s,t) < \infty$ for $t = \infty$.

For the truncated contour $C_{/\alpha}$, if $\alpha \leq s$, then automatically $C(s,-) \geq \alpha$, $C_{/\alpha} = \infty$ and life span is 0. If $s < \alpha \leq e$, then $C_{/\alpha} = C$ until $C(s,-) = \alpha$ after which $C_{/\alpha} = \infty$ and the stable rank goes to 0. Life span is hence given by the inverse of α under injective $C(s,-)$. If $e < \alpha$, stable rank goes to zero before $C(s,-)$ reaches truncation point and the life span is given by the inverse of $C(s,-) = e$.

For translational truncation $C_{/\!/\alpha}$, if $s + \alpha > C(s,-)$, $C_{/\!/\alpha} = C$ and life span is given when $C(s,-) = e$. If $s + \alpha \leq e$, then from the conditions of stable rank we have $C(s,-) \geq e \geq s + \alpha$ which implies that $C_{/\!/\alpha} = \infty$. Stable rank then goes to zero and life span is given by the inverse of $C(s,-) = s + \alpha$. In the condition $s = \infty$ life span is not well defined. ■

Proposition 7.4 together with Theorem 6.8) gives:

■ **Corollary 7.5.** *Let C be a regular contour. If D denotes C, $C_{/\alpha}$ or $C_{/\!/\alpha}$, then:*

$$\widehat{rank}_D \left(\bigoplus_{i \in \bar{n}} K(s(i), e(i)) \right)(t) = |\{i \mid t < life_D K(s(i), e(i))\}|,$$

$$lim \left(\widehat{rank}_D \left(\bigoplus_{i \in \bar{n}} K(s(i), e(i)) \right) \right) = |\{i \mid e(i) = \infty\}|.$$

According to the above proposition, the value of the stable rank at t is the number of bars whose life span with respect to C strictly exceeds t.

7.3 Ampleness

► **Definition 7.6.** Let T be a set of objects from some category. A sequence of pseudometrics $\{d_\alpha\}_{\alpha\in\mathbf{R}_\infty}$ on T is called **ample** for a discrete invariant $I\colon T \to \mathbf{N}$ if:

1. it is admissible for I,
2. the metric stabilization $\overline{I}\colon T \to \mathscr{M}_2$ of I along this sequence has the following property: X and Y in T are isomorphic if and only if $\overline{I}(X) = \overline{I}(Y)$.

◄

Ample sequences of pseudometrics on T for I give stable embeddings of isomorphism classes of objects in T into $\mathscr{M}_2$. With the choice of such an embedding, we can think of T as a subspace of $\mathscr{M}_2$. We can then use the properties of this function space and use e.g. more developed statistical methods to study our data set. Different ample sequences give different embeddings. As with single pseudometric d on T, the idea behind is that by choosing an appropriate such embedding, structural properties of T are reflected in the geometry of the image of $\overline{I}$ described by the various metrics in $\mathscr{M}_2$. The task of a data analyst is then to find the optimal sequence of pseudometrics.

For our final theorem let C be a contour. For every α in $\mathbf{R}_\infty$, we can take the truncations $C_{/\alpha}$ and $C_{//\alpha}$. In this way we get two sequences of contours indexed by $\mathbf{R}_\infty$ such that for any $\alpha < \beta$ in $\mathbf{R}_\infty$:

$$\infty = C_{/0} \geq \cdots \geq C_{/\alpha} \geq \cdots \geq C_{/\beta} \geq \cdots \geq C_{/\infty} = C$$

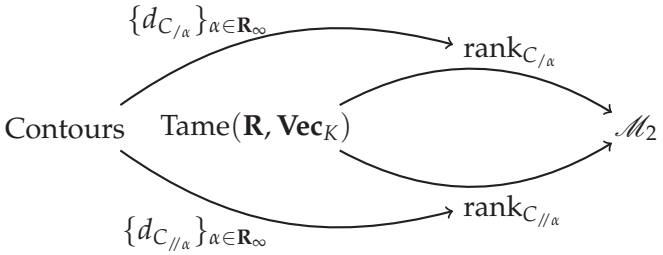
$$\infty = C_{//0} \geq \cdots \geq C_{//\alpha} \geq \cdots \geq C_{//\beta} \geq \cdots \geq C_{/\infty} = C.$$

Each of these contours induces a pseudometric on $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$. In this way, we obtain two sequences of pseudometrics $\{d_{C_{/\alpha}}\}_{\alpha\in\mathbf{R}_\infty}$ and $\{d_{C_{//\alpha}}\}_{\alpha\in\mathbf{R}_\infty}$. By Proposition 5.16, for any V and W in $\mathrm{Tame}(\mathbf{R}, \mathbf{Vec}_K)$ we have the inequalities

$$0 = d_{C_{/0}}(V,W) \leq \ldots \leq d_{C_{/\alpha}}(V,W) \leq \ldots \leq d_{C_{/\beta}}(V,W) \leq \ldots \leq d_C(V,W),$$

$$0 = d_{C_{//0}}(V,W) \leq \ldots \leq d_{C_{//\alpha}}(V,W) \leq \ldots \leq d_{C_{//\beta}}(V,W) \leq \ldots \leq d_C(V,W).$$

Every contour then leads to two non-decreasing sequences of pseudometrics and hence to two metric stabilizations:



■ **Theorem 7.7.** *If C is a regular contour, then the sequences of pseudometrics $\{d_{C_{/\alpha}}\}_{\alpha\in\mathbf{R}_\infty}$ and $\{d_{C_{//\alpha}}\}_{\alpha\in\mathbf{R}_\infty}$ on the category Tame($\mathbf{R}$, $\boldsymbol{Vec}_K$) are ample for the rank function*

$$\mathit{rank}\colon \mathit{Tame}(\mathbf{R}, \boldsymbol{Vec}_K) \to \mathbf{N}.$$

Proof. The argument for the two sequence are entirely analogous so we go through the proof for $\{d_{C_{/\alpha}}\}_{\alpha\in\mathbf{R}_\infty}$.

We have already shown in Proposition 7.2 that non-decreasing sequences of pseudometrics are universally admissible. If V and W in Tame($\mathbf{R}$, $\mathbf{Vec}_K$) are isomorphic, their stable ranks are equal for any $\alpha \in \mathbf{R}_\infty$. So assume that for any $\alpha \in \mathbf{R}_\infty$, $\widehat{\mathrm{rank}}_{C_{/\alpha}} V = \widehat{\mathrm{rank}}_{C_{/\alpha}} W$. We need to show that V and W are isomorphic.

Regularity of C implies that it is closed and $C(t, 0) = t$ for any t. Hence according to Theorem 6.8 V and W have the same rank:

$$\mathrm{rank}(V) = \widehat{\mathrm{rank}}_C V(0) = \widehat{\mathrm{rank}}_C W(0) = \mathrm{rank}(W).$$

We can then assume $V \cong \oplus_{i\in\bar{n}} K(s(i), e(i))$ and $W \cong \oplus_{i\in\bar{n}} K(s'(i), e'(i))$.

Step 1: Reduction to finite bars.
According to Corollary 7.5:

$$|\{i \,|\, e(i) = \infty\}| = \lim\left(\widehat{\mathrm{rank}}_C V\right) = \lim\left(\widehat{\mathrm{rank}}_C W\right) = |\{i \,|\, e'(i) = \infty\}| = n_2.$$

Thus V and W are isomorphic to, respectively:

$$\bigoplus_{i=1}^{n_1} K(s(i), e(i)) \oplus \bigoplus_{j=1}^{n_2} K(s(j), \infty), \qquad \bigoplus_{i=1}^{n_1} K(s'(i), e'(i)) \oplus \bigoplus_{j=1}^{n_2} K(s'(j), \infty).$$

where $e(i), e'(i) < \infty$ for any $i = 1, \ldots, n_1$.

Choose β in $\mathbf{R}$ such that $\beta > e(i), e'(i), s(j), s'(j)$ for any i and j and define:

$$V/\beta = \bigoplus_{i=1}^{n_1} K((i), e(i)) \oplus \bigoplus_{j=1}^{n_2} K(s(j), \beta)$$
$$W/\beta = \bigoplus_{i=1}^{n_1} K(s'(i), e'(i)) \oplus \bigoplus_{j=1}^{n_2} K(s'(j), \beta).$$

Since we are excising the infinite bars beyond any $s(j)$ or $s'(j)$ this does not change the fact that V and W are isomorphic if and only if V/β and W/β are isomorphic. Thus to prove the theorem it is enough to show V/β and W/β are isomorphic.

For any α in $\mathbf{R}$, $\widehat{\text{rank}}_{C_{/\alpha}}(V/\beta) = \widehat{\text{rank}}_{C_{/\alpha}}(W/\beta)$. This follows from the assumption $\widehat{\text{rank}}_{C_{/\alpha}} V = \widehat{\text{rank}}_{C_{/\alpha}} W$, the additivity of stable rank (Theorem 6.8) and Proposition 7.4 which gives that for any $s < \beta$:

$$\text{life}_{C_{/\alpha}} K(s, \beta) = C^{-1}(s, -)(\alpha) = \text{life}_{C_{/\alpha}} K(s, \infty), \text{ if } \alpha \leq \beta$$
$$\text{life}_{C_{/\alpha}} K(s, \beta) = C^{-1}(s, -)(\beta) = \text{life}_{C_{/\beta}} K(s, \infty), \text{ if } \alpha > \beta.$$

The theorem is thus reduced to the case when all the bars in the bar decompositions of V and W are finite.

Step 2: Induction on the rank.
Assume $V \cong \oplus_{i \in \bar{n}} K(s(i), e(i))$, $W \cong \oplus_{i \in \bar{n}} K(s'(i), e'(i))$ and $e(i), e'(i) < \infty$ by reduction to finite bars in Step 1.

We proceed to show by induction on the rank that V and W are isomorphic. This is clear if $\text{rank}(V) = \text{rank}(W) = 0$, since in this case both V and W are isomorphic to 0 by Proposition 3.40.

Assume $n = \text{rank}(V) > 0$. Denote

$$l_i = \text{life}_C K(s(i), e(i)) \text{ and } l_i' = \text{life}_C K(s'(i), e'(i)).$$

From Corollary 7.5 we get that for any t in $\mathbf{R}$,

$$|\{i \mid t < l_i\}| = \widehat{\text{rank}}_C V(t) = \widehat{\text{rank}}_C W(t) = |\{i \mid t < l_i'\}|.$$

It follows that $l_{\max} = \max\{l_i \mid i \in \bar{n}\} = \max\{l_i' \mid i \in \bar{n}\}$.

Let us define the largest ends of bars attaining the maximum lifetime:

$$e_{\max} = \max\{e(i) \mid l_i = l_{\max}\} \text{ and } e'_{\max} = \max\{e'(i) \mid l_i' = l_{\max}\}.$$

It is the case that $e_{\max} = e'_{\max}$. To see this, assume $e_{\max} < e'_{\max}$ and consider the ranks

$$\widehat{\operatorname{rank}}_{C_{/e_{\max}}} V(t) = |\{i \mid t < \operatorname{life}_{C_{/e_{\max}}} K(s(i), e(i))\}|,$$
$$\widehat{\operatorname{rank}}_{C_{/e_{\max}}} W(t) = |\{i \mid t < \operatorname{life}_{C_{/e_{\max}}} K(s'(i), e'(i))\}|.$$

Since $l_{\max}$ was the maximum of lifetimes, for t large enough the stable ranks of V and W reduce to checking life spans under truncation by $e_{\max}$ of those bars attaining $l_{\max}$. Thus there are bars $K(s, e_{\max})$ and $K(s', e'_{\max})$ such that (see Proposition 7.4)

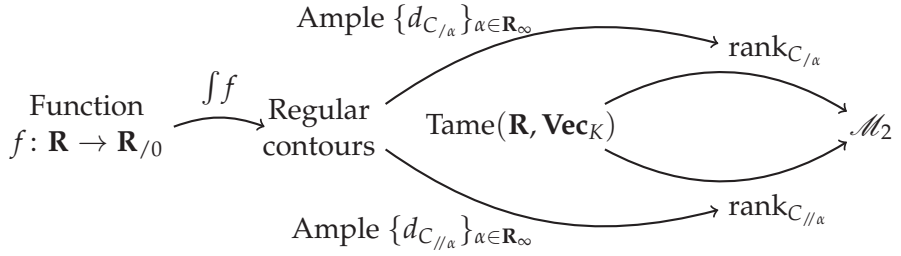
$$\operatorname{life}_{C_{/e_{\max}}} K(s, e_{\max}) = C(s, -)^{-1}(e_{\max}) = l_{\max},$$
$$\operatorname{life}_{C_{/e_{\max}}} K(s', e'_{\max}) = C(s', -)^{-1}(e_{\max}) = l'_{\max} < l_{\max}$$

by the assumption $e_{\max} < e'_{\max}$. It follows that there is $l'_{\max} < t' < l_{\max}$ such that

$$\widehat{\operatorname{rank}}_{C_{/e_{\max}}} V(t') = \widehat{\operatorname{rank}}_{C_{/e_{\max}}} K(s, e_{\max})(t') = 1 >$$
$$0 = \widehat{\operatorname{rank}}_{C_{/e_{\max}}} K(s', e'_{\max})(t') = \widehat{\operatorname{rank}}_{C_{/e_{\max}}} W(t')$$

and we get a contradiction to the assumption of equal stable ranks.
Since C is regular, there is a unique s such that $C(s, l_{\max}) = e_{\max}$. Thus both V and W contain a bar of the form $K(s, e_{\max})$ in their bar decompositions. We can then split off this bar and proceed by induction. ■

Note that the reduction to finite bars in Step 1 of the proof does not change for the sequence $\{d_{C_{/\!/\alpha}}\}_{\alpha \in \mathbf{R}_\infty}$ and the proof goes through analogously. Let us summarize in the diagram below our whole theoretical construction of embedding tame persistent vector spaces into Lebesgue measurable functions:



Chapter 8

Applications

This chapter gathers the theory presented in previous chapters around concrete persistence data analysis. We illustrate how choosing a right contour, the induced stable ranks can lead to improved results in e.g. supervised learning. As outlined in the Introduction, there has been vivid activity in mapping barcodes into different presentations more amenable for machine learning. We perform metric classification using only the metric structure of space $\mathscr{M}$ where rank invariants reside. As mentioned above in the discussion of metric stabilization, the idea is that the information that rank invariants carry about the input data sets is reflected in the geometry of $\mathscr{M}$ by the metrics $d_{\triangleright\!\!\bowtie}$ and L_p. We now make this concrete with actual data analysis. We emphasize that focus here is not on finding an optimal classifier for a specific case. The aim is to show how the choice of metrics is fundamental for persistence and to explain how contours are used in concrete analysis. We also demonstrate that the presented theory leads to a practical TDA pipeline. To generate barcodes needed in the analyses we used the Ripser software [1].

8.1 Visualizing bars and contours

The bars $K(s,e)$ of a tame persistent vector space (recall the bar decomposition from Section 3.5) can be parametrized by the start s and life span with respect to the standard contour C, $\mathrm{life}_C K(s,e) = e - s$. Bars can then be visualized in a $(s, e-s)$-plot as vertical stems. We call this presentation **persistence stem plot** or simply stem plot. Taking into account multiplicity of more than one bar having the same start value we extend the

domain of the stem plot to $\mathbf{R} \times \mathbf{N}$, where $\mathbf{N}$ is used to index bars with the same birth ordered by their life spans. However with real data this is needed basically only for the 0-th homology.
For a fixed ε, the relation $C(s, \varepsilon) < e$ in the second point of Theorem 6.8 describes an area above the parametric curve $\gamma_\varepsilon(s) = (s, C(s, \varepsilon))$ in the (s, e)-plane. Setting $C(s, \varepsilon) = e$ and applying transformation $(s, e) \mapsto (s, e - s)$, we get a curve $\widehat{\gamma}_\varepsilon(s) = (s, C(s, \varepsilon) - s)$. Such curves are typically called contour lines, hence the name contours.
Figure 8.1 illustrates a persistence stem plot along with contour lines of distance and shift contours for few values of ε. Density function used to calculate contours is also shown. Stem plot and contour lines make it easy to understand visually Corollary 7.5: the value of stable rank at ε is the number of those bars that exceed the contour line at ε. Visualizing contours also shows how our framework leads to a rich space of metrics. Stem plot can be an effective tool to gain understanding of stable ranks with respect to different contours and to explore appropriate ones for a given task. This exploratory strategy was used in the analysis tasks of the next sections.

8.2 Point processes

Point processes have gathered interest in TDA community, see for example [3, 26, 42]. We simulated six different classes of point processes on a unit square, see their descriptions below. For each class we produced 500 simulations on average containing 200 points. Let $X \sim PD(k)$ denote that random variable X follows probability disribution PD with parameter k. In particular, Poisson(λ) denotes the Poisson distribution with event rate λ.

Poisson: We first sampled number of events N, where $N \sim$ Poisson(λ). We then sampled N points from a uniform distribution defined on the unit square $[0, 1] \times [0, 1]$. Here $\lambda = 200$.

Normal: Number of events N was sampled from Poisson(λ), $\lambda = 200$. We then created N coordinate pairs (x, y), where both x and y are sampled from normal distribution $N(\mu, \sigma^2)$ with mean μ and standard deviation σ. Here $\mu = 0.5$ and $\sigma = 0.2$.

Matern: Poisson process as above was simulated with event rate κ. Obtained points represent parent points, or cluster centers, on the unit square. For each parent, number of child points N was sampled from Poisson(μ).

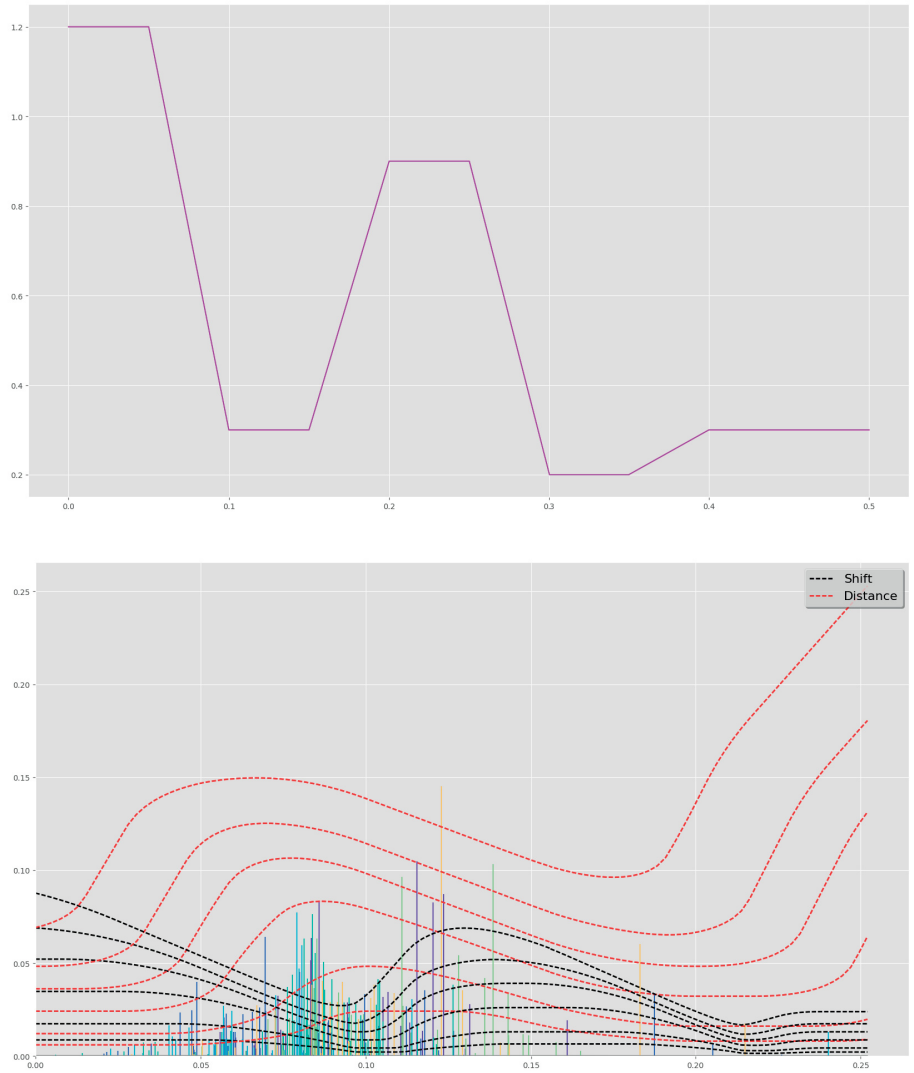


Figure 8.1: Distance and shift type contours coming from the same density function (top) visualized on a stem plot (bottom)

A disk of radius r centered on each parent point was defined. Then for each parent the corresponding number of child points N were placed on the disk. Child points were uniformly distributed on the disks. Note that parent points are not part of the actual data set. We set κ=40, μ=5 and $r = 0.1$.

Thomas: Thomas process is similar to Matern process except that instead of uniform distributions, child points were sampled from bivariate normal distributions defined on the disks. The distributions were centered on the parents and had diagonal covariance $\left[\begin{smallmatrix}\sigma^2 & 0\\ 0 & \sigma^2\end{smallmatrix}\right]$. Here $\sigma = 0.1$.

Baddeley-Silverman: For this process the unit square was divided into equal size squares with side lengths $\frac{1}{14}$. Then for each tile number of points N was sampled, $N \sim$ Baddeley-Silverman. Baddeley-Silverman distribution is a discrete distribution defined on values $(0, 1, 10)$ with probabilities $(\frac{1}{10}, \frac{8}{9}, \frac{1}{90})$. For each tile, associated number of points N were then uniformly distributed on the tile.

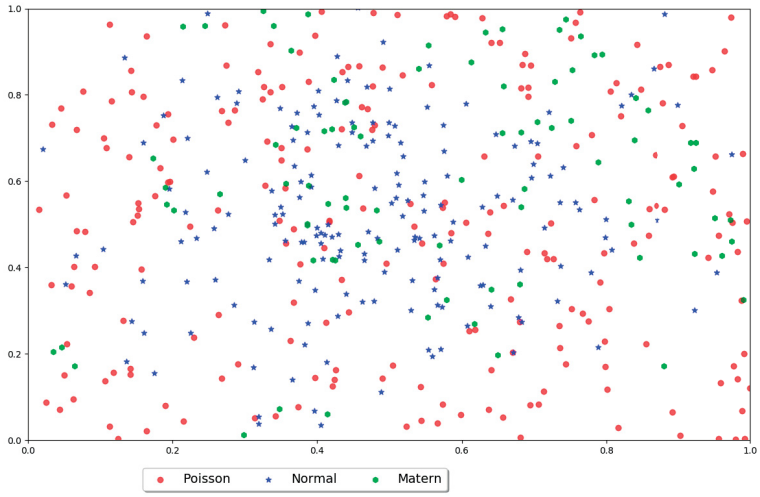
Iterated function system (IFS): We also generated point sets with an iterated function system. For this a discrete distribution is defined on values $(0, 1, 2, 3, 4)$ with corresponding probabilities $(\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$. We denote this distribution by IFS. Starting from an initial point (x_0, y_0) on the unit square, $N \sim \text{Poisson}(200)$ new points are generated by the recursive formula $(x_n, y_n) = f_i(x_{n-1}, y_{n-1})$, where $n \in \{1, ..., N\}$, $i \sim$ IFS and the functions f_i are given as

$$f_0(y, x) = \left(\frac{x}{2}, \frac{y}{2}\right), f_1(y, x) = \left(\frac{x}{2} + \frac{1}{2}, \frac{y}{2}\right), f_2(y, x) = \left(\frac{x}{2}, \frac{y}{2} + \frac{1}{2}\right)$$

$$f_3(y, x) = \left(\left|\frac{x}{2} - 1\right|, \frac{y}{2}\right), f_4(y, x) = \left(\frac{x}{2}, \left|\frac{y}{2} - 1\right|\right).$$

Figure 8.2 shows one realization of the point processes with given parameters. From topological data analysis point of view the point sets hold no distinct large scale topology. It is therefore ideal to study the geometric correlations or features in the filtration captured by homologies in degrees 0 and 1, denoted H_0 and H_1 respectively.

Figure 8.3 shows H_1 stable ranks for one realization of the point processes. Corresponding bars and contours used were shown in Figure 8.1. Note the different character of stable ranks between contours. Distance contour decreases lifespans of bars relative to it making the stable ranks decrease to zero faster compared to stable ranks from shift contour. Also note difference, for example, in Poisson and Baddeley-Silverman stable ranks between contours. Shift contour increases lifespans of Baddeley-Silverman bars around $s = 0.08$ and this can be seen in Baddeley-Silverman dominating Poisson stable rank (Figure 8.3 right). Around $s = 0.08$ the effect of distance contour is opposite, effectively discarding many Baddeley-Silverman bars. After $s = 0.11$ distance contour begins to increase lifes-



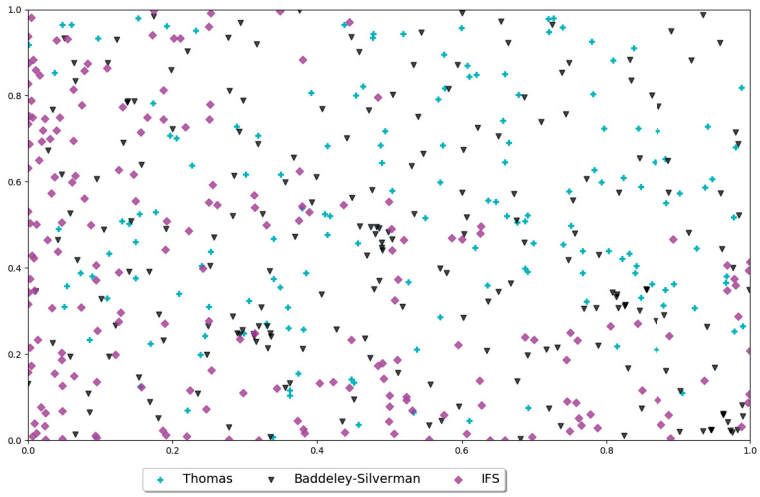


Figure 8.2: Example realizations of point processes on unit square.

pans while shift contour begins to decrease lifespans of the Poisson bars with larger start values. This can be seen in (Figure 8.3 left) with Poisson stable rank having larger support. This example explains how the choice of metric allows analyst to emphasize differently homological features in

persistence analysis depending on what is deemed important in a given task.

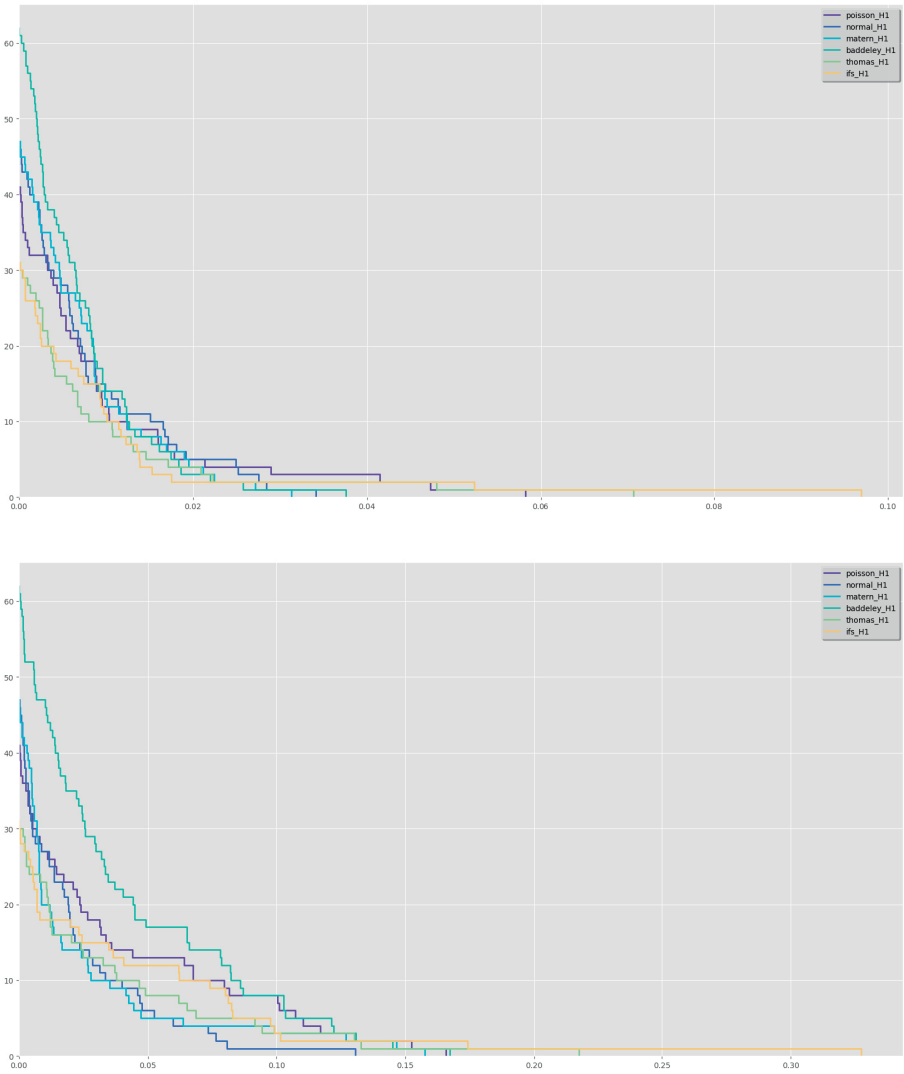


Figure 8.3: Stable ranks in H_1 of point processes for distance (top) and shift (bottom) contours of Figure 8.1. The bars are also shown in the stem plot of Figure 8.1.

Figure 8.4 is a plot of the averages (point-wise means) of H_0 and H_1 stable ranks with respect to the standard contour for 200 simulations of the

point processes. Different point processes are clearly distinguished by their topological signatures. It is worth noting that Matern and Thomas processes are well separated even though in their definition they only differ in the distribution used for point clusters.
To test how well the stable ranks with respect to different contours perform in classifying different point processes we conducted mean classification procedure:

- For each class choose 200 simulations as a training set. Remaining 300 simulations form test set of the class.
- Compute the point-wise means of the training set stable ranks with respect to the chosen contour. These mean invariants are used as classifiers, denoted by $\widehat{C}_{H_\bullet}$, where $H_\bullet$ refers to corresponding homology.
- Denote stable ranks in test set by $T_{H_\bullet}$. Compute distances $L_1(\widehat{C}_{H_\bullet}, T_{H_\bullet})$ between each test element and all classifiers.
- Record found minimum distance by adding 1 to the corresponding pair of the classifier and the test class. Classification is successful if the classifier and the test belong to the same class (in the optimal case the value of the pair (Poisson $\widehat{C}$ class, Poisson T class) would be 300, for example).
- For cross-validation use 20-fold random subsampling: randomly sample 200 stable ranks for classifiers, remaining 300 invariants in each class constitute the test sets. Repeat the classification procedure above 20 times and take the classification accuracy to be the average over the folds.

Obtained cross-validated classification accuracies with standard contour are reported in the confusion matrices of Figure 8.5. The confusion matrices show relative accuracies after dividing by 300 after each fold and averaging after the full cross-validation run. The mean classification accuracy by taking the average over classes (average of the diagonal) is 85% for H_0 and 73% for H_1. The classification procedure performs comparably or better as the hypothesis testing against the homogeneous Poisson process in [3]. Note that no other assumptions or parameter selections were involved in our methodology other than the split between training and test samples (200 and 300, respectively.)

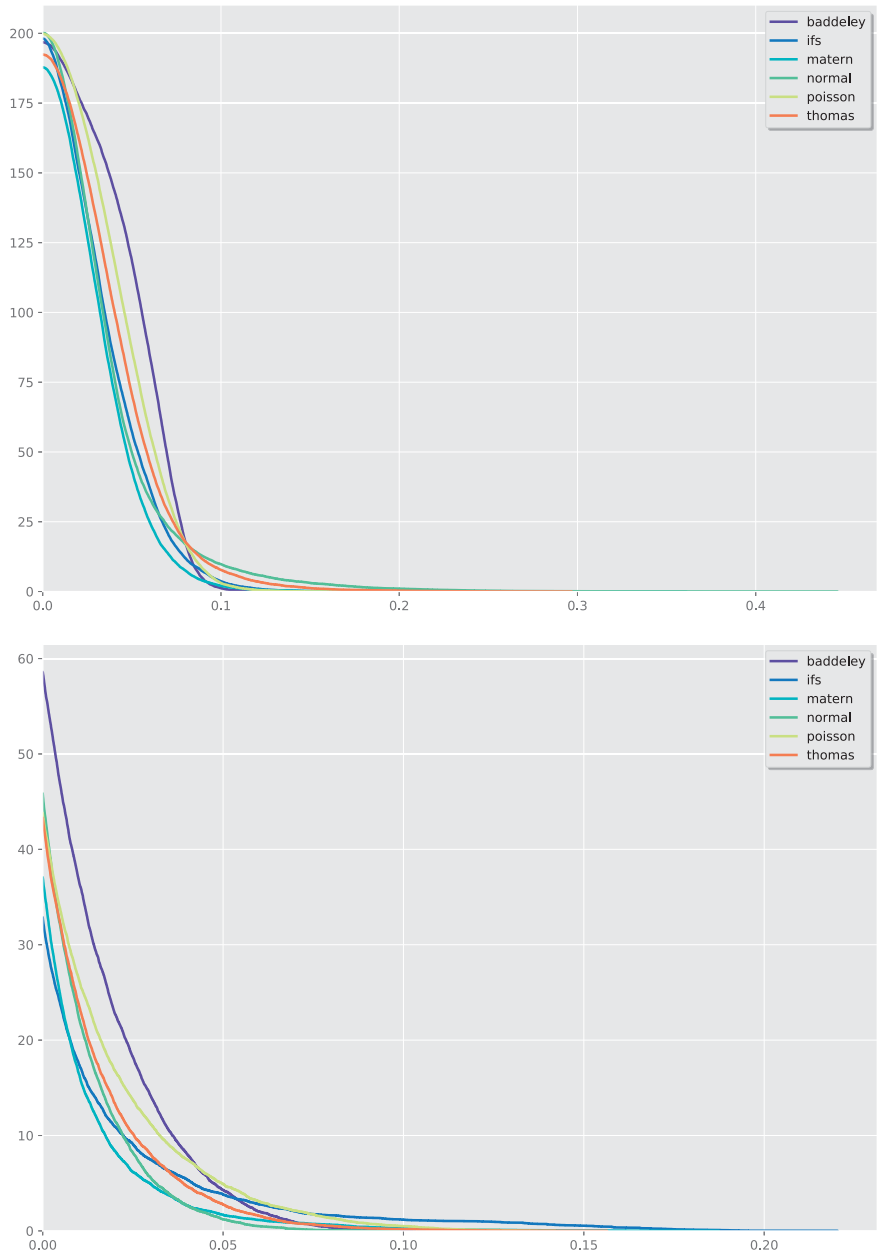


Figure 8.4: Mean stable ranks with respect to the standard contour for H_0 (top) and H_1 (bottom) for 200 simulations.

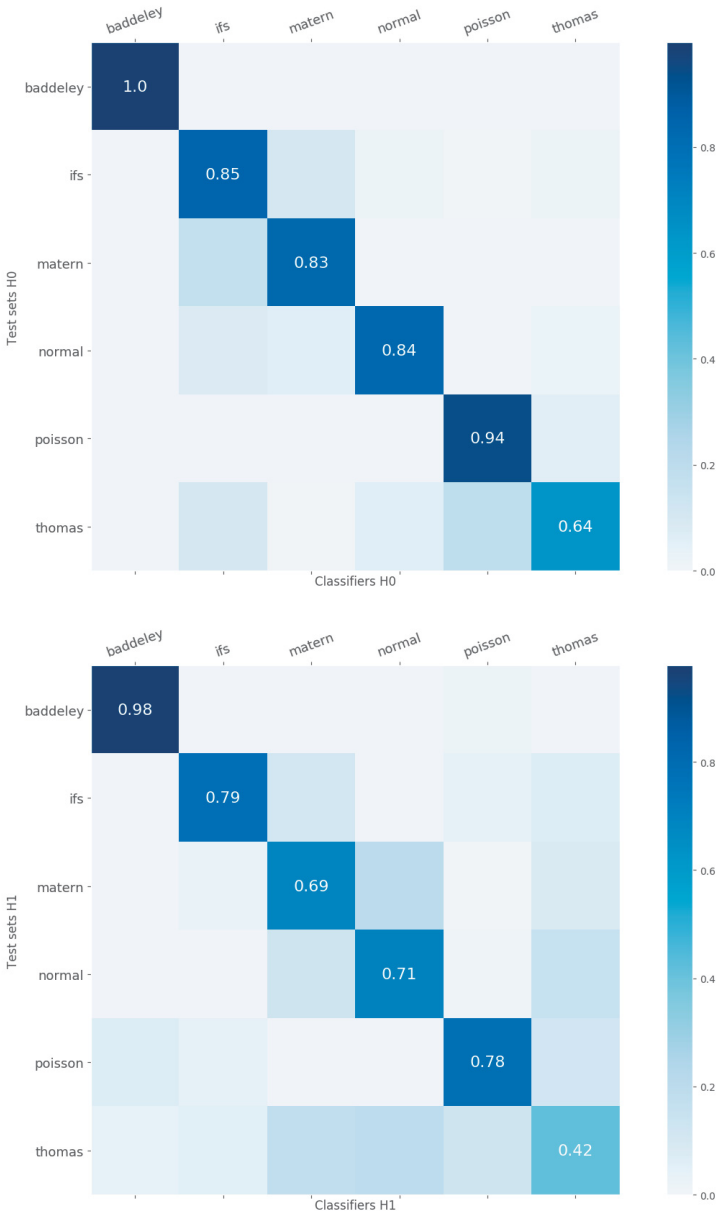


Figure 8.5: Confusion matrices for the point process classification in H_0 (top) and H_1 (bottom) with standard contour.

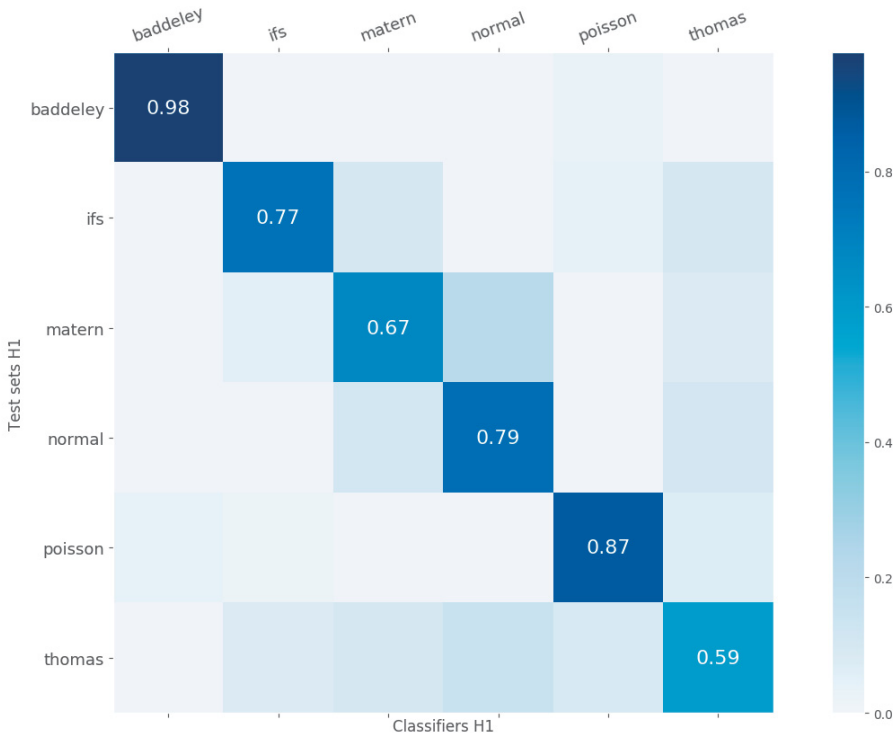


Figure 8.6: Confusion matrix for the classification in H_1 with contour coming from density function in Figure 8.7.

Figure 8.6 shows cross-validated classification accuracies for H_1 stable ranks with shift contour described in Figure 8.7. We thus increase the lifespans of features appearing in the middle of the filtration. The overall classification accuracy increased to 78%. Particularly classification accuracy of the Thomas process was drastically improved as shown in the confusion matrix of Figure 8.6. Also noteworthy is the improvement in the accuracy of normal and Poisson processes. Used shift contour thus captures relevant distinguishing homological information of the point processes better than compared to the standard contour.

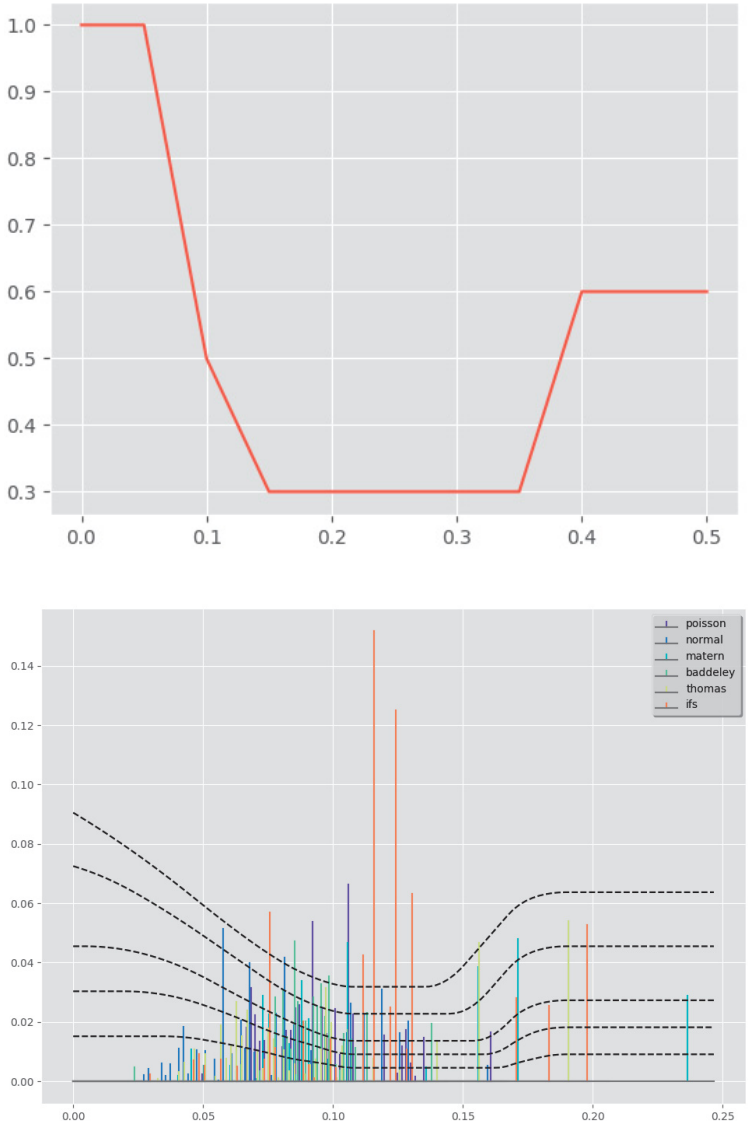


Figure 8.7: Density function used in producing shift contour for point process classification in H_1 (top). Corresponding contour lines and stem plots from H_1 persistence analysis of one realization of the studied point processes (bottom).

8.3 Activity monitoring

As an application to real data we studied activity monitoring of different physical activities. Used data set was PAMAP2 data obtainable from [37]. It makes sense to use all the persistence information, i.e. to combine homologies of different degrees into single classification scheme. In this section we demonstrate how this is enabled by stable ranks and our pipeline.

The data consisted of seven persons from the PAMAP2 data set performing different activities such as walking, cycling, vacuuming or sitting. Test subjects were fitted with three Inertial Measurements Units (IMUs), one on wrist, ankle and chest, and a heart rate monitor. Measurements were registered every 0.1 seconds. Each IMU measured 3D acceleration, 3D gyroscopic and 3D magnetometer data. One data set thus consisted of 28-dimensional data points indexed by 0.1 second timesteps.

We looked at two activities in this case study: ascending and descending stairs. At the outset one would expect these activities to be very similar and therefore difficult to distinguish. For persistence analysis we randomly sampled without replacement 100 points from each data set, repeated 100 times. For each of the 7 subjects we thus obtained 100 resamplings from the activity data. We computed H_0 and H_1 persistence for each sampling. The classification procedure was the same as outlined in Section 8.2 except we combined both homologies in the classifier as follows. We took the mean of 40 out of 100 stable ranks both in H_0 and H_1. We thus obtained 14 classifier pairs $(\widehat{C}_{H_0}, \widehat{C}_{H_1})$ corresponding to all (subject, activity) classes. Remaining 60 signatures formed test data pairs (T_{H_0}, T_{H_1}) in each class. For a pair we then found

$$\min(L_1(\widehat{C}_{H_0}, T_{H_0}) + L_1(\widehat{C}_{H_1}, T_{H_1})).$$

Again the classification is successful if the minimum is obtained with $(\widehat{C}_{H_0}, \widehat{C}_{H_1})$ and (T_{H_0}, T_{H_1}) belonging to the same (subject, activity) class.

Result for 20-fold random subsampling cross-validation is shown in Figure 8.8 for the standard contour. Overall accuracy is 60%.

Classification results using the standard contour for H_0 and the shift contour of Figure 8.9 for H_1 are shown in Figure 8.10. This contour increases lifespans of features appearing with larger filtration scales. Overall accuracy increased to 65%. Note particularly increase in the accuracy of Subject 4. Noteworthy increases in the classification accuracy can be seen also with other subjects such as Subjects 6 and 8. These data thus exhibit

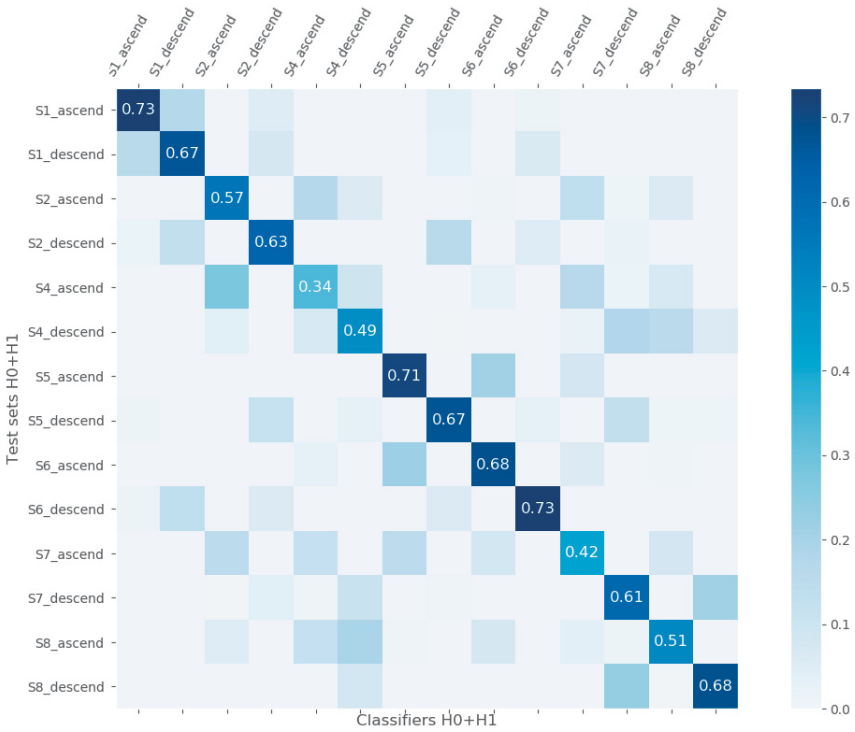


Figure 8.8: Confusion matrix for the classification of ascending and descending stairs activities with standard contour.

some different characterictics in their persistence fingerprint and using an appropriate contour makes this difference more pronounced.
This case study demonstrates that analysis with contours leads to recognizing data that seems to diverge from other similar data. Further analysis by domain expert might e.g. identify characteristics of Subject 4 that make the larger scale features in the persistence output dominate since this is enhanced by the chosen contour. Also noteworthy in the results is that ascendings mainly get confused with ascendings and the same for descendings. Even though not expected, these similar activities differ when looked through persistence analysis.

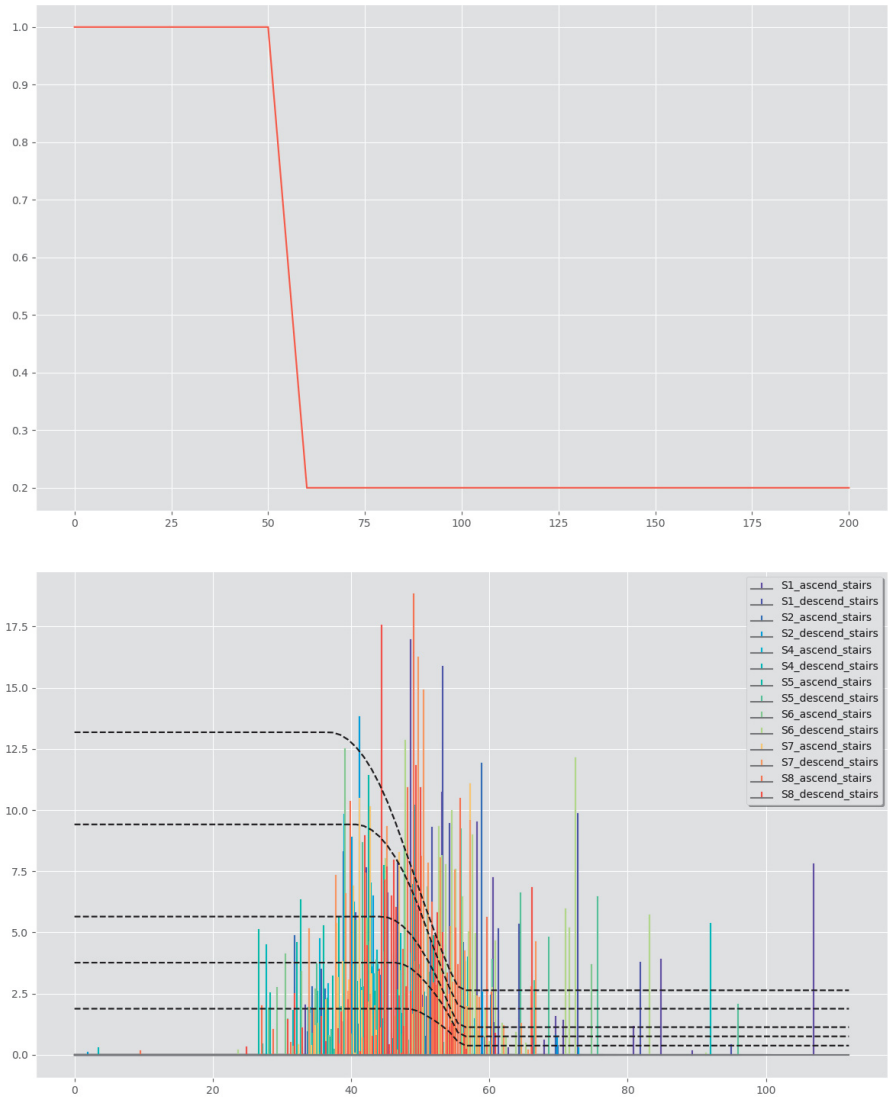


Figure 8.9: Density function used for H_1 stable ranks in the activities classification (top) and contour lines and persistence stem plots for single data sets (bottom).

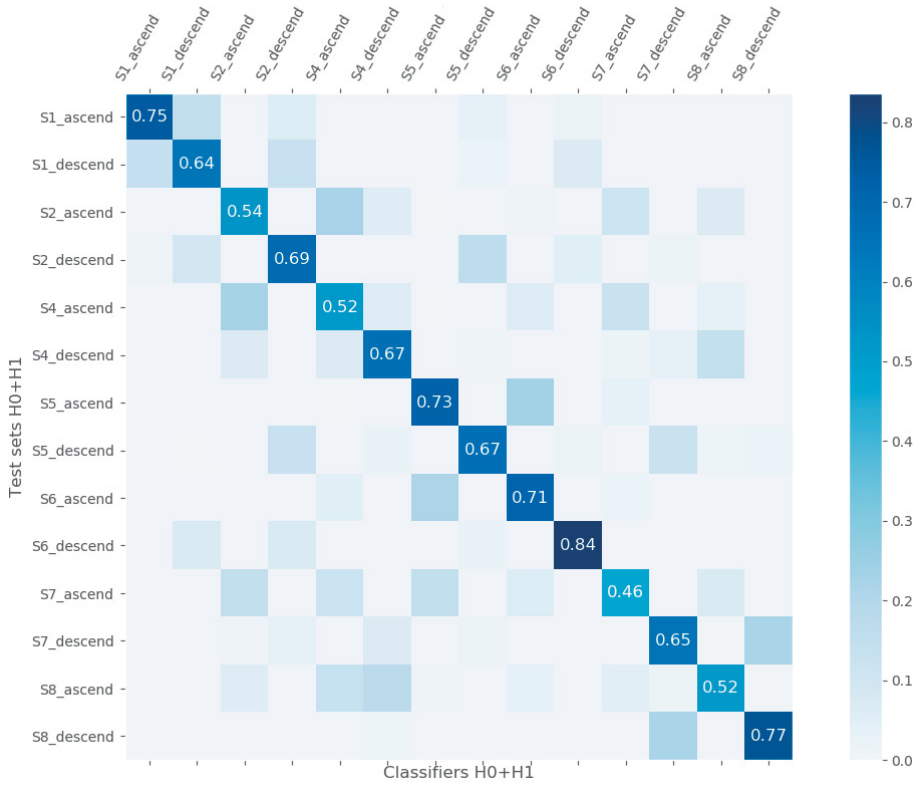


Figure 8.10: Confusion matrix for the classification of ascending and descending stairs activities with contour coming from density function of Figure 8.9.

8.4 Cloud fields and homological density estimation

Cloud field refers to the pattern that clouds form in the sky. Clouds have an important role in the climate system, such as transport of heat and moisture from the Earth's surface and and an impact on solar radiation budget. To take into account these effects clouds need to be parameterized in large-scale climate models. Cloud patterns are, however, very localized and can change rapidly. Due to the small spatiotemporal scales, improving the parameterizations it is necessary to have an efficient notion describing the spatial structure of a cloud field. There are similarities with the study of point processes of Section 8.2 but the spatial structure of cloud fields arises from clustering and pattern formation. This section describes an

initial work in quantifying these phenomena in a time series of cloud patters through stable rank invariant. See [29] for further results and more references.

Cloud fields can be simulated. The data used here was produced by the Dutch Atmospheric Large-Eddy Simulation model. Simulations cover time periods of few hours during one day and the data is saved for analysis at 15 minute intervals. In this study 10 days with different initial conditions were used. The data consists large amount of physical information from which cloud patterns can be extracted. The spatial simulation domain in x, y, z coordinates is $12.8 \times 12.8 \times 5$ in kilometers with horizontal resolution of 50 meters and vertical resolution of 40 meters. To have a manageable data sets points are sampled from cloud clusters with 5% proportional sampling, meaning that 5% of the points in each cloud cluster are sampled to the data set. This ensures that at least one data point is taken from each cloud. The 3D cloud field from the simulation domain was in this case study flattened in the z-direction onto a 2D plane. The cloud fields are then as visualized in black in Figure 8.11 which shows two examples of simulated cloud fields with very different pattern formations. Sampled points are also shown as blue circles. The units are not in physical kilometers but refer to image pixels.

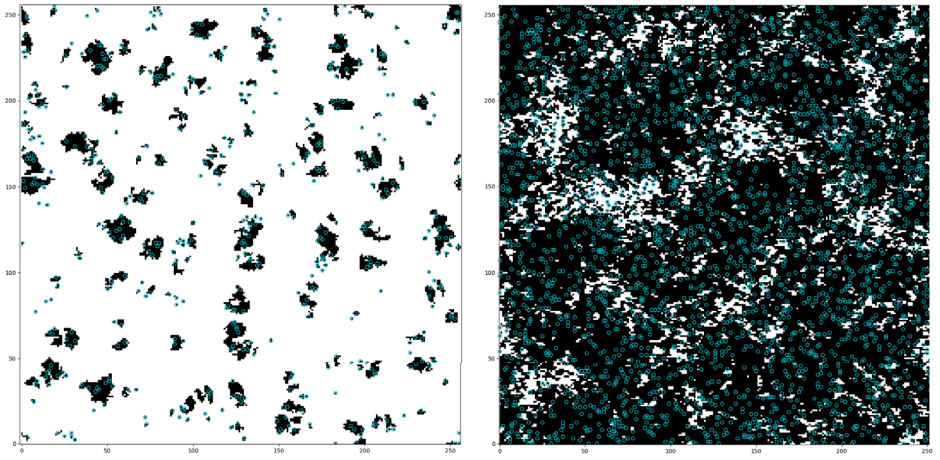


Figure 8.11: Examples of cloud fields (in black) with very different pattern characteristics.

If stable rank function is normalized by dividing by its maximum value at zero, we obtain a function with values decresing from 1 to some limit

value. This limit can be made 0 since stable rank can always be made to decrease to zero by truncation. This function can be seen as a 1-CDF, where CDF refers to a cumulative distribution function of some random variable. In fact this is known as a survival function. The normalized stable rank at t is thus an indication of the relative amount of homological features as measured by the rank that persist beyond t.
This makes it possible to perform homological density estimation from persistence analysis. Given a normalized stable rank from a real data, one then tries to fit a parametric curve onto the stable rank. This was tried in the context of cloud data. Figure 8.12 shows an example of this procedure, homology curve denotes stable rank. The author provided normalized stable ranks with respect to standard contour and the density estimation shown in Figure 8.12 is credited to J. Licón-Saláiz.
The fitting procedure was as follows. For any H_i, accumulate from the barcodes the lengths of persistence intervals in different length classes. Normalize the accumulated lenghts by the size of the population of bars, i.e if there were n_j bars in barcode j, the size of the population is $\sum_j n_j$. This gives the empirical population density function of bar lengths shown in the upper left panel of Figure 8.12. The x-axes are the lengths of bars in actual meters. Fit a parametric density function to the population density by using e.g. maximum likelihood. Fitted parametric densities shown are log-normal

$$f_{\widehat{\text{rank}}}(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln t - \mu)^2}{2\sigma^2}\right)$$

and exponential

$$f_{\widehat{\text{rank}}}(t) = \lambda e^{-\lambda t}.$$

The upper right panel shows the cumulative distribution functions for the fitted densities, along with the empirical CDF from the population density. The survival functions of the fitted densities and the normalized stable rank function as shown in the lower left panel. Lower right panel is the same as the upper right, but with the addition of a cumulative stable rank by integrating the stable rank function. The density estimation procedure is seen to be one possible option to parameterize cloud fields since the fitted curves depend only on few parameters. This also provides a starting point for statistical analysis based on stable ranks.
To study cloud data under different contours we used standard contour and contours visualized in Figure 8.13. These contours are referred to as contour 1 and contour 2. Analyses were performed as follows:

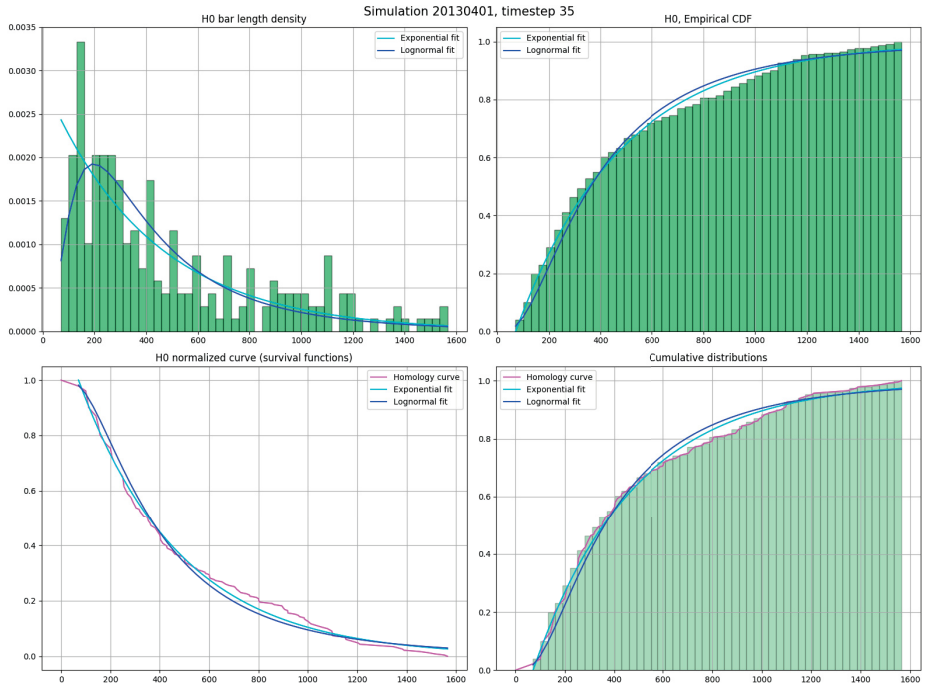


Figure 8.12: Example of homological density estimation with stable ranks.

- To reduce the effect of sampling, cloud field from each simulation time step was resampled 10 times with 5% proportional sampling.
- Stable rank for this cloud field was the mean stable rank of the 10 resamplings, normalized to give the survival function as explained above. Stable ranks were computed in H_0 and H_1 with respect to standard contour, contour 1 and contour 2. Altogether the data consisted of 254 normalized stable ranks for each class of contours.
- Distance matrices using interleaving, L_1- and L_2-metrics were then computed for all the different classes of stable ranks.
- Dendrograms from the distance matrices were visually analyzed to decide on a number of clusters of stable ranks.

Figure 8.14 shows H_1 analysis results for interleaving metric. Contour 2 seems to give the clearest clustering with three clusters. Plots of the stable ranks in different clusters also show that contour 2 gives stable ranks

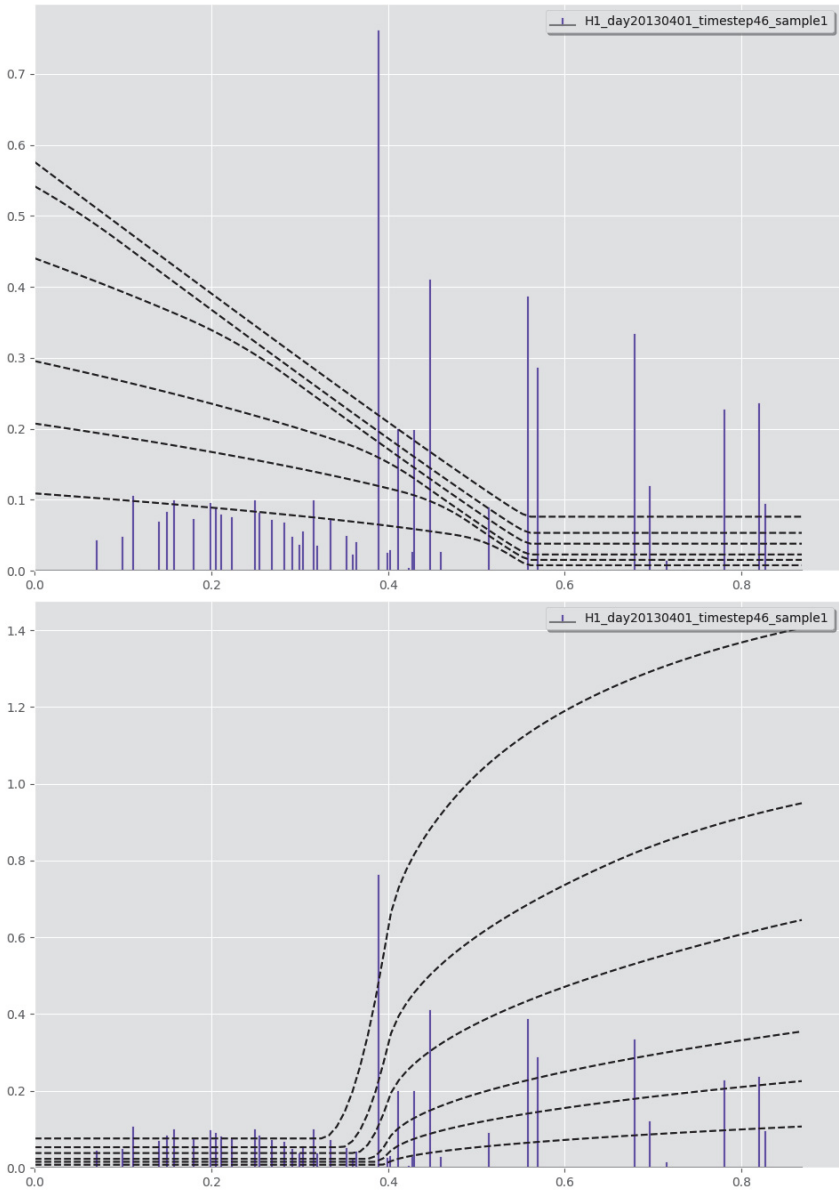


Figure 8.13: Contour 1 (top) and contour 2 (bottom) used in the analysis of cloud fields. Stem plot is from one sampling of a cloud field at one time step.

that are most clearly separated between clusters as compared to standard contour and contour 1. With cloud fields, putting emphasis on the homological features appearing at the beginning of filtration thus seems to be a disinguishing factor, at least with interleaving metric between the induced stable ranks.
This should be compared to point processes of Section 8.2. Figure 8.15 shows H_1 stable ranks by cluster when L_1-metric was used as with point processes. In this case contour 2 again gives the best separation into clusters. With point processes the contour used to enhance classification accuracy increased the life span of features in the middle of filtration, see 8.7. When contour increasing life spans in the beginning of filtration, as contour 2, was used with point processes, classification accuracy was 55%, compared to the 78% found in Section 8.2. Contour analysis thus seems to indicate that different pattern formations occur in cloud fields and spatial point processes.
From the analysis it seems feasible to parameterize cloud fields using stable rank with respect to an appropriate contour. Particularly since through density estimation stable ranks can be represented by simple parametrically fitted curves. Of course further investigation is still needed: what is the optimal contour, how the cloud fields in clusters are distributed to time steps, how the various physical variables correspond to structure of cloud fields in and between the clusters, how the cloud fields corresponding to clusters are similar in some other measure such as cloud cover fraction.

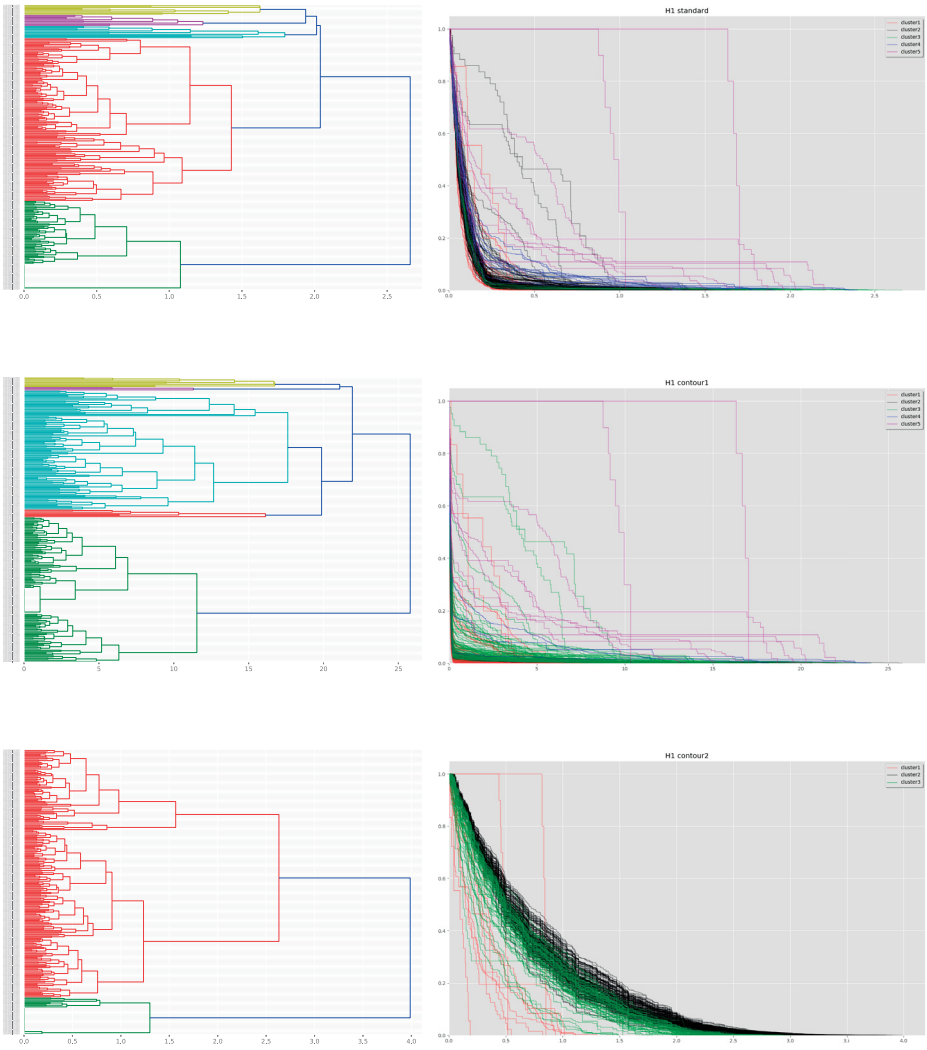


Figure 8.14: Dendrograms from clustering of cloud field H_1 stable ranks with respect to interleaving metric. Stable ranks corresponding to different clusters are shown in the right column. Top standard contour, middle contour 1 and bottom contour 2.

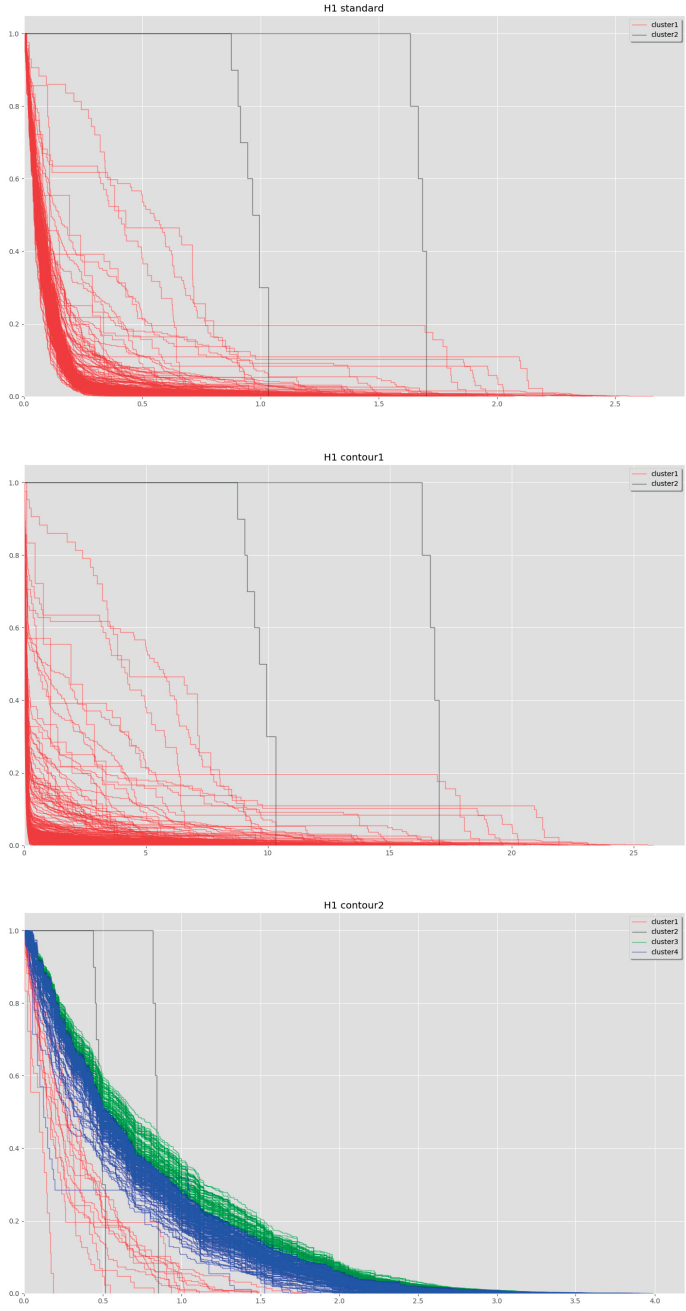


Figure 8.15: Cloud field stable ranks by cluster in H_1 for a clustering with L_1 metric. Top standard contour, middle contour 1 and bottom contour 2.

Chapter 9

End discussion

9.1 Conclusions

This work presented theoretical foundation for stabilizing the rank invariant, or minimal number of generators, of a persistent vector space in one-dimensional case along with a variety of data analysis applications. The framework of metric stabilization in 4 was shown to be a general way of producing continuous invariants from discrete invariants. The flexibility of this approach to persistence arises from the choice of a metric. We showed in Chapter 5 how metrics can be derived from contours. Contour again can be derived from integrating a density function giving rise to a rich space of metrics. In Chapter 6 we proved main properties of stable rank and derived algorithmic way of computing it. In Chapter 7 we proved that stabilization of rank along a sequence of metrics embeds isomorphism classes of persistent vector spaces into the space of Lebesgue measurable functions.

In Chapter 3 we proved main properties of the rank. Decisive property in one-dimensional persistence is the monotonicity of the rank with respect to taking a subspace of a persistent vector space. We showed how the bar decomposition results from this property. We thus claim that rank is a fundamental invariant for persistence. More so since rank and its metric stabilization are defined generally for multi-dimensional persistence whereas monotonicity of the rank only holds in one-dimensional case.

The flexibility of choosing a metric for stabilizing rank is not only of theoretical importance. Major part of the project behind the presented work was to learn to use contours and resulting stable rank invariants in various data analysis tasks. This was achieved to a high degree as

presented in Chapter 8. Stable ranks are functional fingerprints for data sets living in the space of Lebesgue measurable functions $\mathscr{M}$ with various metrics between functions. Given a collection T of persistent vector spaces obtained from input data, the aim of data analysis pipeline based on stable rank is to study how the associated invariants from different choices of metric on T reflect its structural properties relevant to a given data analysis task. The expectation is that these properties should also be reflected by the geometry of $\mathscr{M}$ described by its metrics.
In Chapter 8 we showed with various case studies how the ideas outlined above work in practice. With an appropriate choice of contour/metric we were able to enhance classification accuracies in supervised learning. Indications were also given how statistical analysis with stable ranks might be done. This thesis aimed to be a first step towards data analysis pipeline around stable rank invariant. The presentation was also aimed to be a manual for anyone wanting to implement and use the pipeline.

9.2 Open questions and outlook

Many avenues for future research are still ready for exploration. We list various possible directions for further development. The list is in no particular order of importance or relevance.

Intuition of contours for different data sets

In Chapter 8 we explored different contours with the help of persistence stem plots and visualization of contours. For data analyst using the pipeline it would be advantageous to build an intuition for different contours to get a feeling for an appropriate one for the data at hand. Here the visual tool is helpful. It is also important to understand how the contour modifies bars and what the resulting stable rank picks out from the data.

Learning contours

For a full scale machine learning one needs to develop a learning algorithm for contours: given a collection of data sets from few classes, what is the contour that best separates stable ranks of data from different classes? This might be optimization in the function space of contours. Looking at

visualizations in Chapter 8 and the modification of life spans in Proposition 7.4, contours can also be seen as giving weigths to bars. Is it possible to discretize contours such that they can be learned as weight vectors for bar features?

Analysis of contours

We showed two ways of producing contours by integrating density functions in Section 5.2, namely shift and distance type integration. The aim is to get to pseudometrics through contours, the way of getting there is rather irrelevant. Is there some more universal analytical way of producing contours? What is the relation of shift and distance type contours? Figure 8.1 gives indication of some "inverse derivative" relation. We only gave one example of a contour that is not a monoid action on $(\mathbf{R}, +)$, namely the parabolic contour in Definition 5.8. It would be interesting to study generally how to give contours that are not actions and what is particular of them compared to true actions.

Multidimensional contours

Theory of stable rank originated for multidimensional persistence in [14] and was continued in [23]. Noise systems were defined as general means to give metrics on $\mathrm{Tame}(\mathbf{R}^n, \mathbf{Vec}_K)$. To extend metric stabilization with contours into multidimensional persistence one needs to produce multidimensional contours. There is work in this direction around software TopCat [22].

Statistics for stable rank

In the end one wants to make statistical conclusions from a collection of data sets. This requires development of basic statistics for stable ranks. The notion of homological density estimation discussed in Section 8.4 is also an interesting path to explore for parameterizing outputs from persistence analysis.

Choice of function space

The target space of metric stabilization was the space of Lebesgue measurable functions. This allowed to show universal admissibility for sequences

of pseudometrics in Proposition 7.2. But its very difficult to give non-measurable functions and therefore examples of inadmissible sequences of pseudometrics. Can we use some other function space to restrict what sequences are admissible?

Differential geometric point of view

Contours modify the life spans of bars by Proposition 7.4. Looking at the visualizations in Chapter 8 one comes to see contours as giving a local change of length scale over the filtration domain. Persistent vector space can be seen as parameterized by a manifold $\mathbf{R}$. Is there some differential geometric formulation where contour can be related to how a diffeomorphism $\phi\colon \mathbf{R} \to \mathbf{R}$ changes the metric tensor? And more generally when the parameterization domain is $\mathbf{R}^n$? If we let persistent vector space be parameterized by a general manifold M, contours that are actions are then in fact flows $\Phi\colon M \times \mathbf{R} \to M$ on the manifold.

Stabilizing other invariants

Metric stabilization could be used to generally stabilize discrete invariants. This needs 1) a category of interest, 2) a discrete invariant of interest and, most crucially, 3) a definition of a pseudometric in the category. If this is to be used in practical TDA, all this needs to be computable as well.

Universality of d_C

The interleaving distance d_I is universal on multidimensional persistence modules, i.e. for any other stable metric d, $d \leq d_I$, where stability is with respect to d_∞ in the sense of sublevelset persistence [28]. With respect to the standard contour, d_C and d_I are related [14]. It would be interesting to study universality of d_C with respect to a general contour C and with respect to L_p and $d_{\bowtie}$ stabilities. Moreover, could the universality guide in the selection of contours in applications? We want to thank Claudia Landi for suggesting these questions.

Bibliography

[1] U. Bauer. Ripser software. *github.com/Ripser/ripser*.

[2] P. Bendich, J. Marron, E. Miller, A. Pieloch, and S. Skwerer. Persistent homology analysis of brain artery trees. *The Annals of Applied Statistics*, 10:198–218, 2016.

[3] C. Biscio and J. Møller. The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications. *arXiv:1611.00630*, 2016.

[4] H. Bjerkevik, M. Botnan, and M. Kerber. Computing the interleaving distance is np-hard. *arXiv:1811.09165*, 2018.

[5] M. Botnan and M. Lesnick. Algebraic stability of zigzag persistence modules. *Algebraic & Geometric Topology*, 18:3133–3204, 2018.

[6] P. Bubenik. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16:77–102, 2015.

[7] P. Bubenik, V. de Silva, and J. Scott. Metrics for generalized persistence modules. *Foundations of Computational Mathematics*, 15:1501–1531, 2015.

[8] P. Bubenik and J. Scott. Categorification of persistent homology. *Discrete & Computational Geometry*, 51:600–627, 2014.

[9] G. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46:255–308, 2009.

[10] G. Carlsson. Topological pattern recognition for point cloud data. *Acta Numerica*, 23:289–368, 2014.

[11] G. Carlsson and V. de Silva. Zigzag persistence. *Foundations of Computational Mathematics*, 10:367–405, 2010.

[12] G. Carlsson and A. Zomorodian. The theory of multidimensional persistence. *Discrete & Computational Geometry*, 42:71–93, 2009.

[13] M. Carrière, M. Cuturi, and S. Oudot. Sliced wasserstein kernel for persistence diagrams. *arXiv:1706.03358*, 2017.

[14] W. Chachólski, A. Lundman, R. Ramanujam, M. Scolamiero, and S. Öberg. Multidimensional persistence and noise. *Foundations of Computational Mathematics*, 17:1367–1406, 2017.

[15] W. Chachólski and H. Riihimäki. Metrics and stabilization in one parameter persistence. *submitted to SIAM Journal on Applied Algebra and Geometry*, 2019.

[16] F. Chazal, D. Cohen-Steiner, M. Glisse, L. J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. *Proceedings of the 25th annual symposium on Computational geometry*, SCG09:237–246, 2009.

[17] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37:103–120, 2007.

[18] H. Edelsbrunner and J. Harer. Computational topology-an introduction. *American Mathematical Society*, 2010.

[19] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28:511–533, 2002.

[20] H. Adams et al. Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18, 2017.

[21] P. Frosini. Measuring shapes by size functions. *Intelligent Robots and Computer Vision X: Algorithms and Techniques*, SPIE Vol. 1607:122–133, 1992.

[22] O. Gäfvert. Topcat software. *github.com/olivergafvert/topcat*.

[23] O. Gäfvert and W. Chacholski. Stable invariants for multidimensional persistence. *arXiv:1703.03632*, 2017.

[24] R. Ghrist. Homological algebra and data. *IAS/Park City Mathematics Series*, 25:273–325, 2018.

[25] Y. Hiraoka, T. Nakamura, A. Hirata, E. Escolar, K. Matsue, and Y. Nishiura. Hierarchical structures of amorphous solids characterized by persistent homology. *PNAS*, 113:7035–7040, 2016.

[26] Y. Hiraoka, T. Shirai, and T. K. Duy. Limit theorems for persistence diagrams. *The Annals of Applied Probability*, 28:2740–2780, 2018.

[27] G. Kusano, K. Fukumizu, and Y. Hiraoka. Kernel method for persistence diagrams via kernel embedding and weight factor. *Journal of Machine Learning Researc*, 18:1–41, 2018.

[28] M. Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Foundations of Computational Mathematics*, 15:613–650, 2015.

[29] J. Licón-Saláiz, H. Riihimäki, and T. W. van Laar. Topological characterization of shallow cumulus cloud fields using persistent homology. *Proceedings of the 8th International Workshop on Climate Informatics*, CI 2018:107–110, 2018.

[30] S. MacLane. Homology (reprint of the 1975 edition). *Springer*, 4th printing, 1994.

[31] P. McCullagh. What is a statistical model. *The Annals of Statistics*, 30:1225–1310, 2002.

[32] K. Meehan and D. Meyer. An isometry theorem for generalized persistence modules. *arXiv:1710.02858*, 2017.

[33] C. Moon, N. Giansiracusa, and N. Lazar. Persistence terrace for topological inference of point cloud data. *arXiv:1705.02037*, 2017.

[34] J. Munkres. Elements of algebraic topology. *Addison-Wesley Publishing Company*, 1984.

[35] N. Otter, M. Porter, U. Tillmann, P. Grindrod, and H. Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6, 2017.

[36] S. Oudot. Persistence theory: From quiver representations to data analysis. *American Mathematical Society*, 2015.

[37] PAMAP. Physical activity monitoring for aging people. *www.pamap.org*.

[38] V. Puuska. Erosion distance for generalized persistence modules. *arXiv:1710.01577*, 2017.

[39] J. Reininghaus, S. Huber, U. Bauer, and R. Kwitt. A stable multi-scale kernel for topological machine learning. *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR2015:4741–4748, 2015.

[40] H. Riihimäki and W. Chachólski. Generalized persistence analysis based on stable rank invariant. *arXiv:1807.01217*, 2018.

[41] V. Robins. Towards computing homology from finite approximations. *Topology Proceedings*, 24:503–532, 1999.

[42] A. Robinson and K. Turner. Hypothesis testing for topological data analysis. *Journal of Applied and Computational Topology*, 1:241–261, 2017.

[43] J. Rotman. An introduction to algebraic topology. *Springer*, 1998.

[44] J. Rotman. An introduction to homological algebra 2 ed. *Springer*, 2009.

[45] D. Spivak. Category theory for the sciences. *The MIT Press*, 2014.

[46] B. Stolz, H. Harrington, and M. Porter. Persistent homology of time-dependent functional networks constructed from coupled time series. *Chaos*, 27:047410–1 – 047410–17, 2017.

[47] K. Xia and G.-W. Wei. Persistent homology analysis of protein structure, flexibility, and folding. *International Journal for Numerical Methods in Biomedical Engineering*, 30:814–844, 2014.

[48] K. Xia and G.-W. Wei. Persistent topology for cryo-em data analysis. *International Journal for Numerical Methods in Biomedical Engineering*, 31, 2015.

[49] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33:249–274, 2005.