SPRINGER BRIEFS IN MATHEMATICS

Frédéric Chazal
Vin de Silva
Marc Glisse
Steve Oudot

The Structure and Stability of Persistence Modules

Springer

SpringerBriefs in Mathematics

SpringerBriefs in Mathematics showcases expositions in all areas of mathematics and applied mathematics. Manuscripts presenting new results or a single new result in a classical field, new field, or an emerging topic, applications, or bridges between new results and already published works, are encouraged. The series is intended for mathematicians and applied mathematicians.

More information about this series at http://www.springer.com/series/10030

Frédéric Chazal · Vin de Silva
Marc Glisse · Steve Oudot

The Structure and Stability of Persistence Modules

Springer

Frédéric Chazal
Inria Saclay – Île-de-France
Palaiseau
France

Marc Glisse
Inria Saclay – Île-de-France
Palaiseau
France

Vin de Silva
Pomona College
Claremont, CA
USA

Steve Oudot
Inria Saclay – Île-de-France
Palaiseau
France

ISSN 2191-8198 ISSN 2191-8201 (electronic)
SpringerBriefs in Mathematics
ISBN 978-3-319-42543-6 ISBN 978-3-319-42545-0 (eBook)
DOI 10.1007/978-3-319-42545-0

Library of Congress Control Number: 2016945105

Mathematics Subject Classification (2010): 55-02, 16G20

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

Our intention, at the beginning, was to write a short paper resolving some technical issues in the theory of topological persistence. Specifically, we wished to present a clean easy-to-use framework for continuous-parameter persistence, building on the well-studied case of discrete-parameter persistence. Over a number of years, we gradually abandoned the idea of a short paper. It apparently takes us about a hundred pages to explain things in a simple way. We take the time to develop the concepts that make everything flow: decorated reals, quiver calculus, r-measures, q-tameness. Our main concern has been to make the mathematics lucid. We hope the reader will find the ideas useful and the presentation clarifying.

Palaiseau, France — Frédéric Chazal
Claremont, CA, USA — Vin de Silva
Palaiseau, France — Marc Glisse
Palaiseau, France — Steve Oudot

May 2016

Acknowledgments

Many important ideas in this monograph have their origins in earlier work with David Cohen-Steiner and Leo Guibas [15] and Gunnar Carlsson [8]. We deeply appreciate their intellectual influence.

We are grateful to several people who directly helped improve this monograph. William Crawley-Boevey has been an invaluable source of expertise on module decomposition. The current form of Theorem 2.8, proved by him, is particularly convenient for our exposition. Michael Lesnick provided us with an early version of his work [42], drew our attention to the paper of Webb [48], and corrected our original statement of Theorem 2.8. Peter Landweber read the manuscript closely and gave us many detailed suggestions and corrections. Finally, we thank the anonymous referees whose suggestions have greatly improved this monograph.

This work has been supported by several research grants:

- Digiteo Foundation. C3TTA and a Digiteo Chair (held by VdS).
- European Commission. Computational Geometric Learning (255827).
- Agence Nationale de Recherche. GIGA (ANR-09-BLAN-0331-01).
- DARPA. Sensor Topology and Minimal Planning (HR0011-07-1-0002).
- Simons Foundation. Simons Fellowship (grant #267571 to VdS).

We gratefully acknowledge the support of these funding agencies. We also thank the Institute for Mathematics and its Applications, for hosting and partially supporting a sabbatical (VdS), with funds provided by the National Science Foundation.

Last but not least, we thank our home institutions for their ongoing support: Inria Saclay – Île-de-France (FC, MG, SO) and Pomona College (VdS).

Contents

Chapter 1
Introduction

We intend this monograph to be a self-contained presentation of the theory of persistence modules over the real line. We give the best proofs we know of the most important results. Each theorem is located at an appropriate level of abstraction (we believe).

- Newcomers will find this to be an accessible introduction to the algebraic foundations of persistence. They will learn what persistence modules are, how to access their internal structure, their different degrees of 'tameness', how to construct their persistence diagrams mathematically, and how to show that those diagrams are stable.
- Experts will find that our methods add precision and power to what they already know. We construct persistence modules in great generality and show that they can be manipulated just as easily as the traditional examples. We give evidence that this greater generality occurs naturally 'in the wild'.

Persistence modules are the mathematical object at the heart of the young, rapidly growing field of topological data analysis. This field—a blend of computer science, algebraic topology and statistics—is founded on the assumption that scientific data sets carry information in their internal structure and that sometimes this internal structure is topological. Persistence modules were designed to carry topological information about a data set at many different scales simultaneously. This information can be extracted in the form of an invariant—the *persistence diagram* or *barcode*—that can be computed effectively and is statistically robust.

New researchers in the field have to come to terms with multiple aspects of persistence. They must learn how data can be turned into geometric objects (usually a nested family of simplicial complexes). They must learn to work with the algorithms that turn these geometric objects into barcodes or persistence diagrams. They need algebraic topology to know how to interpret these barcodes. And, increasingly often nowadays, they must know enough statistical theory to draw valid inferences from the results of these calculations. There is by now a great deal of literature covering

F. Chazal et al., *The Structure and Stability of Persistence Modules*,
SpringerBriefs in Mathematics, DOI 10.1007/978-3-319-42545-0_1

these different facets, written by and intended for researchers across a large range of disciplines.

Our focus is narrower. In this monograph, we are concerned almost exclusively with the mathematical properties of persistence modules. We have several reasons for doing this:

- The existing literature is largely built around particular instances of topological persistence, such as the sublevelset persistent homology of a Morse function on a compact manifold. While this can be useful for developing intuition, it does create bias in how the subject is understood. We wish to correct this bias.
- Several components of the theory are algebraic in nature. The fact that most of the usual examples come from topology sometimes creates a different impression. Here we give a purely algebraic presentation of the main ingredients: the persistence diagram, tameness, stability. Topological arguments are used when studying the properties of persistence modules obtained from topological data, but are not needed otherwise.
- Variations such as image and kernel persistence can be treated equally.

More specifically, we are interested in persistence modules indexed by a single real parameter. Much of our work goes to show that a continuous parameter can be handled just as effectively as a discrete parameter; this fills a gap in the literature. On the other hand, we make no attempt to discuss multiparameter persistence. It is a complicated subject and at some point will be ready for its own book.

Within its scope, the monograph is almost entirely self-contained. We do appeal to two off-the-shelf technical results from module theory (Theorems 2.7 and 2.8). When discussing persistent *homology*, we assume that the reader is familiar with the standard properties of simplicial complexes and homology functors. Since it is important to remain connected to the larger world of topological data analysis, we briefly present two applications in Sects. 1.3 and 1.4, and a themed reading list in Sect. 1.5.

1.1 A Brief History of Persistence

The early history of persistence is concerned with the quantity

$$\mathrm{r}_t^s = \mathrm{rank}(\mathrm{H}(X_s) \to \mathrm{H}(X_t))$$

for an object X represented at two different scales s, t, and where H is homology. This appeared in the early 1990s in the work of Frosini [34], with different notation and under the name ‘size function’. Independently, a few years later, Robins [46] introduced the term ‘persistent Betti numbers’ for quantities of the form $\mathrm{r}_{\varepsilon+\rho}^{\varepsilon}$, and noted their stability with respect to Hausdorff distance.

The modern theory of persistence is built on three pillars:

- The persistence diagram, and an algorithm for computing it, were introduced by Edelsbrunner, Letscher and Zomorodian [30]. This gives a compact representation of the size function and an effective way to compute it.
- Zomorodian and Carlsson [50] defined persistence modules in the abstract, indexed by the natural numbers and viewed as graded modules over the polynomial ring $\mathbf{k}[t]$. This introduced tools from commutative algebra.
- Cohen-Steiner, Edelsbrunner and Harer [19] formulated and proved the stability theorem, which guarantees that the persistence diagram is robust to changes in the input data. Robustness is measured in terms of a 'bottleneck distance' between persistence diagrams.

All three papers make the assumption that the data is essentially finite. This is understandable from the perspective of computer science: a finite machine can only handle a finite amount of data. And mathematically it is the natural place to begin. In the realm of continuous topology it is common to make finiteness assumptions: a continuous function on a smooth manifold may be approximated by a Morse function, and on a polyhedron may be approximated by a piecewise-linear map.

The finiteness restrictions were lifted in [15] (and its published conference version [14]), which generalises the main results to persistence modules indexed over the real line, under the relatively mild assumption that $\mathrm{r}_t^s < \infty$ for $s < t$. In the present work, we call these modules 'q-tame'. It turns out that this is a natural condition; large classes of examples are q-tame. Moreover, the formulation of the stability theorem in [15] is purely algebraic, and not tied to any particular geometric situation. The only drawback is that some of the arguments are rather complicated.

In this monograph, we carry out the program of [15] with new arguments. The proofs are now very clean and the methods are versatile. Let us say a few words about why we think it worth the effort to work with continuous-parameter persistence modules. Here are our two main reasons:

- Real-world data sets are always finite, but they may be statistical samples from an underlying continuous object or process. Ideally the persistent homology of a sample will be an approximation to the persistent homology of the continuous model. Formulating this requires a theory of continuous-parameter persistence.
- Continuous-parameter persistence extends the applicability of the theory from finite data sets to more general objects such as compact metric spaces. This widens the applicability of persistence within pure mathematics.

In support of this last point, we observe that in recent years the discrete form of persistence has seen application in various branches of pure mathematics. For example, Ellis and King [33] use persistence to study p-groups; and Pakianathan and Winfree [45] have reformulated a number of famous problems in number theory, including the Riemann Hypothesis, in terms of the persistent homology of certain filtered simplicial complexes.

We draw attention to three recent papers which share our goal of understanding continuous-parameter persistence modules:

- Lesnick [42] gives an extensive algebraic treatment of modules over one or more real parameters. The converse stability inequality, and hence the isometry theorem, appears for the first time in his work. Our present work was carried out largely independently, with one salient exception: it was from Lesnick [42] we learned of results of Webb [48] that resolved a sticking-point for us.
- Bubenik and Scott [5] develop the category-theoretical view of persistence modules. This allows them to formulate and prove stability theorems in great generality. Categories and functors surface occasionally in the present monograph, and can be used to streamline some of the work very effectively and non-trivially.
- Bauer and Lesnick [3] give a completely original proof of the stability theorem for q-tame continuous-parameter persistence modules, making no use of the interpolation lemma that is so crucial to our approach (which is based on the original proof of Cohen-Steiner, Edelsbrunner and Harer [19]). Their work depends on strong results on the decomposition of persistence modules.

All of these works appeared during the writing of this monograph.

1.2 Main Contributions

Many authors have studied persistence modules in recent years, and many of the theorems presented here are not original in themselves. The originality lies in the methods that we use. Our main innovations are these:

- We construct persistence diagrams using measure theory. The existence of a diagram is equivalent to the existence of a certain kind of measure on rectangles in the plane.
- We introduce ‘decorated’ real numbers for two related purposes: to remove the ambiguity about the endpoints of persistence intervals, and to get the measure theory to work.
- We introduce a special notation for calculations on quiver representations. This considerably simplifies the linear algebra (for instance, in proving the ‘box lemma’).
- We define several kinds of ‘tameness’ for a persistence module. These occur naturally in practice. The most restrictive of these, finite type, is what is normally seen in the literature. We show how to work effectively with the less restrictive hypotheses.
- We give a clean proof of the algebraic interpolation theorem of [15].
- We rewrite the algebraic stability theorem of [15] as a theorem about measures. Among other consequences, this leads to diagram stability results for even quite badly behaved persistence modules.

Our goal in introducing these ideas is to enable our readers to define persistence diagrams cleanly and effectively in a wide variety of situations. In the earlier work [15], continuous-parameter persistence diagrams are constructed using a careful limiting process through ever-finer discretisations of the parameter. Unfortunately the limiting arguments turn out to be quite complicated, and the resulting diagrams are difficult to work with. Our approach in the present monograph gives the best of both worlds; we are able to work with broader classes of persistence modules, and we can reason about their diagrams in a clean way using arguments of a finite nature.

1.3 Application: Stable Descriptors for Metric Spaces

In this section and the next, we illustrate the usefulness of the persistence diagram for problems in data analysis and machine learning. The key point is that the persistence diagram is a robust invariant of the underlying geometric situation.

It often happens in classification tasks that the objects to be classified cannot be compared against one another directly but only through descriptors or 'signatures'. Consider the problem of organizing a database of 3-dimensional objects into meaningful classes, as illustrated in Fig. 1.1. Each object—or '3d shape'—is represented as some part of the bounding surface of a 3-dimensional region, and it can be abstracted mathematically as a compact metric space $(P, \mathrm{d_P})$ where the set P is the surface itself and the metric is the geodesic distance along the surface. Comparing two 3d shapes amounts to comparing the corresponding metric spaces.

The natural distance on the space of compact metric spaces is the so-called *Gromov–Hausdorff distance* $\mathrm{d_{GH}}$, a generalisation of the classical Hausdorff distance defined as follows:

$$\mathrm{d_{GH}}(P, Q) = \inf_{S,f,g} \mathrm{d_H}(f(P), g(Q))$$

where $(P, \mathrm{d_P})$ and $(Q, \mathrm{d_Q})$ are the two given compact metric spaces, where $(S, \mathrm{d_S})$ ranges over all metric spaces, where $f : P \to S$ and $g : Q \to S$ range over all isometric embeddings of P and Q into S, and where $\mathrm{d_H}$ denotes the Hausdorff distance in $(S, \mathrm{d_S})$.

The issue with this distance is that its direct computation leads to a quadratic assignment problem that is hard to solve in practice. Upper bounds can be obtained easily, but lower bounds are more difficult. A workaround is to map the metric spaces P, Q to some space of signatures in which distances are easier to compute. Ideally, one would like the distance between signatures to be the same as the Gromov–Hausdorff distance between the initial metric spaces, but this is usually too much to ask. We can at least require that the distance between signatures provide a lower bound on the distance between the metric spaces, so in particular the signatures are provably stable under small perturbations of the spaces.

The general stability theorem stated in the present monograph (Theorem 5.25) makes it possible to derive such a stability guarantee, when the signature of a compact

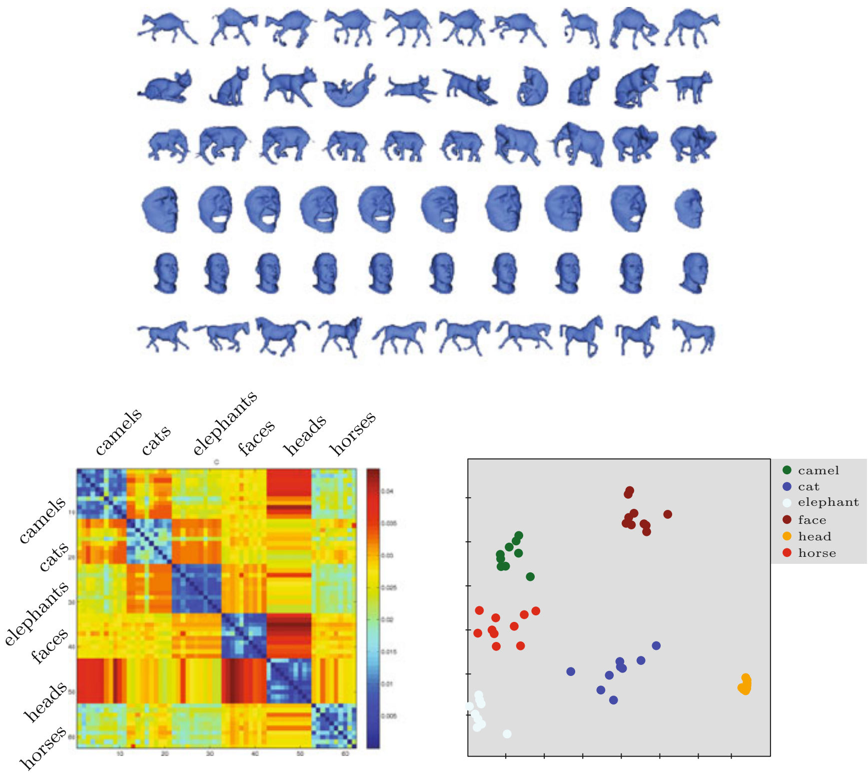


Fig. 1.1 (From [13]) An unsupervised classification task using persistence-based signatures. *Top:* The collection of 60 shapes to be classified into 6 classes (labels unknown). *Bottom-left:* The distance matrix in signature space, with color-coded values (each row and column corresponds to a single shape in the collection). *Bottom-right:* The signatures are embedded into the Euclidean plane using multidimensional scaling. The objects are then classified by a simple k-means clustering procedure applied to this embedding. Label names can be extrapolated to classes if some of the individual objects have known labels

metric space is taken to be the persistence diagram of the homology of its Vietoris–Rips complexes[1]:

Theorem 1.1 ([13, 16]) *For any compact metric spaces* $(P, \mathrm{d_P})$ *and* $(Q, \mathrm{d_Q})$,

$$\mathrm{d_b}(\mathsf{dgm}(\mathrm{H}_*(\mathbb{R}\mathrm{ips}(P))),\ \mathsf{dgm}(\mathrm{H}_*(\mathbb{R}\mathrm{ips}(Q)))) \leq \mathrm{d_{GH}}(P, Q).$$

[1]We do not give the details of this construction here; see [13, 16] for instance. What matters is that the signature $\mathsf{dgm}(\mathrm{H}_*(\mathbb{R}\mathrm{ips}(P)))$ is easily computed, the distance between two signatures is easily computed, the distance is robust in the sense of Theorem 1.1, and that the lower bound in the theorem is sufficiently tight to solve the learning problem under consideration.

As desired, this result provides a lower bound on the Gromov–Hausdorff distance between the metric spaces in terms of the bottleneck distance d_b between their signatures.

It turns out that signatures of this type are rich enough to be used effectively in machine learning applications such as the one depicted in Fig. 1.1. In such applications, the continuous shapes themselves are replaced by finite samples for practical purposes. One can exploit Theorem 1.1 to prove minimax-optimal upper bounds on the convergence rate of the sample signatures to the signatures of the underlying continuous objects [17]. It is also possible to define local versions of the signatures, with similar stability guarantees, for use in partial comparison and matching applications [11].

1.4 Application: Stable Clustering Using Persistence

Unsupervised learning or clustering is an important tool for understanding and interpreting data. Among the wealth of existing approaches, mode seeking is the one most closely related to Morse theory and persistence. The approach assumes that the data points have been drawn from some unknown probability distribution with density f. The idea is to detect the local peaks of f and use them as cluster centers, grouping each data point with the local peak that it eventually reaches by following the gradient vector field of f uphill—assuming f has a well-behaved gradient flow.

A common issue faced by these techniques is that the gradient and extremal points of a density function are notoriously unstable, so their approximation from a density estimator can lead to unpredictable results (Fig. 1.2, top-right). One possible workaround is to smoothe the estimator before launching the hill-climbing procedure; this raises the difficult question of how much smoothing is needed to remove the noise without affecting the signal and still obtain the correct number of clusters. A different workaround, proposed in [18], is to do the hill-climbing anyway and afterwards detect and merge the unstable clusters to regain some stability. This is where persistence comes into play.

The persistence diagram of the density estimator $\hat{f}$ provides a measure of prominence for the peaks of $\hat{f}$, through the distances of their corresponding diagram points to the diagonal. Whenever the diagram satisfies a 'sufficient separation' condition, it is easy to pick a prominence threshold that will separate the relevant peaks of $\hat{f}$ from the irrelevant ones. The clusters associated to the irrelevant peaks can then be merged into the clusters of the relevant peaks, using the hierarchy built by the persistence algorithm. In this way one obtains the correct number of clusters: see Fig. 1.2.

Theorem 1.2 ([18]) *Let c be the Lipschitz constant of the true density f, let η be the approximation error (in the supremum norm) of the estimator $\hat{f}$, and let δ be the neighborhood size parameter used in the hill-climbing procedure. Assume that the peaks of the true density f have prominences at least d. Then, for any prominence threshold within the range $(2(c\delta + \eta),\ d - 3(c\delta + \eta))$, the number of clusters*

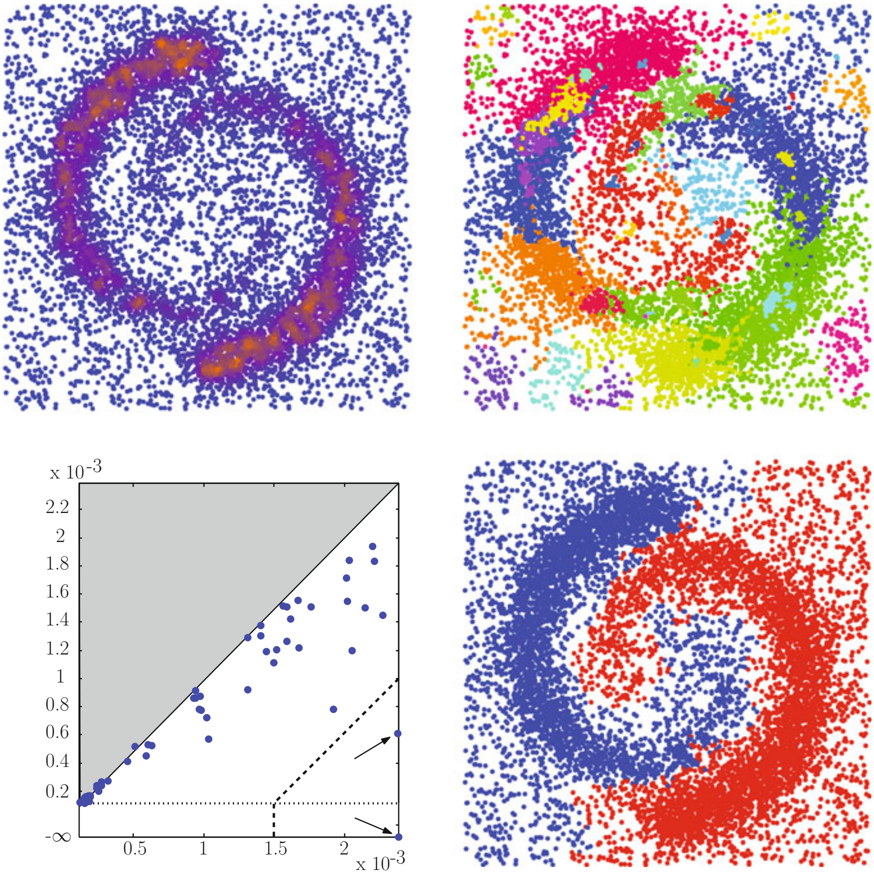


Fig. 1.2 (From [18]) Persistence-based clustering in a nutshell. *Top-left:* The underlying density function f is estimated at the data points. *Top-right:* The hill-climbing procedure applied to the estimated density $\hat{f}$ leads to a large number of unreliable clusters. *Bottom-left:* The persistence diagram of $\hat{f}$ shows 2 points far off the diagonal corresponding to the 2 peaks of the true density f. *Bottom-right:* The final result is obtained by merging the clusters of the other peaks of $\hat{f}$

computed by the above procedure, on an input of n random sample points drawn i.i.d. according to f, is equal to the number of peaks of f with probability at least $1 - e^{-\Omega((c\delta+\eta)n)}$, where the constant in the big-Ω notation depends only on geometric quantities (e.g. volumes of balls) associated with the ambient space.

The proof of this result relies on a partial notion of interleaving and a version of the stability theorem for such partial interleavings (Theorem 6.1). Both follow easily from the framework developed in this monograph.

1.5 Recommended Reading

There is by now a substantial literature on topological persistence, launched by the work of Edelsbrunner, Letscher and Zomorodian [30], with antecedents in papers of Frosini [34] and Robins [46]. Beyond these historical documents, we have some themed suggested readings for the reader seeking orientation in the larger field of topological data analysis.

Applied Algebraic Topology

A substantial grounding in the broad field of applied algebraic topology may be found in the following books, each reflecting the particular tastes of its author(s).

- Tomasz Kaczynski, Konstantin Mischaikow, and Marian Mrozek. *Computational Homology*, volume 157 of *Applied Mathematical Sciences*. Springer, 2004.
- Afra Zomorodian. *Topology for Computing*, volume 16 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, 2005.
- Herbert Edelsbrunner and John L. Harer. *Computational topology: an introduction*. American Mathematical Society, 2010.
- Robert Ghrist. *Elementary Applied Topology*. CreateSpace Independent Publishing Platform, September 2014.
- Steve Y. Oudot. *Persistence Theory: from quiver representations to data analysis*, volume 209 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2015.

Topological Data Analysis

The following survey articles provide different perspectives on the application of topological persistence to data analysis. These articles provide a good introduction for readers new to the field. In particular, they explain how persistence diagrams (or, equivalently, barcodes) are used in practice.

- Robert Ghrist. Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.
- Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.
- Herbert Edelsbrunner and Dmitriy Morozov. Persistent homology: theory and practice. In *European Congress of Mathematics*, pages 31–50. European Mathematical Society, 2012.

The Persistence Algorithm

The following articles deal specifically with the details of the persistence algorithm, an aspect that is not covered in the present book.

- Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28:511–533, 2002.

- Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.
- Vin de Silva, Dmitriy Morozov, and Mikael Vejdemo-Johansson. Dualities in persistent (co)homology. *Inverse Problems*, 27:124003, 2011.

Stability Theorems

The following articles present various versions of the proof of stability for persistence diagrams. We are omitting [15], which served as a basis for the present work.

- David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.
- David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have Lp-stable persistence. *Foundations of Computational Mathematics*, 10(2):127–139, 2010.
- Ulrich Bauer and Michael Lesnick. Induced matchings of barcodes and the algebraic stability of persistence. In *Proceedings of the 30th Annual Symposium on Computational Geometry (SoCG)*, pages 355–364, Kyoto, Japan, June 2014.

Geometric Complexes on Compact Metric Spaces

The following paper uses the tools introduced in the present book to derive stable topological signatures for compact metric spaces.

- Frédéric Chazal, Vin de Silva, and Steve Oudot. Persistence stability of geometric complexes. *Geometriae Dedicata*, 173:193–214, 2014.

It follows previous work in the non-persistent context:

- Jean-Claude Hausmann. On the Vietoris-Rips complexes and a cohomology theory for metric spaces. In *Prospects in Topology*, volume 138 of *Annals of Mathematical Studies*, pages 175–188. Princeton University Press, Princeton, New Jersey, 1995.
- Janko Latschev. Vietoris-Rips complexes of metric spaces near a closed Riemannian manifold. *Archiv der Mathematik*, 77:522–528, 2001.

Variations on Persistence

The following articles introduce several variations of 1-dimensional persistence: vineyards, extended persistence, image persistence, zigzag persistence. The tools introduced in the present book can be used to simplify some parts of their analysis.

- David Cohen-Steiner, Herbert Edelsbrunner, and Dmitriy Morozov. Vines and vineyards by updating persistence in linear time. In *Proceedings of the 22nd Annual Symposium on Computational Geometry (SoCG)*, pages 119–126, 2006.
- David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Extending persistence using Poincaré and Lefschetz duality. *Foundations of Computational Mathematics*, 9:79–103, 2008.

- David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Dmitriy Morozov. Persistent homology for kernels, images, and cokernels. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1011–1020, 2009.
- Gunnar Carlsson and Vin de Silva. Zigzag persistence. *Foundations of Computational Mathematics*, 10(4):367–405, 2010.

Multidimensional Persistence

The following articles extend the theory to persistence modules indexed over multidimensional index sets. There the situation is much less clear than for 1-dimensional index sets.

- Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. In *Proceedings of the 23rd Annual Symposium on Computational Geometry (SoCG)*, Gyeongju, South Korea, June 2007.
- Gunnar Carlsson, Gurjeet Singh, and Afra Zomorodian. Computing multidimensional persistence. *Journal of Computational Geometry*, 1(1):72–100, 2010.
- Michael Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Foundations of Computational Mathematics*, 15(3):613–650, 2015.

Categorified Persistence

Finally, the following articles build more abstract versions of the theory using the language of category theory. They complement the ideas developed in the present book, providing an abstract framework that some readers may find congenial.

- Peter Bubenik and Jonathan A. Scott. Categorification of persistent homology. *Discrete & Computational Geometry*, 51:600–627, 2013.
- Peter Bubenik, Vin de Silva, and Jonathan Scott. Metrics for generalized persistence modules. *Foundations of Computational Mathematics*, 15(6):1501–1531, 2015.

1.6 Organisation

The remaining chapters of the monograph are organised as follows.

Chapter 2 introduces the basic properties of persistence modules. These can be defined over any partially ordered set; we are primarily interested in persistence modules over the real line. In the best case a persistence module can be expressed as a direct sum of ‘interval modules’, which can be thought of as the atomic building blocks of the theory. Decorated real numbers are introduced here. Not all persistence modules decompose into interval modules, so we spend much of the monograph developing techniques that work without this assumption. These techniques depend on a thorough understanding of finitely-indexed persistence modules known as 'A_n-quiver representations' [35, 27]. We introduce a special notation for performing calculations on these quiver representations. This 'quiver calculus' is used throughout.[2]

Chapter 3 addresses the question of how to define the diagram of a persistence module. This is easy for modules which decompose into intervals. To handle the general case, we establish an equivalence between diagrams and a certain kind of measure defined on rectangles in the plane. Whenever a persistence diagram is sought, therefore, it suffices to construct the corresponding persistence measure. Theorems about a diagram can be replaced by simpler-to-prove theorems about its measure. The diagram exists wherever the measure takes finite values. This leads to several different notions of 'tameness'. There are large classes of examples of naturally occurring persistence modules which are tame enough for their diagrams to be defined everywhere or almost everywhere. Some elementary 'vanishing lemmas' facilitate the explicit calculation of persistence modules. We finish by showing how our abstractly defined diagrams agree with the diagrams that are produced by the standard algorithms [30, 50] when working in finite situations derived from real data.

Chapter 4 is concerned with interleavings. An interleaving is an approximate isomorphism between two persistence modules. They occur naturally in applications when the input data are known only up to some bounded error. After presenting the basic properties, we give a clean proof of the technical lemma (from [15]) that two interleaved modules can be interpolated by a 1-parameter family.

Chapter 5 is devoted to the isometry theorem, which asserts that the interleaving distance between two persistence modules is equal to the bottleneck distance between their persistence diagrams. The two inequalities that comprise this result are treated separately. One direction is the celebrated stability theorem of [19]. The more recent converse inequality appears in [42]. We formulate the stability theorem as a statement about measures and their diagrams. The proof of this more abstract result closely follows the original proof in [19]. Our version of the isometry theorem supposes that the persistence modules are q-tame. We also provide a more general version of the stability theorem which allows us to compare diagrams of persistence modules with no assumptions on their tameness: wherever the two diagrams are defined, they must be close to each other.

Chapter 6, finally, contains two worked examples. We show how the theory developed in this monograph can be used in practice to define various forms of persistence and prove the needed theorems and lemmas. We hope these examples illustrate the strength and flexibility of our approach.

[2] Readers who wish to adopt our notation are invited to contact us for the LaTeX macros.

1.7 Multisets

Persistence diagrams are multisets rather than sets. For our purposes, a multiset is a pair $\mathsf{A} = (S, \mathrm{m})$ where S is a set and

$$\mathrm{m} : S \to \{1, 2, 3, \dots\} \cup \{\infty\}$$

is the multiplicity function, which tells us how many times each element of S occurs in A. Here are our conventions regarding multisets:

- The cardinality of $\mathsf{A} = (S, \mathrm{m})$ is defined to be

$$\operatorname{card} \mathsf{A} = \sum_{s \in S} \mathrm{m}(s)$$

  which takes values in $\{0, 1, 2, \dots\} \cup \{\infty\}$. We do not distinguish between different infinite cardinals.
- We never form the intersection of two multisets, but we will sometimes restrict a multiset A to a set B:

$$\mathsf{A}|_B = (S \cap B, \mathrm{m}|_{S \cap B})$$

  We may write this as $\mathsf{A} \cap B$ when $\mathsf{A}|_B$ is typographically inconvenient.
- A pair (B, m) where

$$\mathrm{m} : B \to \{0, 1, 2, \dots\} \cup \{\infty\}$$

  is implicitly regarded as defining a multiset $\mathsf{A} = (S, \mathrm{m}|_S)$ where $S = B - \mathrm{m}^{-1}(0)$ is the support of m.
- If $\mathsf{A} = (S, \mathrm{m})$ is a multiset and $f : S \to B$ where B is a set, then the notation

$$\{f(a) \mid a \in \mathsf{A}\}$$

  is interpreted as the multiset in B with multiplicity function

$$\mathrm{m}'(b) = \sum_{s \in f^{-1}(b)} \mathrm{m}(s)$$

Except in definitions like these, we seldom refer explicitly to S.

Chapter 2
Persistence Modules

All vector spaces are taken to be over an arbitrary field $\mathbf{k}$, fixed throughout.

2.1 Persistence Modules Over a Real Parameter

A persistence module $\mathbb{V}$ over the real numbers $\mathbf{R}$ is defined to be an indexed family of vector spaces

$$(V_t \mid t \in \mathbf{R}),$$

and a doubly-indexed family of linear maps

$$(v_t^s : V_s \to V_t \mid s \leq t)$$

which satisfy the composition law

$$v_t^s \circ v_s^r = v_t^r$$

whenever $r \leq s \leq t$, and where v_t^t is the identity map on V_t.

Remark 2.1 Equivalently, a persistence module is a functor [43] from the real line (viewed as a category with a unique morphism $s \to t$ whenever $s \leq t$) to the category of vector spaces. The uniqueness of the morphism $s \to t$ corresponds to the fact that all possible compositions

$$v_t^{s_{n-1}} \circ v_{s_{n-1}}^{s_{n-2}} \circ \cdots \circ v_{s_2}^{s_1} \circ v_{s_1}^{s}$$

from V_s to V_t are equal to each other, and in particular to v_t^s.

Here is the standard class of examples from topological data analysis. Let X be a topological space and let $f : X \to \mathbf{R}$ be a function (not necessarily continuous). Consider the sublevelsets:

F. Chazal et al., *The Structure and Stability of Persistence Modules*,
SpringerBriefs in Mathematics, DOI 10.1007/978-3-319-42545-0_2

$$X^t = (X, f)^t = \{x \in X \mid f(x) \le t\}$$

The inclusion maps $(i_t^s : X^s \to X^t \mid s \le t)$ trivially satisfy the composition law and the identity map condition. Collectively this information is called the **sublevelset filtration** of (X, f) and we call it $\mathbb{X}_{\text{sub}}$ or $\mathbb{X}_{\text{sub}}^f$.

Remark 2.2 Here we are using **closed sublevelsets** $\{x \in X \mid f(x) \le t\}$, but one might instead choose to work with **open sublevelsets** $\{x \in X \mid f(x) < t\}$.

We can obtain a persistence module by applying to this filtration any functor from topological spaces to vector spaces. For example, let $\mathrm{H} = \mathrm{H}_k(-; \mathbf{k})$ be the functor 'k-dimensional singular homology with coefficients in $\mathbf{k}$'. We define a persistence module $\mathbb{V}$ by setting

$$V_t = \mathrm{H}(X^t) \quad \text{and} \quad v_t^s = \mathrm{H}(i_t^s) : V_s \to V_t$$

using the fact that functors operate on maps as well as objects. We can express this definition concisely by writing $\mathbb{V} = \mathrm{H}(\mathbb{X}_{\text{sub}})$.

In the applied topology literature, there are many examples (X, f) whose persistent homology is of interest. Very often X is a finite simplicial complex and each X^t is a subcomplex. It follows that the vector spaces $\mathrm{H}(X^t)$ are finite-dimensional; and as t increases there are finitely many 'critical values' at which the complex changes, growing by one or more new cells. Suppose these critical values are

$$a_1 < a_2 < \cdots < a_n.$$

Then all the information in the persistence module is contained in the finite diagram

$$\mathrm{H}(X^{a_1}) \longrightarrow \mathrm{H}(X^{a_2}) \longrightarrow \ldots \longrightarrow \mathrm{H}(X^{a_n})$$

of finite-dimensional vector spaces and linear maps. In this situation,

- the isomorphism type of $\mathrm{H}(\mathbb{X}_{\text{sub}})$ admits a compact description [30, 50];
- there is a fast algorithm for computing this description [30, 50];
- the description is continuous (indeed 1-Lipschitz) with respect to f [19].

This description is the famous **persistence diagram**, or **barcode**. It encodes the structure of the diagram as a list of intervals of the form $[b, d) = [a_i, a_j)$ or $[a_i, +\infty)$. Each such interval represents a 'feature' that is 'born' at b and 'dies' at d.

There are good grounds for extending the results of [30, 50, 19] beyond the case of finite diagrams. For example, theoretical guarantees are commonly formulated in terms of an idealised model; for instance the sampled data may be an approximation to an underlying continuous space. Finiteness becomes unnatural and difficult to enforce in these ideal models, but one still wants the main results to be true.

Here is (what we believe to be) a good notion of tameness: a persistence module $\mathbb{V}$ is **q-tame** if

$$\mathrm{r}_t^s = \mathrm{rank}(v_t^s) < \infty \quad \text{whenever } s < t.$$

The definition is taken from [15], where such modules are simply called 'tame'. Since that word is overloaded with too many different meanings in the persistence literature, we say 'q-tame' instead (see Sect. 3.8 for the etymology).

It is shown in [15] that persistence diagrams can be constructed for q-tame persistence modules, and that these diagrams are stable with respect to certain natural metrics. We reproduce these results here, using different methods for many of the arguments. We complete the picture by showing that the map from q-tame persistence modules to persistence diagrams is an isometry. This isometry theorem is due to Lesnick [42].

We believe that q-tame persistence modules are a good class of objects for two complementary reasons: (i) we can prove almost everything we want to prove about q-tame modules and their persistence diagrams; and (ii) they occur in practice. For example, a continuous function on a finite simplicial complex has q-tame sublevelset persistent homology (Theorem 3.33). See [16] for many other examples.

2.2 Index Posets

We can define a persistence module over any partially ordered set, or *poset*, $\mathbf{T}$, in the same way as for $\mathbf{R}$, by specifying indexed families

$$(V_t \mid t \in \mathbf{T}) \quad \text{and} \quad (v_t^s \mid s, t \in \mathbf{T},\ s \leq t)$$

of vector spaces and linear maps, for which $v_t^s \circ v_s^r = v_t^r$ whenever $r \leq s \leq t$, and where v_t^t is the identity on V_t. The resulting collection of data is called a **T-persistence module** or a persistence module **over T**.

If $\mathbb{V}$ is a $\mathbf{T}$-persistence module and $\mathbf{S} \subset \mathbf{T}$, then we get an $\mathbf{S}$-persistence module by considering only those spaces and maps with indices in $\mathbf{S}$. This is the **restriction** of $\mathbb{V}$ to $\mathbf{S}$, and may be written $\mathbb{V}_{\mathbf{S}}$ or $\mathbb{V}|_{\mathbf{S}}$. Most commonly, we work with finite subsets $\mathbf{T} \subset \mathbf{R}$. We collect information about an $\mathbf{R}$-persistence module by considering its restriction to different finite subsets. This works well because persistence modules over $\{1, 2, \ldots, n\}$ are well understood.

In Chap. 4, we make use of certain posets that are subsets of $\mathbf{R}^2$.

2.3 Module Categories

A homomorphism Φ between two $\mathbf{T}$-persistence modules $\mathbb{U}, \mathbb{V}$ is a collection of linear maps $(\phi_t : U_t \to V_t \mid t \in \mathbf{T})$ such that the diagram

$$\begin{array}{ccc} U_s & \xrightarrow{u_t^s} & U_t \\ {\scriptstyle\phi_s}\downarrow & & \downarrow{\scriptstyle\phi_t} \\ V_s & \xrightarrow{v_t^s} & V_t \end{array}$$

commutes for all $s \leq t$. Composition is defined in the obvious way, as are identity homomorphisms. This makes the collection of persistence modules into a category. The category contains kernel, image, and cokernel objects for every map Φ, and there is a zero object. Write

$$\begin{aligned} \operatorname{Hom}(\mathbb{U}, \mathbb{V}) &= \{\text{homomorphisms } \mathbb{U} \to \mathbb{V}\}, \\ \operatorname{End}(\mathbb{V}) &= \{\text{homomorphisms } \mathbb{V} \to \mathbb{V}\}. \end{aligned}$$

Note that $\operatorname{End}(\mathbb{V})$ is a $\mathbf{k}$-algebra. Later we consider homomorphisms that shift the index, in order to define the interleaving relation between persistence modules.

2.4 Interval Modules

The building blocks of persistence are the interval modules. One seeks to understand a persistence module by decomposing it into intervals. This is not always possible, but it is sufficiently possible for our purposes.

An **interval** in a totally ordered set $\mathbf{T}$ is a subset $J \subseteq \mathbf{T}$ such that if $r \in J$ and $t \in J$ and $r < s < t$ then $s \in J$. For any nonempty interval $J \subseteq \mathbf{T}$, the **interval module** $\mathbb{I} = \mathbf{k}^J$ is defined to be the $\mathbf{T}$-persistence module with vector spaces

$$I_t = \begin{cases} \mathbf{k} \text{ if } t \in J \\ 0 \text{ otherwise} \end{cases}$$

and linear maps

$$i_t^s = \begin{cases} 1 \text{ if } s, t \in J \\ 0 \text{ otherwise} \end{cases}$$

In informal language, the module $\mathbf{k}^J$ represents a 'feature' which 'persists' over exactly the interval J and nowhere else. We write $\mathbf{k}^J_{\mathbf{T}}$ when we wish to name the index set explicitly.

Intervals in a finite set $\mathbf{T} = \{a_0 < a_1 < \cdots < a_n\}$ are usually written as closed intervals $[a_i, a_j]$, and sometimes as half-open intervals $[a_i, a_{j+1})$ with the convention that $a_{n+1} = +\infty$. We often lower the superscript when naming the corresponding modules, writing $\mathbf{k}[a_i, a_j]$ rather than $\mathbf{k}^{[a_i,a_j]}$ for ease of reading.

Intervals in the real line $\mathbf{R}$ merit a special notation of their own. Each non-empty real interval has endpoints (possibly $\pm\infty$) defined by its infimum and supremum, and it may or may not attain its finite endpoints. To distinguish the various cases, we introduce **decorated reals**, written as ordinary real numbers with a superscript $^+$ (plus) or $^-$ (minus). For finite intervals we adopt the following dictionary:

$$\begin{aligned}
(p^-, q^-) &\quad \text{means} \quad [p, q) \\
(p^-, q^+) &\quad \text{means} \quad [p, q] \\
(p^+, q^-) &\quad \text{means} \quad (p, q) \\
(p^+, q^+) &\quad \text{means} \quad (p, q]
\end{aligned}$$

We require $p < q$ except for the special case (r^-, r^+) which represents the 1-point interval $[r, r]$. For infinite intervals we use the symbols $-\infty^+$ and $+\infty^-$:

$$\begin{aligned}
(-\infty^+, q^-) &\quad \text{means} \quad (-\infty, q) \\
(-\infty^+, q^+) &\quad \text{means} \quad (-\infty, q] \\
(p^-, +\infty^-) &\quad \text{means} \quad [p, +\infty) \\
(p^+, +\infty^-) &\quad \text{means} \quad (p, +\infty) \\
(-\infty^+, +\infty^-) &\quad \text{means} \quad (-\infty, +\infty)
\end{aligned}$$

When we wish to refer to a decorated number but don't know what the decoration is, we use an asterisk. Thus p^* means p^+ or p^-.

The collection of decorated and undecorated numbers is totally ordered by setting

$$p^- < p < p^+ < q^- < q < q^+ \tag{2.1}$$

for all $p < q$. One advantage of doing this is that nonempty real intervals now correspond exactly to pairs (p^*, q^*) such that $-\infty < p^* < q^* < +\infty$, with the single statement

$$(p^*, q^*) \quad \text{means} \quad \{t \in \mathbf{R} \mid p^* < t < q^*\} \tag{2.2}$$

uniformly replacing the nine dictionary definitions given above. Sometimes it is helpful to extend membership of a real interval to decorated real numbers. We adopt the convention that

$$t^* \in (p^*, q^*) \quad \text{means} \quad p^* < t^* < q^* \tag{2.3}$$

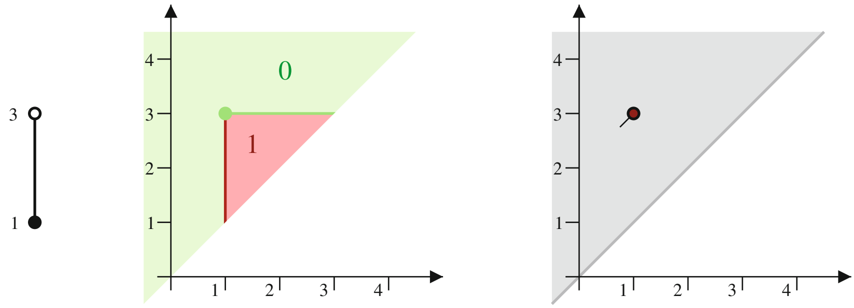


Fig. 2.1 The interval (*left*), rank function (*middle*), and decorated point (*right*) representations of the interval module $\mathbf{k}[1, 3) = \mathbf{k}(1^-, 3^-)$

for any t^* and (p^*, q^*). The interval itself continues to be a set of undecorated real numbers; we are simply overloading the symbol '$\in$' with an additional meaning.

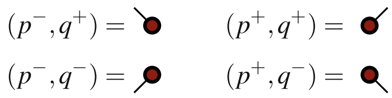
We finish with some visual conventions for interval modules over $\mathbf{R}$. Let

$$\mathcal{H} = \{(p, q) \mid p \leq q\}$$

be the half-plane of points in $\mathbf{R}^2$ which lie on or above the diagonal. A finite interval module $\mathbf{k}(p^*, q^*)$ may be represented in several different ways (see Fig. 2.1):

- as an interval in the real line;
- as a function $\mathcal{H} \to \{0, 1\}$, defined by $(s, t) \mapsto \operatorname{rank}(i_t^s)$;
- as a point (p, q) in $\mathcal{H}$, with a tick to specify the decoration.

Here are the four tick directions explicitly:



The convention is that the tick points into the quadrant suggested by the decorations. We can represent infinite intervals by working in the extended half-plane

$$\overline{\mathcal{H}} \;=\; \mathcal{H} \;\cup\; \{-\infty\} \times \mathbf{R} \;\cup\; \mathbf{R} \times \{+\infty\} \;\cup\; \{(-\infty, +\infty)\}.$$

This can be drawn schematically as a triangle; see Fig. 2.2.

Remark 2.3 Persistence diagrams have traditionally been drawn without ticks. This is adequate for most purposes, and indeed in most traditional examples the intervals that occur are half-open intervals $[p, q) = (p^-, q^-)$ and there is no need to consider

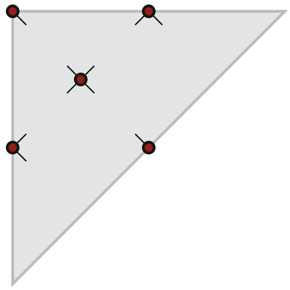


Fig. 2.2 The extended half-plane $\overline{\mathcal{H}}$ with examples of each interval type drawn as points with ticks. Points on the *left* and *top* edges correspond to intervals that are unbounded below and above, respectively. Points on the *diagonal* correspond to singleton intervals $(r^-, r^+) = [r, r] = \{r\}$

other possibilities. In the present work, the extra precision provided by decorations is essential to the correspondence between diagrams and measures.

2.5 Interval Decomposition

The **direct sum** $\mathbb{W} = \mathbb{U} \oplus \mathbb{V}$ of two persistence modules $\mathbb{U}$, $\mathbb{V}$ is defined as follows:

$$W_t = U_t \oplus V_t, \quad w_t^s = u_t^s \oplus v_t^s$$

This generalises immediately to arbitrary (finite or infinite) direct sums.

A persistence module $\mathbb{W}$ is **indecomposable** if the only decompositions $\mathbb{W} = \mathbb{U} \oplus \mathbb{V}$ are the trivial decompositions $\mathbb{W} \oplus 0$ and $0 \oplus \mathbb{W}$.

Direct sums play both a *synthetic* role and an *analytic* role in our theory. On the one hand, given an indexed family of intervals $(J_\ell \mid \ell \in L)$ we can synthesise a persistence module

$$\mathbb{V} = \bigoplus_{\ell \in L} \mathbf{k}^{J_\ell}$$

whose isomorphism type depends only on the multiset $\{J_\ell \mid \ell \in L\}$. In light of the direct-sum decomposition, we can think of $\mathbb{V}$ as having an independent feature for each $\ell \in L$, supported over the interval J_ℓ. On the other hand, we can attempt to analyse a given persistence module $\mathbb{V}$ by decomposing it into submodules isomorphic to interval modules.

Remark 2.4 The decomposition of a persistence module is frequently described in metaphorical terms. The index $t \in \mathbf{R}$ is interpreted as 'time'. Each interval summand $\mathbf{k}^J$ represents a 'feature' of the module which is 'born' at time $\inf(J)$ and 'dies' at time $\sup(J)$.

We now present the necessary theory. A 'building block' in a module category can be characterised by having a comparatively simple endomorphism ring. Interval modules have the simplest possible:

Proposition 2.5 *Let* $\mathbb{I} = \mathbf{k}^J_{\mathbf{T}}$ *be an interval module over* $\mathbf{T} \subseteq \mathbf{R}$*; then* $\operatorname{End}(\mathbb{I}) = \mathbf{k}$.

Proof Any endomorphism of $\mathbb{I}$ acts on each nonzero $I_t = \mathbf{k}$ by scalar multiplication. By the commutative square for morphisms, it is the same scalar for each t. □

Proposition 2.6 *Interval modules are indecomposable.*

Proof Given a decomposition $\mathbb{I} = \mathbb{U} \oplus \mathbb{V}$, the projection maps onto $\mathbb{U}$ and $\mathbb{V}$ are idempotent endomorphisms.[1] The only idempotents in $\operatorname{End}(\mathbb{I}) = \mathbf{k}$ are 0 and 1. □

Theorem 2.7 (Krull–Remak–Schmidt–Azumaya) *Suppose a persistence module over* $\mathbf{T} \subseteq \mathbf{R}$ *can be expressed as a direct sum of interval modules in two different ways:*

$$\mathbb{V} \cong \bigoplus_{\ell \in L} \mathbf{k}^{J_\ell} \cong \bigoplus_{m \in M} \mathbf{k}^{K_m}$$

Then there is a bijection $\sigma : L \to M$ *such that* $J_\ell = K_{\sigma(\ell)}$ *for all* ℓ.

Proof This is from Azumaya [2] (Theorem 1), along with the trivial observation that $\mathbf{k}^J \cong \mathbf{k}^K$ implies $J = K$. The theorem requires a 'locality' condition on the endomorphism ring of each possible interval module: if $\alpha, \beta \in \operatorname{End}(\mathbb{I})$ are non-isomorphisms then $\alpha + \beta$ is a non-isomorphism. Since each $\operatorname{End}(\mathbb{I}) = \mathbf{k}$, the only non-isomorphism is the zero map and the condition is satisfied. □

In other words, provided we can decompose a given persistence module $\mathbb{V}$ as a direct sum of interval modules, then the multiset of intervals is an isomorphism invariant of $\mathbb{V}$. But when does such a decomposition exist?

Theorem 2.8 (Gabriel, Auslander, Ringel–Tachikawa, Webb, Crawley-Boevey) *Let* $\mathbb{V}$ *be a persistence module over* $\mathbf{T} \subseteq \mathbf{R}$. *Then* $\mathbb{V}$ *can be decomposed as a direct sum of interval modules in either of the following situations:*

(1) $\mathbf{T}$ *is a finite set; or*
(2) *each* V_t *is finite-dimensional.*

On the other hand, (3) *there exists a persistence module over* $\mathbf{Z}$ *(indeed, over the nonpositive integers) which does not admit an interval decomposition.*

Proof (1) The decomposition of a diagram

$$V_1 \longrightarrow V_2 \longrightarrow \dots \longrightarrow V_n$$

into interval summands, when each $\dim(V_i)$ is finite, is one of the simpler instances of Gabriel's theorem [35]; see [50] or [8] for a concrete explanation. The extension to

[1] An endomorphism e is idempotent if it satisfies $ee = e$.

infinite-dimensional modules follows abstractly from a theorem of Auslander [1] and, independently, Ringel and Tachikawa [47]. Alternatively, observe that the argument given in [8] does not require finite-dimensionality (although it is presented as such).

(2) The result for $\mathbf{T} = \mathbf{Z}$, and therefore for any locally finite $\mathbf{T} \subset \mathbf{R}$, follows from Propositions 2 and 3 and Theorem 3 of Webb [48]. This was generalised to $\mathbf{T} = \mathbf{R}$, and therefore to any $\mathbf{T} \subseteq \mathbf{R}$, more recently by Crawley-Boevey [25].

(3) Webb [48] gives this example, indexed over the nonpositive integers $-\mathbf{N}$:

$$\begin{aligned} W_0 &= \{\text{sequences } (x_1, x_2, x_3, \dots) \text{ of scalars}\} \\ W_{-n} &= \{\text{such sequences with } x_1 = \cdots = x_n = 0\} \qquad (n \geq 1) \end{aligned}$$

The w_{-n}^{-m} are taken to be the canonical inclusion maps. We can succinctly describe this module as an infinite product $\mathbb{W} = \prod_{n \geq 0} \mathbf{k}[-n, 0]$.

Suppose $\mathbb{W}$ has an interval decomposition. Since each map w_{-n}^{-n-1} is injective, all of the intervals must be of the form $[-n, 0]$ or $(-\infty, 0]$. The multiplicity of $[-n, 0]$ may then be calculated as $\dim(W_{-n}/W_{-n-1}) = 1$. The multiplicity of $(-\infty, 0]$ is zero, because any summand of that type requires a nonzero element of W_0 that is in the image of w_0^{-n} for all $n \geq 0$, but $\bigcap_{n \geq 0} W_{-n} = \{0\}$ so such an element doesn't exist. All of this implies that $\mathbb{W} \cong \bigoplus_{n \geq 0} \mathbf{k}[-n, 0]$. This contradicts the fact that $\dim(W_0)$ is uncountable[2] so $\mathbb{W}$ does not admit an interval decomposition after all. $\square$

In Examples 3.31 and 3.40, we show what we *can* do with the Webb module.

Remark 2.9 Here are other examples of persistence modules that lack an interval decomposition. Crawley-Boevey [24] proposed the infinite product $\prod_{n \geq 1} \mathbf{k}[0, 1/n]$. A dimension count implies that any interval decomposition must include uncountably many copies of $\mathbf{k}[0, 0]$, but this contradicts the fact that $\bigcap_{t>0} \ker(v_t^0)$ is trivial. Nor is uncountable dimensionality a necessary feature. Lesnick [41] proposed the following example that has countable dimension at every index in $-\mathbf{N}$:

$$\begin{aligned} L_0 &= \mathbf{k} \\ L_{-1} &= \{\text{eventually-zero sequences } (x_1, x_2, x_3, \dots) \text{ of scalars}\} \\ L_{-n} &= \{\text{such sequences with } x_1 = \cdots = x_{n-1} = 0\} \qquad (n \geq 2) \end{aligned}$$

The ℓ_{-n}^{-m} are taken to be the canonical inclusion maps when $n \geq 1$, while ℓ_0^{-m} is the 'augmentation map' which takes the sum of the entries of the sequence. Given an interval decomposition, consider the unique summand that meets L_0 nontrivially. We can rule out $\mathbf{k}(-\infty, 0]$ since $\bigcap_{n \geq 1} L_{-n} = \{0\}$, so it is isomorphic to some $\mathbf{k}[-m, 0]$. No other summands reach L_0, so ℓ_0^{-m-1} must be the zero map; but it isn't.

[2] No countable sequence of vectors $w_1, w_2, w_3, \dots$ can span W_0. Consider a vector $x = (x_1, x_2, x_3, \dots)$ where for all $k \geq 1$ the $k+1$ terms $x_{k^2}, x_{k^2+1}, \dots, x_{k^2+k}$ have been chosen to guarantee that x is not a linear combination of $w_1, \dots, w_k$. Then $x \notin \operatorname{span}(w_1, w_2, w_3, \dots)$.

For a persistence module which decomposes into intervals, the way is now clear to define its persistence diagram: it is simply the list of intervals, with multiplicity, that occur in the decomposition. Theorem 2.7 tells us that this multiset is an isomorphism invariant.

Given that an arbitrary persistence module over **R** is not guaranteed an interval decomposition, here are three possible ways to proceed:

- Work in restricted settings to ensure that the structure of $\mathbb{V}$ depends only on finitely many index values $t \in \mathbf{R}$. For example, if X is a compact manifold and f is a Morse function, then $\mathrm{H}(\mathbb{X}_{\mathrm{sub}})$ is determined by the finite sequence

$$\mathrm{H}(X^{a_1}) \longrightarrow \mathrm{H}(X^{a_2}) \longrightarrow \dots \longrightarrow \mathrm{H}(X^{a_n})$$

  where $a_1, a_2, \dots, a_n$ are the critical values of f. This is the traditional approach. In this setting, the word 'tame' typically refers to pairs (X, f) for which $\mathrm{H}(\mathbb{X}_{\mathrm{sub}})$ is determined by a finite diagram of finite-dimensional vector spaces.
- Sample the persistence module $\mathbb{V}$ over a finite grid. Consider limits as the grid converges to the whole real line. This is the strategy adopted in [15], where it is shown that the q-tame hypothesis is sufficient to guarantee good limiting behaviour.
- Show that the persistence intervals (in the decomposable case) can be inferred from the behaviour of $\mathbb{V}$ on finite index sets.[3] Apply this indirect definition to define the persistence diagram in the non-decomposable case. This is the method of 'rectangle measures' developed in this monograph.

The first method is adequate for computational applications, at least on a first pass. The second and third methods both entail a certain amount of analytic work. The advantage of the third method is that this work is black-boxed as a technical result (Theorem 3.12) that allows one to move freely between rectangle measures and their corresponding persistence diagrams. The end-user is protected from the analytic details.

2.6 The Decomposition Persistence Diagram

If a persistence module $\mathbb{V}$ indexed over **R** can be decomposed

$$\mathbb{V} \cong \bigoplus_{\ell \in L} \mathbf{k}(p_\ell^*, q_\ell^*)$$

then we define the **decorated persistence diagram** to be the multiset

$$\mathsf{Dgm}(\mathbb{V}) = \mathsf{Int}(\mathbb{V}) = \{(p_\ell^*, q_\ell^*) \mid \ell \in L\}$$

[3] We consider index sets of length 4 to define the persistence measure, length 5 to show that it is additive, and length 8 to prove the stability theorem.

and the **undecorated persistence diagram** to be the multiset

$$\mathsf{dgm}(\mathbb{V}) = \mathsf{int}(\mathbb{V}) = \{(p_\ell, q_\ell) \mid \ell \in L\} - \Delta$$

where $\Delta = \{(r, r) \mid r \in \mathbf{R}\}$ is the diagonal in the plane.

These are the **decomposition persistence diagrams**. In Sect. 3.7 we give a different definition of persistence diagrams based on the persistence measure. Often they coincide, but occasionally we need to distinguish them. In that case we use the alternate names Int, int for the diagrams defined here.

Theorem 2.7 implies that $\mathsf{Dgm}(\mathbb{V})$ and $\mathsf{dgm}(\mathbb{V})$ are independent of the choice of decomposition of $\mathbb{V}$. Notice that Dgm is a multiset of decorated points in $\overline{\mathcal{H}}$, whereas dgm is a multiset of undecorated points in the interior of $\overline{\mathcal{H}}$. Here 'interior' means that we exclude the diagonal but keep the points at infinity. The information retained by dgm is the information we care about later when we discuss bottleneck distances. See Chap. 5.

Example 2.10 Consider the curve in $\mathbf{R}^2$ shown in Fig. 2.3, filtered by the height function. The topology (that is, the homotopy type) of the sublevelsets of f is empty over $(-\infty, a)$ and constant over the intervals $[a, b)$, $[b, c)$, $[c, d)$, $[d, e)$, $[e, f)$ and $[f, +\infty)$, so it is enough to consider the 6-term persistence modules obtained by restricting $\mathrm{H}_*(\mathbb{X}_{\mathrm{sub}})$ to the six critical values.

$$\begin{array}{ll} \mathrm{H}_0: & \mathbf{k} \longrightarrow \mathbf{k}^2 \longrightarrow \mathbf{k} \longrightarrow \mathbf{k}^2 \longrightarrow \mathbf{k} \longrightarrow \mathbf{k} \\ \mathrm{H}_1: & 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow \mathbf{k} \end{array}$$

To decompose the H_0 diagram we need knowledge of the maps. Let $[a]$, $[b]$, $[d]$ denote the 0-homology classes associated to the connected components born at times a, b, d respectively. When two components merge at index c we get the

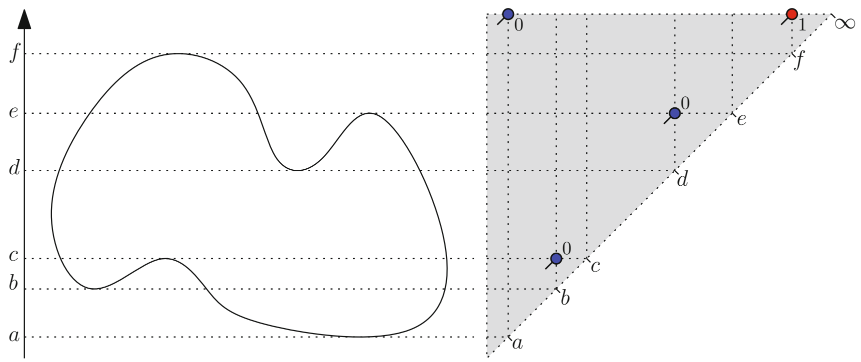


Fig. 2.3 A traditional example. *Left:* X is a smoothly embedded curve in the plane, and f is its y-coordinate or 'height' function. *Right:* The decorated persistence diagram of $\mathrm{H}(\mathbb{X}_{\mathrm{sub}})$: there are three intervals in H_0 (*blue dots*, marked 0) and one interval in H_1 (*red dot*, marked 1)

relation $[a] = [b]$. This becomes $[a] = [b] = [d]$ at index e. It follows that $\mathrm{H}_0(\mathbb{X}_{\mathrm{sub}})$ decomposes as follows.

$$\begin{array}{ll}
[a] \quad : & \mathbf{k} \longrightarrow \mathbf{k} \longrightarrow \mathbf{k} \longrightarrow \mathbf{k} \longrightarrow \mathbf{k} \longrightarrow \mathbf{k} \\
[b]-[a]: & 0 \longrightarrow \mathbf{k} \longrightarrow 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow 0 \\
[d]-[a]: & 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow \mathbf{k} \longrightarrow 0 \longrightarrow 0
\end{array}$$

The generator of each summand is shown on the left. Each generator has precisely the lifetime indicated by its interval module, and at each index the existing surviving generators form a basis for the homology at that index.

The 1-homology is already an interval module with no further decomposition necessary. It is generated by the 1-cycle $[f]$ which appears at index f:

$$[f] \quad : \qquad 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow \mathbf{k}$$

It follows that the full persistent homology of $\mathbb{X}_{\mathrm{sub}}$ looks like this:

$$\begin{aligned}
\mathrm{H}_0(\mathbb{X}_{\mathrm{sub}}) &\cong \mathbf{k}[a, +\infty) \oplus \mathbf{k}[b, c) \oplus \mathbf{k}[d, e) \\
\mathrm{H}_1(\mathbb{X}_{\mathrm{sub}}) &\cong \mathbf{k}[f, +\infty)
\end{aligned}$$

The decorated persistence diagram is drawn in Fig. 2.3.

Remark 2.11 For a Morse function on a compact manifold with critical values (a_i), the intervals are always half-open, of type $[a_i, a_j) = (a_i^-, a_j^-)$, since the homotopy type of the sublevelsets is constant over the intervals $[a_i, a_{i+1})$. Compare Sect. 3.12. The persistence algorithm of Edelsbrunner, Letscher and Zomorodian [30], later presented in general form by Zomorodian and Carlsson [50], computes the interval decomposition and therefore persistence diagram for any example of this type.

Remark 2.12 Because of torsion phenomena in homology, different choices of field $\mathbf{k}$ can lead to different persistence diagrams for a given geometric object.

2.7 Quiver Calculations

We now set up the notation and algebraic tools for handling persistence modules over a finite index set.

A persistence module $\mathbb{V}$ indexed over a finite subset

$$\mathbf{T}: \quad a_1 < a_2 < \cdots < a_n$$

of the real line can be thought of as a diagram of n vector spaces and $n - 1$ linear maps:

$$\mathbb{V}: \quad V_{a_1} \longrightarrow V_{a_2} \longrightarrow \cdots \longrightarrow V_{a_n}$$

Such a diagram is a **representation** of the following **quiver**:

$$\bullet \longrightarrow \bullet \longrightarrow \cdots \longrightarrow \bullet$$

We have seen (Theorem 2.8) that $\mathbb{V}$ decomposes as a finite sum of interval modules $\mathbf{k}[a_i, a_j]$. When n is small, we can represent these interval modules pictorially. The following example illustrates how.

Example 2.13 Let $a < b < c$. There are six interval modules over $\{a, b, c\}$, namely:

$$\mathbf{k}[a,a] = \bullet_a \!-\!\!\!-\! \circ_b \!-\!\!\!-\! \circ_c \qquad \mathbf{k}[a,b] = \bullet_a \!-\!\!\!-\! \bullet_b \!-\!\!\!-\! \circ_c \qquad \mathbf{k}[a,c] = \bullet_a \!-\!\!\!-\! \bullet_b \!-\!\!\!-\! \bullet_c$$
$$\mathbf{k}[b,b] = \circ_a \!-\!\!\!-\! \bullet_b \!-\!\!\!-\! \circ_c \qquad \mathbf{k}[b,c] = \circ_a \!-\!\!\!-\! \bullet_b \!-\!\!\!-\! \bullet_c$$
$$\mathbf{k}[c,c] = \circ_a \!-\!\!\!-\! \circ_b \!-\!\!\!-\! \bullet_c$$

A filled circle $\bullet$ indicates a copy of the 1-dimensional vector space $\mathbf{k}$; an empty circle $\circ$ indicates the zero vector space. A map between two filled circles is always the identity; all other maps are by necessity zero.

Now let $\mathbb{V}$ be a persistence module indexed over $\mathbf{R}$. For any finite set of indices

$$\mathbf{T}: \quad a_1 < a_2 < \cdots < a_n$$

and any interval $[a_i, a_j] \subseteq \mathbf{T}$, we define the multiplicity of $[a_i, a_j]$ in $\mathbb{V}_\mathbf{T}$ to be the number of copies of $\mathbf{k}[a_i, a_j]$ to occur in the interval decomposition of $\mathbb{V}_\mathbf{T}$. This takes values in the set $\{0, 1, 2, \ldots, \infty\}$. (We do not distinguish different infinite cardinals.)

It is useful to have notation for these multiplicities. Again, we define by example.

Example 2.14 We write

$$\langle [b,c] \mid \mathbb{V}_{a,b,c} \rangle \quad \text{or} \quad \langle\, \circ_a \!-\!\!\!-\! \bullet_b \!-\!\!\!-\! \bullet_c \mid \mathbb{V} \,\rangle$$

for the multiplicity of $\circ_a \!-\!\!\!-\! \bullet_b \!-\!\!\!-\! \bullet_c$ in the following 3-term module:

$$\mathbb{V}_{a,b,c}: \quad V_a \longrightarrow V_b \longrightarrow V_c$$

When $\mathbb{V}$ is clear from the context, we may simply write

$$\langle\, \circ_a \!-\!\!\!-\! \bullet_b \!-\!\!\!-\! \bullet_c \,\rangle.$$

The abbreviation $\langle [b,c] \rangle$ is not permitted since it is ambiguous. For example, $\langle [b,c] \mid \mathbb{V}_{b,c} \rangle$ and $\langle [b,c] \mid \mathbb{V}_{a,b,c} \rangle$ are not generally the same. See Proposition 2.17 and Example 2.18.

Example 2.15 The invariants of a single linear map $V_a \xrightarrow{v} V_b$ are:

$$\begin{aligned}\operatorname{rank}(v) &= \langle \bullet_a\!\!-\!\!\bullet_b \mid \mathbb{V} \rangle \\ \operatorname{nullity}(v) &= \langle \bullet_a\!\!-\!\!\circ_b \mid \mathbb{V} \rangle \\ \operatorname{conullity}(v) &= \langle \circ_a\!\!-\!\!\bullet_b \mid \mathbb{V} \rangle\end{aligned}$$

To see this when V_a, V_b are finite dimensional, observe that by elementary linear algebra we can find bases

$$e_1, \ldots, e_r, f_1, \ldots, f_n \quad \text{and} \quad e'_1, \ldots, e'_r, g_1, \ldots, g_c$$

for V_a and V_b respectively, such that $v(e_i) = e'_i$ and $v(f_j) = 0$ for all i, j. The basis elements define a decomposition of the module $(V_a \xrightarrow{v} V_b)$ into interval summands of the three types

$$\big(\operatorname{span}(e_i) \longrightarrow \operatorname{span}(e'_i)\big) \quad \text{and} \quad \big(\operatorname{span}(f_j) \longrightarrow 0\big) \quad \text{and} \quad (0 \longrightarrow \operatorname{span}(g_k))$$

which are respectively isomorphic to $\bullet_a\!\!-\!\!\bullet_b$ and $\bullet_a\!\!-\!\!\circ_b$ and $\circ_a\!\!-\!\!\bullet_b$.

Proposition 2.16 (direct sums) *Suppose a persistence module* $\mathbb{V}$ *can be written as a direct sum*

$$\mathbb{V} = \bigoplus_{\ell \in L} \mathbb{V}^\ell$$

Then

$$\langle [a_i, a_j] \mid \mathbb{V}_{\mathbf{T}} \rangle = \sum_{\ell \in L} \langle [a_i, a_j] \mid \mathbb{V}^\ell_{\mathbf{T}} \rangle$$

for any index set $\mathbf{T} = \{a_1, a_2, \ldots, a_n\}$ *and interval* $[a_i, a_j] \subseteq \mathbf{T}$.

Proof Each summand $\mathbb{V}^\ell_{\mathbf{T}}$ can be decomposed separately into interval modules. Putting these together we get an interval decomposition of $\mathbb{V}_{\mathbf{T}}$. The number of summands of a given type in $\mathbb{V}_{\mathbf{T}}$ is then equal to the total number of summands of that type in all of the $\mathbb{V}^\ell_{\mathbf{T}}$. □

Often we wish to compare multiplicities of intervals in different finite restrictions of $\mathbb{V}$. The principle is very simple:

Proposition 2.17 (restriction principle) *Let* $\mathbf{S}, \mathbf{T}$ *be finite index sets with* $\mathbf{S} \subset \mathbf{T}$. *Then*

$$\langle \mathbb{I} \mid \mathbb{V}_{\mathbf{S}} \rangle = \sum_{\mathbb{J}} \langle \mathbb{J} \mid \mathbb{V}_{\mathbf{T}} \rangle$$

where the sum is over those intervals $\mathbb{J} \subseteq \mathbf{T}$ *which restrict over* $\mathbf{S}$ *to* $\mathbb{I}$.

Proof Take an arbitrary interval decomposition of $\mathbb{V}_{\mathbf{T}}$. This induces an interval decomposition of $\mathbb{V}_{\mathbf{S}}$. Summands of $\mathbb{V}_{\mathbf{S}}$ of type $\mathbb{I}$ arise precisely from those summands of $\mathbb{V}_{\mathbf{T}}$ of types $\mathbb{J}$ as above. □

Example 2.18 Suppose $a < p < b < q < c$. Then we have

$$\langle \circ_a \text{———} \bullet_b \text{———} \bullet_c \rangle = \langle \circ_a \text{———} \bullet_b \text{—} \bullet_q \text{—} \bullet_c \rangle$$
$$\langle \circ_a \text{———} \bullet_b \text{———} \bullet_c \rangle = \langle \circ_a \text{—} \circ_p \text{—} \bullet_b \text{———} \bullet_c \rangle + \langle \circ_a \text{—} \bullet_p \text{—} \bullet_b \text{———} \bullet_c \rangle$$

and

$$\langle \circ_a \text{———} \circ_b \text{———} \bullet_c \rangle = \langle \circ_a \text{—} \circ_p \text{—} \circ_b \text{———} \bullet_c \rangle$$
$$\langle \circ_a \text{———} \circ_b \text{———} \bullet_c \rangle = \langle \circ_a \text{———} \circ_b \text{—} \circ_q \text{—} \bullet_c \rangle + \langle \circ_a \text{———} \circ_b \text{—} \bullet_q \text{—} \bullet_c \rangle$$

for instance. The extra term occurs when the inserted new index occurs between a clear node and a filled node, because then there are two possible intervals which restrict to the original interval.

Example 2.19 For any finite list of indices in which a, b and later c, d occur as adjacent pairs, the restriction principle gives

$$\langle \ldots \text{—} \circ \text{—} \circ_a \text{—} \bullet_b \text{—} \bullet \text{—} \ldots \text{—} \bullet \text{—} \bullet_c \text{—} \circ_d \text{—} \circ \text{—} \ldots \rangle = \langle \circ_a \text{—} \bullet_b \text{—} \bullet_c \text{—} \circ_d \rangle.$$

When $\mathrm{rank}(V_b \to V_c) < \infty$, this observation combines with Proposition 3.6 to give an easy expression for any interval multiplicity.

We will make frequent use of the restriction principle. Here is a simple illustration, to serve as a template for similar arguments that we will encounter later on.

Example 2.20 Consider the elementary fact that $\mathrm{rank}(V_b \to V_c) \geq \mathrm{rank}(V_a \to V_d)$ when $a \leq b \leq c \leq d$. The proof using quiver notation runs as follows:

$$\begin{aligned}
\mathrm{rank}(V_b \to V_c) &= \langle \text{——} \bullet_b \text{—} \bullet_c \text{——} \rangle \\
&= \langle \bullet_a \text{—} \bullet_b \text{—} \bullet_c \text{—} \bullet_d \rangle + \text{three other terms} \\
&\geq \langle \bullet_a \text{—} \bullet_b \text{—} \bullet_c \text{—} \bullet_d \rangle \\
&= \langle \bullet_a \text{————} \bullet_d \rangle \\
&= \mathrm{rank}(V_a \to V_d)
\end{aligned}$$

The ‘three other terms’ are

$$\langle \circ_a \text{—} \bullet_b \text{—} \bullet_c \text{—} \bullet_d \rangle, \quad \langle \bullet_a \text{—} \bullet_b \text{—} \bullet_c \text{—} \circ_d \rangle, \quad \langle \circ_a \text{—} \bullet_b \text{—} \bullet_c \text{—} \circ_d \rangle$$

as indicated by the restriction principle.

Chapter 3
Rectangle Measures

For a decomposable **R**-persistence module

$$\mathbb{V} \cong \bigoplus_{\ell \in L} \mathbf{k}(p_\ell^*, q_\ell^*),$$

we have defined the decorated persistence diagram to be the multiset

$$\mathsf{Dgm}(\mathbb{V}) = \mathsf{Int}(\mathbb{V}) = \{(p_\ell^*, q_\ell^*) \mid \ell \in L\},$$

and the undecorated persistence diagram to be the multiset

$$\mathsf{dgm}(\mathbb{V}) = \mathsf{int}(\mathbb{V}) = \{(p_\ell, q_\ell) \mid \ell \in L\}.$$

If we don't know that $\mathbb{V}$ is decomposable then we have to proceed differently. We are guided by the following heuristic: if we know how many points of Dgm belong to each rectangle in the half-space, then we know Dgm itself. For persistence modules, counting points in rectangles turns out to be easy.

The language of measure theory is well suited to this argument. We will see that a persistence module defines an integer-valued measure on rectangles. If the module is tame then this measure is finite-valued, and therefore (Theorem 3.12) it is concentrated at a discrete set of points. These points, taken with their multiplicities, constitute the persistence diagram. In the case where the module is decomposable, the persistence diagram constructed this way agrees with the persistence diagram defined earlier. When the module is not known to be decomposable, we can proceed regardless.

Remark 3.1 The persistence measures that we construct are not true measures on subsets of $\mathbf{R}^2$: they are additive in the sense of tiling rather than in the usual sense of disjoint set union. The discrepancy arises when we split a rectangle into two: what

F. Chazal et al., *The Structure and Stability of Persistence Modules*,
SpringerBriefs in Mathematics, DOI 10.1007/978-3-319-42545-0_3

happens to the points on the common edge? To which rectangle do they belong? Decorated points resolve this question, and moreover the formalism fits perfectly with our use of decorations to distinguish open and closed ends of intervals.

3.1 The Persistence Measure

Let $\mathbb{V}$ be a persistence module. The **persistence measure** of $\mathbb{V}$ is the function

$$\mu_{\mathbb{V}}(R) = \langle \circ_a\!\!-\!\!\bullet_b\!\!-\!\!\bullet_c\!\!-\!\!\circ_d \mid \mathbb{V} \rangle$$

defined on **rectangles** $R = [a, b] \times [c, d]$ in the plane with $a < b \leq c < d$.

For a decomposable persistence module, there is a clear relationship between $\mu_{\mathbb{V}}$ and the interval summands of $\mathbb{V}$. Let us first consider the case of an interval module.

Proposition 3.2 *Let* $\mathbb{V} = \mathbf{k}^J$ *where* $J = (p^*, q^*)$ *is a real interval. Let* $R = [a, b] \times [c, d]$ *where* $a < b \leq c < d$. *Then*

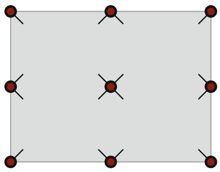
$$\mu_V(R) = \begin{cases} 1 & \textit{if } [b, c] \subseteq J \subseteq (a, d) \\ 0 & \textit{otherwise} \end{cases}$$

Proof It is clear that $\mathbf{k}^J$ restricted to $\{a, b, c, d\}$ is an interval or is zero. Thus, $\mu_{\mathbb{V}}(R) \leq 1$. Moreover $\mu_{\mathbb{V}}(R) = 1$ precisely when

$$\mathbf{k}^J_{\{a,b,c,d\}} = \circ_a\!\!-\!\!\bullet_b\!\!-\!\!\bullet_c\!\!-\!\!\circ_d,$$

which happens if and only if $b, c \in J$ and $a, d \notin J$. This is equivalent to the condition $[b, c] \subseteq J \subseteq (a, d)$. □

Proposition 3.2 has a graphical interpretation. Represent the interval $J \subseteq \mathbf{R}$ as a decorated point in the extended plane. The following picture indicates exactly which decorated points (p^*, q^*) are detected by $\mu_{\mathbb{V}}(R)$:



If (p, q) is in the interior of R then (p^*, q^*) is always detected regardless of the decoration. If (p, q) is on the boundary, then (p^*, q^*) is detected if the tick is directed inwards.

We formalise this by defining a membership relation between decorated points and rectangles.

Definition 3.3 Let $R = [a, b] \times [c, d]$ where $a < b \leq c < d$, and consider a decorated point (p^*, q^*) with $p^* < q^*$. We write $(p^*, q^*) \in R$ if any of the following equivalent statements is true:

- We have $p^* \in [a, b]$ and $q^* \in [c, d]$ in the notation of Eq. (2.3).
- We have $a < p^* < b$ and $c < q^* < d$ in the total order of Eq. (2.1).
- We have $a^+ \leq p^* \leq b^-$ and $c^+ \leq q^* \leq d^-$ in the total order of Eq. (2.1).
- The real interval (p^*, q^*) is sandwiched $[b, c] \subseteq (p^*, q^*) \subseteq (a, d)$.
- The point-with-tick (p^*, q^*) lies in the closed rectangle R.

The set $R^\times = \{(p^*, q^*) \mid (p^*, q^*) \in R\}$ is called the **r-interior** of R. We also make use of its undecorated counterpart, $R^\circ = (a, b) \times (c, d)$, the **interior** in the standard sense of the finite rectangle $R = [a, b] \times [c, d]$.

Remark 3.4 The expressions $(p^*, q^*) \in R$ and $(p^*, q^*) \in R^\times$ mean the same thing but we will tend to prefer the former. In the same spirit, we write $|_R$ to indicate the restriction of a multiset of decorated points to (the r-interior of) the rectangle R.

Corollary 3.5 *Suppose $\mathbb{V}$ is a decomposable persistence module over* $\mathbf{R}$*:*

$$\mathbb{V} = \bigoplus_{\ell \in L} \mathbf{k}(p^*_\ell, q^*_\ell)$$

Then

$$\mu_{\mathbb{V}}(R) = \mathrm{card}\,(\mathsf{Dgm}(\mathbb{V})|_R) \tag{3.1}$$

for every rectangle $R = [a, b] \times [c, d]$ with $a < b \leq c < d$.

Proof This follows from Propositions 3.2 and 2.16 (direct sums). □

We now have a strategy for defining the persistence diagram without assuming a decomposition: having constructed $\mu_{\mathbb{V}}$, we look for a multiset of decorated points $\mathsf{Dgm}(\mathbb{V})$ which satisfies Eq. (3.1) for all rectangles. For this to work, we need to know that such a multiset exists and is unique. Theorem 3.12 will take care of this under the hypothesis that $\mu_{\mathbb{V}}$ is finite and additive; it is a sort of 'Riesz representation theorem' for measures on rectangles. By Corollary 3.5, the new and old definitions agree in the case where $\mathbb{V}$ is decomposable.

3.2 The Persistence Measure (Continued)

We call $\mu_{\mathbb{V}}$ a measure because it is additive with respect to splitting a rectangle into two rectangles. We prove this shortly. First we consider the 'alternating sum' formula for $\mu_{\mathbb{V}}(R)$ that appears in [19]:

Proposition 3.6 *Let $\mathbb{V}$ be a persistence module, and let $a < b \le c < d$. If the vector spaces V_a, V_b, V_c, V_d are finite-dimensional, or less stringently if $\mathrm{r}_c^b < \infty$, then*

$$\langle \circ_a — \bullet_b — \bullet_c — \circ_d \mid \mathbb{V} \rangle = \mathrm{r}_c^b - \mathrm{r}_c^a - \mathrm{r}_d^b + \mathrm{r}_d^a.$$

(Here as before $\mathrm{r}_t^s = \operatorname{rank}(v_t^s : V_s \to V_t)$.)

Proof Decompose the 4-term module $\mathbb{V}_{\{a,b,c,d\}}$ into intervals. The left-hand side counts intervals of type $[b, c]$. By the restriction principle, the four terms on the right-hand side evaluate as follows:

$$\begin{aligned}
\mathrm{r}_c^b &= \langle \circ_a - \bullet_b - \bullet_c - \circ_d \rangle + \langle \bullet_a - \bullet_b - \bullet_c - \circ_d \rangle + \langle \circ_a - \bullet_b - \bullet_c - \bullet_d \rangle + \langle \bullet_a - \bullet_b - \bullet_c - \bullet_d \rangle \\
\mathrm{r}_c^a &= \langle \bullet_a - \bullet_b - \bullet_c - \circ_d \rangle + \langle \bullet_a - \bullet_b - \bullet_c - \bullet_d \rangle \\
\mathrm{r}_d^b &= \langle \circ_a - \bullet_b - \bullet_c - \bullet_d \rangle + \langle \bullet_a - \bullet_b - \bullet_c - \bullet_d \rangle \\
\mathrm{r}_d^a &= \langle \bullet_a - \bullet_b - \bullet_c - \bullet_d \rangle
\end{aligned}$$

These expressions are all finite: the hypothesis $\mathrm{r}_c^b < \infty$ implies that the other three ranks are finite too (Example 2.20). We can legitimately take the alternating sum, and all terms on the right-hand side cancel except for the $\langle \circ_a — \bullet_b — \bullet_c — \circ_d \rangle$. □

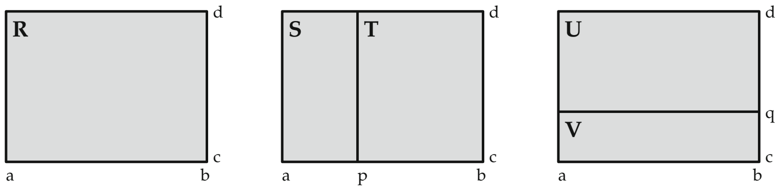
We give three proofs of additivity. The first is completely general, whereas the other two work under restricted settings but are illuminating in their own way.

Proposition 3.7 *$\mu_{\mathbb{V}}$ is additive under vertical and horizontal splitting, meaning that*

$$\begin{aligned}
\mu_{\mathbb{V}}([a, b] \times [c, d]) &= \mu_{\mathbb{V}}([a, p] \times [c, d]) + \mu_{\mathbb{V}}([p, b] \times [c, d]) \\
\mu_{\mathbb{V}}([a, b] \times [c, d]) &= \mu_{\mathbb{V}}([a, b] \times [c, q]) + \mu_{\mathbb{V}}([a, b] \times [q, d])
\end{aligned}$$

whenever $a < p < b \le c < q < d$.

This additivity property is illustrated by the following figure



where the claim is that $\mu_{\mathbb{V}}(R) = \mu_{\mathbb{V}}(S) + \mu_{\mathbb{V}}(T) = \mu_{\mathbb{V}}(U) + \mu_{\mathbb{V}}(V)$.

Proof (*first version*) Let $a < p < b \le c < q < d$. Then we calculate

$$\begin{aligned}
\mu_{\mathbb{V}}([a, b] \times [c, d]) &= \langle \circ_a ——— \bullet_b — \bullet_c — \circ_d \rangle \\
&= \langle \circ_a — \bullet_p — \bullet_b — \bullet_c — \circ_d \rangle + \langle \circ_a — \circ_p — \bullet_b — \bullet_c — \circ_d \rangle \\
&= \langle \circ_a — \bullet_p ——— \bullet_c — \circ_d \rangle + \langle ——— \circ_p — \bullet_b — \bullet_c — \circ_d \rangle \\
&= \mu_{\mathbb{V}}([a, p] \times [c, d]) + \mu_{\mathbb{V}}([p, b] \times [c, d])
\end{aligned}$$

for additivity with respect to a horizontal split, and

$$\begin{aligned}
\mu_{\mathbb{V}}([a,b]\times[c,d]) &= \langle \circ_a\text{---}\bullet_b\text{---}\bullet_c\text{---------}\circ_d \rangle \\
&= \langle \circ_a\text{---}\bullet_b\text{---}\bullet_c\text{---}\circ_q\text{---}\circ_d \rangle + \langle \circ_a\text{---}\bullet_b\text{---}\bullet_c\text{---}\bullet_q\text{---}\circ_d \rangle \\
&= \langle \circ_a\text{---}\bullet_b\text{---}\bullet_c\text{---}\circ_q\text{------} \rangle + \langle \circ_a\text{---}\bullet_b\text{---------}\bullet_q\text{---}\circ_d \rangle \\
&= \mu_{\mathbb{V}}([a,b]\times[c,q]) + \mu_{\mathbb{V}}([a,b]\times[q,d])
\end{aligned}$$

for additivity with respect to a vertical split. □

Proof (*second version, assuming* $\mathrm{r}_c^b < \infty$) The alternating sum formula (Proposition 3.6) gives

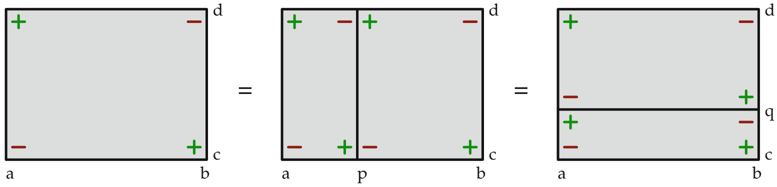
$$\mathrm{r}_c^b - \mathrm{r}_c^a - \mathrm{r}_d^b + \mathrm{r}_d^a = (\mathrm{r}_c^p - \mathrm{r}_c^a - \mathrm{r}_d^p + \mathrm{r}_d^a) + (\mathrm{r}_c^b - \mathrm{r}_c^p - \mathrm{r}_d^b + \mathrm{r}_d^p)$$

and

$$\mathrm{r}_c^b - \mathrm{r}_c^a - \mathrm{r}_d^b + \mathrm{r}_d^a = (\mathrm{r}_c^b - \mathrm{r}_c^a - \mathrm{r}_q^b + \mathrm{r}_q^a) + (\mathrm{r}_q^b - \mathrm{r}_q^a - \mathrm{r}_d^b + \mathrm{r}_d^a)$$

as required. Note that $\mathrm{r}_c^b < \infty$ implies $\mathrm{r}_c^p < \infty$ and $\mathrm{r}_q^b < \infty$, so the formula is valid for all the rectangles in question. □

This second proof is particularly transparent when drawn geometrically in the plane: the + and − signs at the corners of the rectangles cancel in a pleasant way:



Proof (*third version, assuming* $\mathbb{V}$ *is decomposable*) By Corollary 3.5, the measure of a rectangle counts the interval summands whose corresponding decorated points lie in the rectangle. Additivity follows from the observation that a decorated point in R belongs to exactly one of its subrectangles S and T, and to exactly one of its subrectangles U and V. □

Here are two further descriptions of $\mu_{\mathbb{V}}([a,b]\times[c,d])$.

Proposition 3.8 *We have the following formulae:*

$$\begin{aligned}
\langle \circ_a\text{---}\bullet_b\text{---}\bullet_c\text{---}\circ_d \mid \mathbb{V} \rangle &= \dim\left[\frac{\mathrm{im}(v_c^b)\cap\ker(v_d^c)}{\mathrm{im}(v_c^a)\cap\ker(v_d^c)}\right] \\
&= \dim\left[\frac{\ker(v_d^b)}{\ker(v_c^b)+\mathrm{im}(v_b^a)\cap\ker(v_d^b)}\right]
\end{aligned}$$

Proof This is covered, for instance, in the localisation discussion in Sect. 5.1 of [8]. The two formulae are obtained by localising at c and b, respectively. □

Proposition 3.8 expresses the measure of a rectangle as the dimension of a vector space constructed functorially from $\mathbb{V}$. (Ostensibly there are two vector spaces, one for each formula, but the map v_c^b induces a natural isomorphism between them.) Functoriality has its advantages, but in other regards this characterisation is harder to use. For instance, additivity is not as obvious in this formulation.

3.3 Abstract r-Measures

We now consider rectangle measures more abstractly. Persistence measures are of course our primary example, but the general formulation allows for many other situations. For ease of exposition, we initially work in the plane $\mathbf{R}^2$ rather than the extended plane $\overline{\mathbf{R}}^2$. The picture is completed in Sect. 3.6 when we discuss the points at infinity.

Definition 3.9 Let $\mathcal{D}$ be a subset of $\mathbf{R}^2$. Define

$$\text{Rect}(\mathcal{D}) = \{[a, b] \times [c, d] \subset \mathcal{D} \mid a < b \text{ and } c < d\}$$

(the set of closed rectangles contained in $\mathcal{D}$). A **rectangle measure** or **r-measure** on $\mathcal{D}$ is a function

$$\mu : \text{Rect}(\mathcal{D}) \to \{0, 1, 2, \dots\} \cup \{\infty\}$$

which is additive under vertical and horizontal splitting (as in Proposition 3.7).

Proposition 3.10 *Let μ be an r-measure on $\mathcal{D} \subseteq \mathbf{R}^2$. Then:*

- *If $R \in \text{Rect}(\mathcal{D})$ can be written as a union $R = R_1 \cup \dots \cup R_k$ of rectangles with disjoint interiors, then $\mu(R) = \mu(R_1) + \dots + \mu(R_k)$.*
- *If $R \subseteq S$ then $\mu(R) \leq \mu(S)$.*

In other words, μ is finitely additive and monotone.

Proof (*Finitely additive*) Let $R = [a, b] \times [c, d]$. By induction and the vertical splitting property, it follows that finite additivity holds for decompositions of the form

$$R = \bigcup_i R_i$$

where $R_i = [a_i, a_{i+1}] \times [c, d]$ with $a = a_1 < a_2 < \dots < a_m = b$. By induction and the horizontal splitting property, it then follows that finite additivity holds for ‘product’ decompositions

$$R = [a, b] \times [c, d] = \bigcup_{i,j} R_{ij}$$

where $R_{ij} = [a_i, a_{i+1}] \times [c_j, c_{j+1}]$ with $a = a_1 < a_2 < \cdots < a_m = b$ and $c = c_1 < c_2 < \cdots < c_n = d$. For an arbitrary decomposition $R = R_1 \cup \cdots \cup R_k$, finally, the result follows by considering a product decomposition of R by which each R_i is itself product-decomposed.

(Monotone) Decompose S into a collection of rectangles R and $R_1, \ldots, R_{k-1}$ which are interior-disjoint. (This can be done with at most 9 rectangles using a product decomposition.) Then

$$\begin{aligned} \mu(S) &= \mu(R) + \mu(R_1) + \cdots + \mu(R_{k-1}) \\ &\geq \mu(R) \end{aligned}$$

by finite additivity and the fact that $\mu \geq 0$. □

Here is one more plausible-and-also-true statement about abstract r-measures.

Proposition 3.11 (Subadditivity) *Let μ be an r-measure on $\mathcal{D} \subseteq \mathbf{R}^2$. If a rectangle $R \in \mathrm{Rect}(\mathcal{D})$ is contained in a finite union*

$$R \subseteq R_1 \cup \cdots \cup R_k$$

of rectangles $R_i \in \mathrm{Rect}(\mathcal{D})$, then

$$\mu(R) \leq \mu(R_1) + \cdots + \mu(R_k).$$

Proof Let

$$a_1 < a_2 < \cdots < a_m$$

include all the x-coordinates of the corners of all the rectangles, and let

$$c_1 < c_2 < \cdots < c_n$$

include all the y-coordinates. Each rectangle is then tiled as a union of pieces

$$[a_i, a_{i+1}] \times [c_j, c_{j+1}]$$

with disjoint interiors, and the measure of the rectangle is the sum of the measures of its tiles, by additivity. Since each tile belonging to R must also belong to one or more of the R_i, the inequality follows. □

3.4 Equivalence of Measures and Diagrams

We wish to establish a correspondence between r-measures and decorated diagrams. The task of defining a continuous persistence diagram can then be replaced by the simpler task of defining an r-measure. This works best when the measure is finite; in Sect. 3.5 we consider measures that are not finite.

The **r-interior** of a region $\mathcal{D} \subseteq \mathbf{R}^2$ is defined as follows:

$$\mathcal{D}^{\times} = \left\{ (p^*, q^*) \mid \exists R \in \mathrm{Rect}(\mathcal{D}) \text{ such that } (p^*, q^*) \in R \right\}.$$

This is the set of decorated points that can be 'accessed' by some rectangle in $\mathcal{D}$. The decorated diagram will be a multiset in $\mathcal{D}^{\times}$. Clearly, an r-measure in $\mathcal{D}$ cannot tell us what happens outside $\mathcal{D}^{\times}$. The **interior** of $\mathcal{D}$ in the classical sense is written $\mathcal{D}^{\circ}$. In terms of rectangles, we have

$$\mathcal{D}^{\circ} = \left\{ (p, q) \mid \exists R \in \mathrm{Rect}(\mathcal{D}) \text{ such that } (p, q) \in R^{\circ} \right\},$$

where we recall that $R^{\circ} = (a, b) \times (c, d)$ denotes the interior of the closed rectangle $R = [a, b] \times [c, d]$. The undecorated diagram will be a multiset in $\mathcal{D}^{\circ}$.

Theorem 3.12 (The equivalence theorem) *Let $\mathcal{D} \subseteq \mathbf{R}^2$. There is a bijective correspondence between:*

- *Finite r-measures μ on $\mathcal{D}$. 'Finite' means that $\mu(R) < \infty$ for every $R \in \mathrm{Rect}(\mathcal{D})$.*
- *Locally finite multisets A in $\mathcal{D}^{\times}$. 'Locally finite' means that $\mathrm{card}(\mathsf{A}|_R) < \infty$ for every $R \in \mathrm{Rect}(\mathcal{D})$.*

The measure μ corresponding to a multiset A is related to it by the formula

$$\mu(R) = \mathrm{card}(\mathsf{A}|_R) \tag{3.2}$$

for every $R \in \mathrm{Rect}(\mathcal{D})$.

Remark 3.13 We can write Eq. (3.2) equivalently as

$$\mu(R) = \sum_{(p^*, q^*) \in R} \mathrm{m}(p^*, q^*), \tag{3.3}$$

where

$$\mathrm{m} : \mathcal{D}^{\times} \to \{0, 1, 2, \dots\}$$

is the multiplicity function for A.

Assuming the theorem, we define the persistence diagrams of a measure.

Definition 3.14 Let μ be a finite r-measure on a region $\mathcal{D} \subseteq \mathbf{R}^2$.

- The **decorated diagram** of μ is the unique locally finite multiset $\mathsf{Dgm}(\mu)$ in $\mathcal{D}^\times$ such that

$$\mu(R) = \mathrm{card}(\mathsf{Dgm}(\mu)|_R)$$

for every $R \in \mathrm{Rect}(\mathcal{D})$.
- The **undecorated diagram** of μ is the locally finite multiset in $\mathcal{D}^\circ$

$$\mathsf{dgm}(\mu) = \left\{(p, q) \mid (p^*, q^*) \in \mathsf{Dgm}(\mu)\right\} \cap \mathcal{D}^\circ$$

obtained by forgetting the decorations on the points and restricting to the interior.

Remark 3.15 Note that dgm is locally finite in $\mathcal{D}^\circ$, but not necessarily locally finite in $\mathbf{R}^2$—it may have accumulation points on the boundary of $\mathcal{D}$.

Proof (*Theorem* 3.12) One direction of the correspondence is easy. If A is a multiset on $\mathcal{D}^\times$ then the function $\mu(R)$ on rectangles defined by Eq. (3.2) is indeed an r-measure. It is finite if A is locally finite. To verify additivity, suppose that a rectangle R is split vertically or horizontally into two rectangles R_1, R_2. Notice that every decorated point $(p^*, q^*) \in R$ belongs to exactly one of R_1, R_2. It follows that

$$\mu(R) = \mathrm{card}(\mathsf{A}|_R) = \mathrm{card}(\mathsf{A}|_{R_1}) + \mathrm{card}(\mathsf{A}|_{R_2}) = \mu(R_1) + \mu(R_2),$$

as required.

The reverse direction takes more work. Given an r-measure μ we will (1) construct a multiset A in $\mathcal{D}^\times$, (2) show that μ and A are related by Eq. (3.2), and (3) show that A is unique. In practice, we construct the multiplicity function m and establish Eq. (3.3), rather than referring to A directly.

Step 1. (Multiplicity formula.) Let μ be a finite r-measure on $\mathcal{D}$. Define

$$\mathrm{m}(p^*, q^*) = \min\left\{\mu(R) \mid R \in \mathrm{Rect}(\mathcal{D}),\ (p^*, q^*) \in R\right\} \tag{3.4}$$

for (p^*, q^*) in $\mathcal{D}^\times$. Note that the minimum is attained because the set is nonempty and μ takes values in the natural numbers.

Here is an alternative characterisation. Instead of minimising over all rectangles, we take the limit through a decreasing sequence of rectangles:

Lemma 3.16 *Let* (ξ_i) *and* (η_i) *be non-increasing sequences of positive real numbers which tend to zero as* $i \to \infty$. *Then*

$$\mathrm{m}(p^+, q^+) = \lim_{i\to\infty} \mu([p, p + \xi_i] \times [q, q + \eta_i]),$$

and similarly

$$\mathrm{m}(p^+, q^-) = \lim_{i\to\infty} \mu([p, p+\xi_i] \times [q-\eta_i, q]),$$
$$\mathrm{m}(p^-, q^+) = \lim_{i\to\infty} \mu([p-\xi_i, p] \times [q, q+\eta_i]),$$
$$\mathrm{m}(p^-, q^-) = \lim_{i\to\infty} \mu([p-\xi_i, p] \times [q-\eta_i, q]).$$

Proof The key observation is that the sequence of rectangles $R_i = [p, p+\xi_i] \times [q, q+\eta_i]$ is cofinal in the set of rectangles R containing (p^+, q^+). In other words, for any such R we have $R_i \subseteq R$ for all sufficiently large i.

By monotonicity, the sequence of nonnegative integers $\mu(R_i)$ is non-increasing, and hence eventually stabilises to a limit. Then

$$\mathrm{m}(p^+, q^+) \leq \min_i \mu(R_i) = \lim_{i\to\infty} \mu(R_i) \leq \mu(R)$$

for any R containing (p^+, q^+). Taking the minimum over all R, the right-hand side becomes $\mathrm{m}(p^+, q^+)$ and hence by squeezing

$$\mathrm{m}(p^+, q^+) = \lim_{i\to\infty} \mu(R_i).$$

The other three cases of the lemma are similar. □

We return to the main proof.

Step 2. Having defined $\mathrm{m}(p^*, q^*)$, we now show that this is the 'correct' definition, meaning that Eq. (3.3) is satisfied. We have seen already that m corresponds to an r-measure

$$\nu(R) = \sum_{(p^*,q^*)\in R} \mathrm{m}(p^*, q^*), \tag{3.5}$$

and it remains to show (for this step) that $\nu = \mu$. We prove this by induction on $k = \mu(R)$.

Base case. $\mu(R) = 0$. Then for every $(p^*, q^*) \in R$ we have

$$0 \leq \mathrm{m}(p^*, q^*) \leq \mu(R) = 0$$

so $\nu(R) = 0$.

Inductive step. Suppose $\mu(R) = \nu(R)$ for every rectangle R with $\mu(R) < k$. Consider a rectangle R_0 with $\mu(R_0) = k$. We must show that $\nu(R_0) = k$.

Split the rectangle into four equal quadrants S_1, S_2, S_3, S_4. Certainly

$$\mu(R_0) = \mu(S_1) + \mu(S_2) + \mu(S_3) + \mu(S_4)$$
$$\nu(R_0) = \nu(S_1) + \nu(S_2) + \nu(S_3) + \nu(S_4)$$

by finite additivity (Proposition 3.10). If every quadrant satisfies $\mu(S_i) < k$, then by induction we deduce that $\mu(R_0) = \nu(R_0)$. Otherwise, one of the quadrants has

$\mu = k$ and the other three quadrants satisfy $\mu = 0$ (and hence $\nu = 0$). Let R_1 be the distinguished quadrant, so $\mu(R_1) = k$. It is now enough to show that $\nu(R_1) = k$.

We repeat the argument. Subdivide R_i into four equal quadrants. Either all four quadrants satisfy the inductive hypothesis $\mu < k$, in which case we are done. Otherwise we find a quadrant R_{i+1} with $\mu(R_{i+1}) = k$, and we are reduced to showing that $\nu(R_{i+1}) = k$.

In the worst case—the remaining unresolved case—this iteration never terminates and we obtain a sequence of closed rectangles

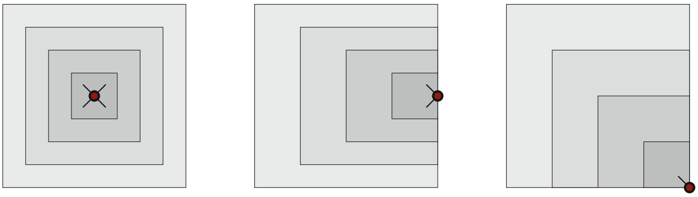
$$R_0 \supset R_1 \supset R_2 \supset \cdots$$

each being a quadrant of the previous one, with $\mu(R_i) = k$. Since the diameters of the rectangles tend to zero, their intersection $\bigcap_i R_i$ contains a single point (r, s).

We are now in a position to show that $\nu(R_0) = k$, by evaluating the sum explicitly over all decorated points in R_0.

First of all, consider decorated points that eventually leave the sequence (R_i). Specifically, suppose that $(p^*, q^*) \in R_0$ but $(p^*, q^*) \in R_{i-1} - R_i$ for some i. This means that (p^*, q^*) belongs to one of the three quadrants of R_{i-1} for which $\mu = 0$. It follows immediately that $\mathrm{m}(p^*, q^*) = 0$.

Thus, the only contribution to $\nu(R_0)$ comes from decorated points (p^*, q^*) which belong to every rectangle in the sequence (R_i). Clearly these must be decorated versions (r^*, s^*) of the intersection point (r, s). There are 4, 2 or 1 of them depending on how the nested sequence of rectangles converges to its limit. Here we illustrate the three cases:



Suppose first that (r, s) lies in the interior of every rectangle R_i, so that all four decorated points (r^+, s^+), (r^+, s^-), (r^-, s^+), (r^-, s^-) belong to every R_i. Divide each R_i into 4 subrectangles R_i^{++}, R_i^{+-}, R_i^{-+}, R_i^{--}, which share a common corner at (r, s) so that each of the four decorated points (r^*, s^*) belongs to one of the subrectangles in the obvious notation. By Lemma 3.16,

$$\mathrm{m}(r^+, s^+) = \lim_{i\to\infty} \mu(R_i^{++}), \quad \mathrm{m}(r^+, s^-) = \lim_{i\to\infty} \mu(R_i^{+-}),$$
$$\mathrm{m}(r^-, s^+) = \lim_{i\to\infty} \mu(R_i^{-+}), \quad \mathrm{m}(r^-, s^-) = \lim_{i\to\infty} \mu(R_i^{--}),$$

and moreover each of these decreasing integer sequences eventually stabilises at its limiting value. Thus, for sufficiently large i,

$$\begin{aligned}\nu(R_0) &= \mathrm{m}(r^+, s^+) + \mathrm{m}(r^+, s^-) + \mathrm{m}(r^-, s^+) + \mathrm{m}(r^-, s^-) \\ &= \mu(R_i^{++}) \quad + \mu(R_i^{+-}) \quad + \mu(R_i^{-+}) \quad + \mu(R_i^{--}) \quad = \mu(R_i) \ = k\end{aligned}$$

as required.

A similar argument (with fewer terms) can be made in the cases where only 2 or 1 of the decorated points (r^*, s^*) belong to every R_i. For instance, if (r, s) lies on the interior of the right-hand edge of the rectangles (R_i) for all sufficiently large i, we split each rectangle into two parts R_i^{-+} and R_i^{--} and obtain

$$\nu(R_0) = \mathrm{m}(r^-, s^+) + \mathrm{m}(r^-, s^-) = \mu(R_i^{-+}) + \mu(R_i^{--}) = \mu(R_i) = k$$

in the same way. In this case (r^+, s^+) and (r^+, s^-) eventually leave (or were never in) the sequence (R_i) and therefore do not contribute to $\nu(R_0)$. We omit the details of the remaining cases, which are equally straightforward.

This completes the inductive step. Thus $\mu(R) = \nu(R)$ for every $R \in \mathrm{Rect}(\mathcal{D})$.

Step 3. Suppose $\mathrm{m}'(p^*, q^*)$ is some other multiplicity function on $\mathcal{D}^\times$ whose associated r-measure

$$\nu'(R) = \sum_{(p^*,q^*)\in R} \mathrm{m}'(p^*, q^*)$$

satisfies $\mu = \nu'$. We must show that $\mathrm{m} = \mathrm{m}'$.

Consider an arbitrary decorated point $(p^*, q^*) \in \mathcal{D}^\times$. Let R be a rectangle which contains (p^*, q^*) at its corner. Since

$$\nu(R) = \nu'(R) = \mu(R) < \infty,$$

there are only finitely many other decorated points $(r^*, s^*) \in R$ with positive multiplicity in m or m′. By making R smaller, we can therefore assume that (p^*, q^*) is the only decorated point in R with positive multiplicity in either measure. Then

$$\mathrm{m}(p^*, q^*) = \nu(R) = \mu(R) = \nu'(R) = \mathrm{m}'(p^*, q^*).$$

Since (p^*, q^*) was arbitrary it follows that $\mathrm{m} = \mathrm{m}'$.

This completes the proof of Theorem 3.12. □

3.5 Non-finite Measures

If an r-measure is not everywhere finite, we restrict our attention to the parts of the plane where it is finite. Define the **finite r-interior** of an r-measure μ to be the set of decorated points

$$\mathcal{F}^\times(\mu) = \left\{(p^*, q^*) \mid \exists R \in \mathrm{Rect}(\mathcal{D}) \text{ such that } (p^*, q^*) \in R \text{ and } \mu(R) < \infty\right\}.$$

The **finite interior** is the set of undecorated points

$$\mathcal{F}^{\circ}(\mu) = \left\{(p, q) \mid \exists R \in \text{Rect}(\mathcal{D}) \text{ such that } (p, q) \in R^{\circ} \text{ and } \mu(R) < \infty\right\}.$$

This is an open subset of the plane, being a union of open rectangles. It is easy to see that $(p, q) \in \mathcal{F}^{\circ}(\mu)$ if and only if $(p^*, q^*) \in \mathcal{F}^{\times}(\mu)$ for all possible decorations. A (decorated) point that is not in the finite (r-)interior may be called **singular**.

Although these interiors are defined in a pointwise sense, the next two propositions show that the finiteness extends to any rectangle contained in them.

Proposition 3.17 *Let* $R \in \text{Rect}(\mathcal{D})$. *If* $R^{\times} \subseteq \mathcal{F}^{\times}(\mu)$ *then* $\mu(R) < \infty$.

Proof We show that each $(p, q) \in R$ is contained in the interior relative to R of a rectangle $S \subseteq R$ of finite measure. If (p, q) lies in the interior of R, then each of the four decorated points (p^*, q^*) belongs to $\mathcal{F}^{\times}(\mu)$ so we can find four finite-measure rectangles containing them. The union of these rectangles contains a neighbourhood of (p, q), and we can take $S \subseteq R$ to be a rectangle contained in this union with (p, q) in its interior. It has finite measure, by subadditivity (Proposition 3.11). The other cases are similar: if (p, q) lies on the interior of an edge, we take two finite-measure rectangles containing a relative neighbourhood of (p, q); and if (p, q) is a corner point we take just one rectangle.

To finish, we note that R, being compact, is the union of finitely many of these rectangles; therefore by subadditivity it has finite measure. □

The undecorated version is an immediate consequence.

Proposition 3.18 *Let* $R \in \text{Rect}(\mathcal{D})$. *If* $R \subseteq \mathcal{F}^{\circ}(\mu)$ *then* $\mu(R) < \infty$. □

Theorem 3.19 *Let* μ *be an r-measure on* $\mathcal{D} \subseteq \mathbf{R}^2$. *Then there is a uniquely defined locally finite multiset* $\mathsf{Dgm}(\mu)$ *in* $\mathcal{F}^{\times}(\mu)$ *such that*

$$\mu(R) = \text{card}(\mathsf{Dgm}(\mu)|_R) \tag{3.6}$$

for every $R \in \text{Rect}(\mathcal{D})$ *with* $R^{\times} \subseteq \mathcal{F}^{\times}(\mu)$.

Proof Within each rectangle S of finite measure, Theorem 3.12 provides a multiset in $S^{\times}$ such that Eq. (3.6) holds for all subrectangles $R \subseteq S$. Uniqueness implies that the multisets for overlapping rectangles agree on the common intersection. Thus we obtain a multiset defined in the union of these $S^{\times}$, which by definition is equal to $\mathcal{F}^{\times}(\mu)$, with the property that Eq. (3.6) holds for all rectangles R of finite measure. By Proposition 3.17, this means all rectangles with $R^{\times} \subseteq \mathcal{F}^{\times}(\mu)$. □

Now we can define the persistence diagrams of a general r-measure.

Definition 3.20 Let μ be an r-measure on a region $\mathcal{D} \subseteq \mathbf{R}^2$.

- The **decorated diagram** of an r-measure μ is the pair $(\mathsf{Dgm}(\mu), \mathcal{F}^{\times}(\mu))$, where $\mathsf{Dgm}(\mu)$ is the multiset in $\mathcal{F}^{\times}(\mu)$ described in Theorem 3.19.

- The **undecorated diagram** is the pair $(\mathsf{dgm}(\mu), \mathcal{F}^\circ(\mu))$, where

$$\mathsf{dgm}(\mu) = \big\{(p, q) \mid (p^*, q^*) \in \mathsf{Dgm}(\mu)\big\} \cap \mathcal{F}^\circ(\mu)$$

  is the locally finite[1] multiset in $\mathcal{F}^\circ(\mu)$ obtained by forgetting the decorations in $\mathsf{Dgm}(\mu)$ and restricting to the finite interior.

Remark 3.21 To make this backwards-compatible with the previously defined persistence diagrams of a finite r-measure on a region $\mathcal{D}$, we can regard $\mathsf{Dgm}(\mu)$ and $\mathsf{dgm}(\mu)$ as abbreviations for $(\mathsf{Dgm}(\mu), \mathcal{D}^\times)$ and $(\mathsf{dgm}(\mu), \mathcal{D}^\circ)$.

Remark 3.22 It is sometimes useful to adopt the **region extension convention**: an r-measure defined on a subset $\mathcal{D} \subset \mathbf{R}^2$ can be interpreted as an r-measure on the whole plane $\mathbf{R}^2$, by agreeing that $\mu(R) = \infty$ for any rectangle that meets $\mathbf{R}^2 - \mathcal{D}$. The extension has the same diagram as the original r-measure.

3.6 Measures and Diagrams in the Extended Plane

We now consider r-measures in the extended plane $\overline{\mathbf{R}}^2$. In Sect. 3.7, we will use this to define the decorated and undecorated diagrams of an arbitrary persistence module. The points at infinity take account of possible infinite interval summands of the module. The development proceeds almost exactly as with the standard plane. What is new is that we use infinite rectangles and we admit $-\infty^+$ and $+\infty^-$ as possible coordinates of decorated points.

A rectangle in the extended plane is a set of the form

$$R = [a, b] \times [c, d]$$

where now $-\infty \leq a < b \leq +\infty$ and $-\infty \leq c < d \leq +\infty$. Thus there are various types of infinite rectangle. The r-interior of a rectangle $R = [a, b] \times [c, d]$ is the set of decorated points

$$R^\times = \{(p^*, q^*) \mid p^* \in [a, b] \text{ and } q^* \in [c, d]\}$$

exactly as before, with the understanding that $-\infty^+$ and $+\infty^-$ are permissible values of p^*, q^*. The interior of R requires a little care:

$$R^\circ = \text{relative interior of } R \text{ as a subspace of } \overline{\mathbf{R}}^2$$

For instance, if $R = [-\infty, b] \times [c, d]$ where b, c, d are finite, then $R^\circ = [-\infty, b) \times (c, d)$. We adopt this convention because we do not wish to lose the points at infinity when we pass from a decorated diagram to its undecorated counterpart.

[1] As before, this does not rule out accumulation points on the boundary of $\mathcal{F}^\circ(\mu)$.

For $\mathcal{D} \subseteq \overline{\mathbf{R}}^2$, let $\mathrm{Rect}(\mathcal{D})$ denote the set of rectangles $R \subseteq \mathcal{D}$. The r-interior and interior of $\mathcal{D}$ are formally the same as before (with R° as above):

$$\mathcal{D}^\times = \left\{(p^*, q^*) \mid \exists R \in \mathrm{Rect}(\mathcal{D}) \text{ such that } (p^*, q^*) \in R\right\},$$
$$\mathcal{D}^\circ = \left\{(p, q) \mid \exists R \in \mathrm{Rect}(\mathcal{D}) \text{ such that } (p, q) \in R^\circ\right\}.$$

An r-measure on $\mathcal{D}$ is a function

$$\mu : \mathrm{Rect}(\mathcal{D}) \to \{0, 1, 2, \dots\} \cup \{\infty\}$$

which is additive with respect to the vertical or horizontal splitting of a rectangle into two rectangles. Propositions 3.10 and 3.11 (finite additivity, monotonicity, subadditivity) follow as before. The finite r-interior and finite interior are

$$\mathcal{F}^\times(\mu) = \left\{(p^*, q^*) \mid \exists R \in \mathrm{Rect}(\mathcal{D}) \text{ such that } (p^*, q^*) \in R \text{ and } \mu(R) < \infty\right\},$$
$$\mathcal{F}^\circ(\mu) = \left\{(p, q) \mid \exists R \in \mathrm{Rect}(\mathcal{D}) \text{ such that } (p, q) \in R^\circ \text{ and } \mu(R) < \infty\right\};$$

if μ is finite, then $\mathcal{F}^\times(\mu) = \mathcal{D}^\times$ and $\mathcal{F}^\circ(\mu) = \mathcal{D}^\circ$.

Claim *Theorems 3.12 and 3.19 are valid, verbatim, for r-measures on $\mathcal{D} \subseteq \overline{\mathbf{R}}^2$. The multiplicity of a point in $\mathsf{Dgm}(\mu)$ may be computed using Eq. (3.4).*

Proof The statements (and indeed the proofs) of Theorems 3.12, 3.19, and Eq. (3.4) are invariant under reparametrisations of the plane of the form

$$x' = f(x), \quad y' = g(y),$$

where f, g are homeomorphic embeddings. We can view $\overline{\mathbf{R}}^2$ as a rectangle in $\mathbf{R}^2$ via a transformation of this type; for instance

$$x' = \arctan(x), \quad y' = \arctan(y)$$

identifies $\overline{\mathbf{R}}^2$ with the rectangle $[-\pi/2, \pi/2] \times [-\pi/2, \pi/2]$ in $\mathbf{R}^2$. Through this hack, the original theorems are transferred to the extended plane. □

Definition 3.23 The **decorated** and **undecorated diagrams** of an r-measure μ are the ordered pairs $(\mathsf{Dgm}(\mu), \mathcal{F}^\times(\mu))$ and $(\mathsf{dgm}(\mu), \mathcal{F}^\circ(\mu))$, where $\mathsf{Dgm}(\mu)$ is given by Theorem 3.19, and $\mathsf{dgm}(\mu)$ is obtained from $\mathsf{Dgm}(\mu)$ by forgetting the decorations and restricting to $\mathcal{F}^\circ(\mu)$.

So far we have treated the extended plane almost exactly like the standard plane. When we come to prove the stability theorem for persistence diagrams (Sects. 5.6–5.8), where metric properties become relevant, we end up considering the points at infinity separately from the points in the standard plane. For this, we make use of certain 'measures at infinity' that we derive now.

Consider an r-measure μ on $\overline{\mathbf{R}}^2$. (If μ is given on a proper sub-region $\mathcal{D} \subset \overline{\mathbf{R}}^2$, extend it to $\overline{\mathbf{R}}^2$ using the region extension convention of Remark 3.22.) The extended plane has 4 lines at infinity

$$(-\infty, \mathbf{R}), \quad (+\infty, \mathbf{R}), \quad (\mathbf{R}, -\infty), \quad (\mathbf{R}, +\infty),$$

and 4 points at infinity

$$(-\infty, -\infty), \quad (+\infty, -\infty), \quad (-\infty, +\infty), \quad (+\infty, +\infty).$$

There are measures defined on each of these. On the four lines at infinity, they are 'interval measures' (the 1-dimensional analogue of r-measures, defined on closed intervals and additive with respect to splitting an interval into two). At the four points at infinity, each measure is simply a number. We write out the three cases of direct relevance to persistence modules. The other five cases are analogous.

- **the line** $(-\infty, \mathbf{R})$:

$$\mu(-\infty, [c, d]) = \lim_{b \to -\infty} \mu([-\infty, b] \times [c, d]) = \min_{b} \mu([-\infty, b] \times [c, d])$$

  for any interval $[c, d] \subseteq \mathbf{R}$.
- **the line** $(\mathbf{R}, +\infty)$:

$$\mu([a, b], +\infty) = \lim_{c \to +\infty} \mu([a, b] \times [c, +\infty]) = \min_{c} \mu([a, b] \times [c, +\infty])$$

  for any interval $[a, b] \subseteq \mathbf{R}$.
- **the point** $(-\infty, +\infty)$:

$$\mu(-\infty, +\infty) = \lim_{e \to +\infty} \mu([-\infty, -e] \times [e, +\infty]) = \min_{e} \mu([-\infty, -e] \times [e, +\infty])$$

Monotonicity of μ guarantees that each limit exists. Each measure has a straightforward interpretation in terms of $\mathsf{Dgm}(\mu)$. We give two sample propositions.

Proposition 3.24 *(1) If $\mu(-\infty, [c, d])$ is finite, then it counts the decorated points of $\mathsf{Dgm}(\mu)$ of the form $(-\infty^+, q^*)$ where $q^* \in [c, d]$.*
(2) If $\mu(-\infty, [c, d])$ is infinite, then $(-\infty^+, q^*) \notin \mathcal{F}^\times(\mu)$ for some $q^* \in [c, d]$.

Proof (1) If the limit is finite, then for some finite b_0 the rectangle $[-\infty, b_0] \times [c, d]$ has finite measure and therefore contains finitely many points of $\mathsf{Dgm}(\mu)$. For all sufficiently small b, then, the rectangle $[-\infty, b] \times [c, d]$ contains exactly those points with first coordinate $-\infty^+$.
(2) Suppose the conclusion failed. Then, arguing as in Proposition 3.17, there would be a finite collection of rectangles $[-\infty, b_i] \times [c_i, d_i]$ of finite measure which together cover $\{-\infty\} \times [c, d]$. Then the rectangle $[-\infty, \min(b_i)] \times [c, d]$ would have finite measure, contradicting the hypothesis. □

Proposition 3.25 *(1) If* $\mu(-\infty, +\infty)$ *is finite, then it counts the multiplicity of* $(-\infty^+, +\infty^-)$ *in* $\mathsf{Dgm}(\mu)$.
(2) If $\mu(-\infty, +\infty)$ *is infinite, then* $(-\infty^+, +\infty^-) \notin \mathcal{F}^\times(\mu)$.

Proof Similar, but easier (especially (2)). □

3.7 The Measure Persistence Diagram

We obtain the measure persistence diagrams of a persistence module $\mathbb{V}$ by defining its persistence measure $\mu_{\mathbb{V}}$ on the extended half-plane $\overline{\mathcal{H}}$.

Definition 3.26 (*persistence measure in extended plane*) Let $\mathbb{V}$ be a persistence module, and let $-\infty \le a < b \le c < d \le +\infty$. We define

$$\mu_{\mathbb{V}}([a, b] \times [c, d]) = \langle \circ_a \!—\! \bullet_b \!—\! \bullet_c \!—\! \circ_d \mid \mathbb{V} \rangle$$

where for infinite rectangles we take $V_{-\infty} = 0$ and $V_{+\infty} = 0$ as needed.

It is easy to see (directly, or by using the arctan trick) that this extended version of $\mu_{\mathbb{V}}$ satisfies the same properties as before: additivity, monotonicity, and so on. In particular, the alternating sum formula of Proposition 3.6 becomes:

$$\begin{array}{llll} \mu_{\mathbb{V}}(\,[-\infty, b] \times [c, +\infty]\,) = \langle \bullet_b \!—\! \bullet_c \mid \mathbb{V} \rangle & = \mathrm{r}_c^b & \\ \mu_{\mathbb{V}}(\,[a, b] \times [c, +\infty]\,) = \langle \circ_a \!—\! \bullet_b \!—\! \bullet_c \mid \mathbb{V} \rangle & = \mathrm{r}_c^b - \mathrm{r}_c^a & (\text{if } \mathrm{r}_c^a < \infty) \\ \mu_{\mathbb{V}}(\,[-\infty, b] \times [c, d]\,) = \langle \bullet_b \!—\! \bullet_c \!—\! \circ_d \mid \mathbb{V} \rangle & = \mathrm{r}_c^b - \mathrm{r}_d^b & (\text{if } \mathrm{r}_d^b < \infty) \end{array}$$

The first of these corresponds to the 'k-triangle lemma' of [19].

In this way $\mu_{\mathbb{V}}$ becomes an r-measure on the extended half-plane

$$\overline{\mathcal{H}} = \{(p, q) \mid -\infty \le p \le q \le +\infty\}$$

with its diagram $\mathsf{Dgm}(\mu_{\mathbb{V}})$ defined on the subset of the r-interior

$$\overline{\mathcal{H}}^\times = \{(p^*, q^*) \mid -\infty^+ \le p^* < q^* \le +\infty^-\}$$

over which $\mu_{\mathbb{V}}$ is finite. Proposition 3.2 and Corollary 3.5 extend straightforwardly to infinite rectangles:

Corollary 3.27 *If* $\mathbb{V}$ *is decomposable into interval modules, then the measure* $\mu_{\mathbb{V}}(R)$ *of any rectangle* R *in the extended half-plane precisely counts the interval summands corresponding to decorated points which lie in* R. □

Definition 3.28 Let $\mathbb{V}$ be a persistence module. Its **measure persistence diagrams** are the decorated diagram

$$\mathsf{Dgm}(\mathbb{V}) = (\mathsf{Dgm}(\mu_{\mathbb{V}}), \mathcal{F}^{\times}(\mu_{\mathbb{V}})),$$

and the undecorated diagram

$$\mathsf{dgm}(\mathbb{V}) = (\mathsf{dgm}(\mu_{\mathbb{V}}), \mathcal{F}^{\circ}(\mu_{\mathbb{V}})).$$

We work in the extended half-plane, so that the finite r-interior and finite interior $\mathcal{F}^{\times}(\mu_{\mathbb{V}})$, $\mathcal{F}^{\circ}(\mu_{\mathbb{V}})$ are subsets of $\overline{\mathcal{H}}^{\times}$, $\overline{\mathcal{H}}^{\circ}$, respectively. When $\mathcal{F}^{\times}(\mu_{\mathbb{V}})$, $\mathcal{F}^{\circ}(\mu_{\mathbb{V}})$ are clear from the context, we may allow ourselves to abuse notation and omit them.

The relationship between the measure and decomposition diagrams is explained in the following proposition:

Proposition 3.29 *If* $\mathbb{V}$ *is decomposable into intervals, then* $\mathsf{Int}(\mathbb{V})$ *agrees with* $\mathsf{Dgm}(\mu_{\mathbb{V}})$ *where the latter is defined, that is, on* $\mathcal{F}^{\times}(\mu_{\mathbb{V}})$.

Proof By Corollary 3.27 we have

$$\mathrm{card}(\mathsf{Int}(\mathbb{V})|_R) = \mu_{\mathbb{V}}(R)$$

for all rectangles. On the other hand, we have

$$\mathrm{card}(\mathsf{Dgm}(\mu_{\mathbb{V}})|_R) = \mu_{\mathbb{V}}(R)$$

for all rectangles with $\mu_{\mathbb{V}}(R) < \infty$. By uniqueness, it follows that $\mathsf{Int}(\mathbb{V})$ and $\mathsf{Dgm}(\mu_{\mathbb{V}})$ must be the same multiset when restricted to $\mathcal{F}^{\times}(\mu_{\mathbb{V}})$. □

Neither definition strictly outperforms the other, as the following examples show.

Example 3.30 (*decomposition* $\geq$ *measure*) Let

$$\mathbb{V} = \bigoplus_{\ell \in L} \mathbf{k}(p_\ell^*, q_\ell^*)$$

where the undecorated pairs (p_ℓ, q_ℓ) form a dense subset of the half-plane $\overline{\mathcal{H}}$. Then $\mathsf{Int}(\mathbb{V})$ is defined; but $\mu_{\mathbb{V}}(R) = \infty$ for every rectangle, so $\mathcal{F}^{\times}(\mu_{\mathbb{V}})$ is the empty set and $\mathsf{Dgm}(\mu_{\mathbb{V}})$ is nowhere defined.

Example 3.31 (*measure* $\geq$ *decomposition*) We adapt the example of Webb [48], from our Theorem 2.8, to be indexed over **R**. Let $\mathbb{W}$ be the persistence module with vector spaces

$$\begin{aligned}
W_t &= 0 && \text{for } t > 0 \\
W_0 &= \{\text{sequences } (x_1, x_2, x_3, \dots) \text{ of scalars}\} \\
W_t &= \{\text{sequences with } x_n = 0 \text{ for all } n \leq |t|\} && \text{for } t < 0
\end{aligned}$$

and maps w_t^s taken to be the canonical inclusion $W_s \subseteq W_t$ when $t \leq 0$ and zero otherwise. This is not decomposable into intervals because its restriction to $-\mathbf{N}$ is not decomposable into intervals (Theorem 2.8(3)). On the other hand, we have

$$\langle \circ_a — \bullet_b \mid \mathbb{W} \rangle = \text{conullity}(W_a \to W_b) < \infty$$

except when $a = -\infty$, and

$$\langle \bullet_c — \circ_d \mid \mathbb{W} \rangle = \text{nullity}(W_c \to W_d) < \infty$$

except when $c \leq 0 < d$. Each of these terms dominates

$$\langle \circ_a — \bullet_b — \bullet_c — \circ_d \mid \mathbb{W} \rangle$$

which is therefore finite for all rectangles that do not contain $(-\infty^+, 0^+)$. Thus, the measure persistence diagram $\mathsf{Dgm}(\mu_{\mathbb{W}})$ is defined everywhere except at that decorated point. We will complete the calculation of $\mathsf{Dgm}(\mu_{\mathbb{W}})$ in Sect. 3.10.

3.8 Tameness

We now describe several different levels of 'tameness' for a persistence module, beginning with the most docile.

- A persistence module is of **finite type** if it is a direct sum of finitely many interval modules. This is the notion of tameness most commonly used in the persistence literature, either explicitly or by implication. Traditionally these modules have simply been called 'tame'.
- A persistence module is **locally finite** if it is a direct sum of interval modules, and satisfies the condition that any bounded subset of $\mathbf{R}$ meets only finitely many of the intervals. By a compactness argument, it is sufficient to require that every $t \in \mathbf{R}$ has a neighbourhood which meets at most finitely many of the intervals.
- A persistence module $\mathbb{V}$ is **pointwise finite-dimensional (pfd)** if each vector space V_t is finite dimensional. As indicated in Theorem 2.8, all such modules have been shown to admit an interval decomposition by Crawley-Boevey [25]. This class of modules is favoured in the work of Lesnick and Bauer [3, 42].

For a module $\mathbb{V}$ that is of finite type, locally finite, or pfd, it is easy to see that $\mathcal{F}^{\times}(\mu_{\mathbb{V}}) = \overline{\mathcal{H}}$. The measure and decomposition diagrams agree and are defined everywhere on $\overline{\mathcal{H}}$, including on the diagonal and at infinity. Any diagonal points are decorated (p^-, p^+), since only these belong to rectangles in $\overline{\mathcal{H}}$.

We introduce four more kinds of tameness. The assumptions here concern the finiteness of $\mu_{\mathbb{V}}$ over different types of rectangle: quadrants, horizontal strips, vertical strips, and bounded rectangles (Fig. 3.1). Each condition guarantees the existence of the persistence diagram over a certain subset of the extended half-plane. The finite

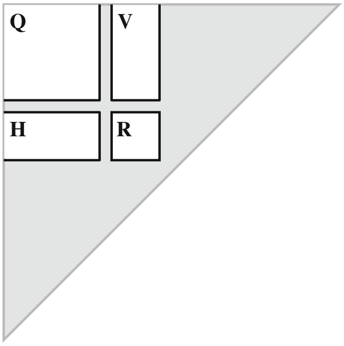


Fig. 3.1 A quadrant, horizontal strip, vertical strip, and finite rectangle in $\overline{\mathcal{H}}$

part of the plane (except the diagonal) is always included; it is at infinity that the four conditions differ.

- We say that $\mathbb{V}$ is **q-tame**, if $\mu_{\mathbb{V}}(Q) < \infty$ for every quadrant Q not touching the diagonal. In other words

$$\langle \bullet_b \!—\! \bullet_c \mid \mathbb{V}\rangle < \infty$$

(that is, $\mathrm{r}_c^b < \infty$) for all $b < c$. The persistence diagram $\mathsf{Dgm}(\mu_{\mathbb{V}})$ is defined over the set:

$$\{(p^*, q^*) \mid -\infty \leq p < q \leq +\infty\} =$$ 

- We say that $\mathbb{V}$ is **h-tame**, if $\mu_{\mathbb{V}}(H) < \infty$ for every horizontally infinite strip H not touching the diagonal. In other words,

$$\langle \bullet_b \!—\! \bullet_c \!—\! \circ_d \mid \mathbb{V}\rangle < \infty$$

for all $b < c < d$. The persistence diagram $\mathsf{Dgm}(\mu_{\mathbb{V}})$ is defined over the set:

$$\{(p^*, q^*) \mid -\infty \leq p < q < +\infty\} =$$ 

- We say that $\mathbb{V}$ is **v-tame**, if $\mu_{\mathbb{V}}(V) < \infty$ for every vertically infinite strip V not touching the diagonal. In other words,

$$\langle \circ_a \!—\! \bullet_b \!—\! \bullet_c \mid \mathbb{V}\rangle < \infty$$

for all $a < b < c$. The persistence diagram $\mathsf{Dgm}(\mu_{\mathbb{V}})$ is defined over the set:

$$\{(p^*, q^*) \mid -\infty < p < q \leq +\infty\} =$$ 

- We say that $\mathbb{V}$ is **r-tame**, if $\mu_{\mathbb{V}}(R) < \infty$ for every finite rectangle R not touching the diagonal. In other words,

$$\langle \circ_a \text{—} \bullet_b \text{—} \bullet_c \text{—} \circ_d \mid \mathbb{V} \rangle < \infty$$

for all $a < b < c < d$. The persistence diagram $\mathsf{Dgm}(\mu_{\mathbb{V}})$ is defined over the set:

$$\{(p^*, q^*) \mid -\infty < p < q < +\infty\} =$$ 

Here is the diagram of implications between the different conditions:

$$\begin{array}{ccccccccc} \text{finite type} & \Rightarrow & \text{locally finite} & \Rightarrow & \text{pfd} & \Rightarrow & \text{q-tame} & \Rightarrow & \text{v-tame} \\ & & & & & & \Downarrow & & \Downarrow \\ & & & & & & \text{h-tame} & \Rightarrow & \text{r-tame} \end{array}$$

One can show that all of the implications are 'strict', in the sense that they cannot be reversed; examples are easily found. The consequent implications

$$\text{q-tame} \;\Rightarrow\; (\text{h-tame } \textbf{and} \text{ v-tame})\,, \qquad (\text{h-tame } \textbf{or} \text{ v-tame}) \;\Rightarrow\; \text{r-tame}$$

are also both strict; examples are suggested by the diagrams in Fig. 3.2. The second of these examples is no surprise, and a better question is this: does every r-tame module decompose as the direct sum of an h-tame module and a v-tame module? This is certainly true if the module is decomposable into intervals, but the general situation seems more subtle and we do not know the answer.

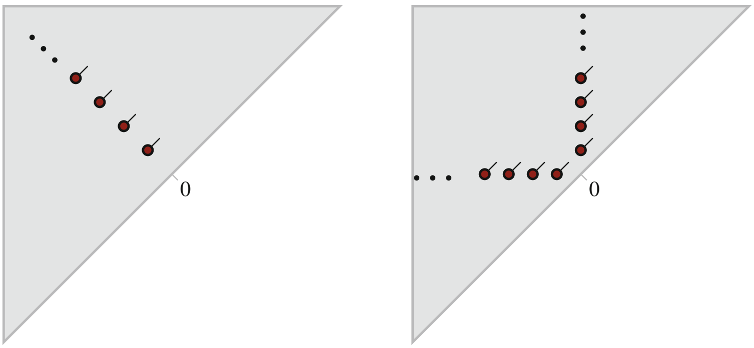


Fig. 3.2 Diagrams of persistence modules which are: (*left*) h-tame and v-tame but not q-tame; (*right*) r-tame but not h-tame or v-tame

Remark 3.32 Later we show that the class of 'q-tame' modules may be interpreted as the closure of the class of 'locally finite' modules: a persistence module is q-tame if and only if it can be approximated arbitrarily well by locally finite modules. See Theorem 5.21.

3.9 Tameness (Continued)

Many naturally occurring persistence modules are q-tame, particularly those arising from persistent homology constructions on compact spaces. We consider some typical examples in this section, using algebraic topology arguments.

A general result of the following kind was published by Cagliari and Landi [6]:

Theorem 3.33 *Let X be a compact polyhedron,*[2] *and let $f : X \to \mathbf{R}$ be a continuous function. Then the persistent homology $\mathrm{H}(\mathbb{X}_{\mathrm{sub}})$ of the sublevelset filtration of (X, f) is q-tame.*

Proof For any $b < c$ we must show that

$$\mathrm{H}(X^b) \longrightarrow \mathrm{H}(X^c)$$

has finite rank. Begin with any triangulation of X, and subdivide it repeatedly until no simplex meets both $f^{-1}(b)$ and $f^{-1}(c)$. If we define Y to be the union of the closed simplices which meet X^b, then we have

$$X^b \subseteq Y \subseteq X^c$$

and hence the factorisation

$$\mathrm{H}(X^b) \longrightarrow \mathrm{H}(Y) \longrightarrow \mathrm{H}(X^c).$$

Since Y is a compact polyhedron, $\mathrm{H}(Y)$ is finite dimensional and so $\mathrm{H}(X^b) \to \mathrm{H}(X^c)$ has finite rank. □

Corollary 3.34 *Let X be a locally compact polyhedron, and let $f : X \to \mathbf{R}$ be a proper*[3] *continuous function which is bounded below. Then $\mathrm{H}(\mathbb{X}_{\mathrm{sub}})$ is q-tame.*

Proof To show that $\mathrm{H}(X^b) \to \mathrm{H}(X^c)$ has finite rank, it is enough to find a compact subpolyhedron of X that contains X^c, because we can then apply Theorem 3.33 in this subpolyhedron. Accordingly, choose a locally finite triangulation of X and consider the closed simplices that meet X^c. There are only finitely many of them because

---

[2] By 'polyhedron' we mean the realisation of a simplicial complex as a topological space. A compact (resp. locally compact) polyhedron is the realisation of a finite (resp. locally finite) complex.

[3] By 'proper' we mean that the preimage $f^{-1}(K)$ of every compact set $K \subset \mathbf{R}$ is compact.

$X^c = f^{-1}[\min(f), c]$ is compact. The union of these simplices is the required subpolyhedron. □

Corollary 3.35 (persistent homology of offsets) *Let A be a nonempty compact subset of $X = \mathbf{R}^n$ and let f be the 'distance from A' in some norm, so $f(x) = \min_{a \in A} \|x - a\|$. It follows from Corollary 3.34 that* $\mathrm{H}(\mathbb{X}_{\text{sub}})$ *is q-tame.* □

Sublevelsets of the 'distance from A' function are generally known as **offsets** of the compact set A in the given norm, and written $A^\varepsilon = f^{-1}(-\infty, \varepsilon]$. There is a rich body of results in the computational geometry literature which govern the topology of these offsets, usually for small ε. These results generally assume that A is 'sufficiently regular', and indeed an important part of that work is to formulate effective regularity conditions that guarantee that offsets are well-behaved. In contrast, Corollary 3.35 tells us that no regularity conditions are needed to guarantee that the persistent homology be q-tame.

Remark 3.36 Under some circumstances, we can obtain stronger tameness results for offsets. If A is a polyconvex set—that is, a finite union of compact convex sets—then $\mathrm{H}(\mathbb{X}_{\text{sub}})$ is of finite type. Indeed, the topology of the offsets can be modelled by a finite filtered simplicial complex, specifically the nerve of the family of offsets of the original convex sets. This works in any norm. On the other hand, the result does not extend to submanifolds. One can manufacture a smooth embedding of the circle in $\mathbf{R}^2$ such that the distance function has infinitely many critical points.

We can drop 'bounded below' in Corollary 3.34 without losing too much:

Theorem 3.37 *Let X be a locally compact polyhedron, and let $f : X \to \mathbf{R}$ be a proper continuous function. Then the persistent homology* $\mathrm{H}(\mathbb{X}_{\text{sub}})$ *of the sublevelset filtration of (X, f) is h-tame and v-tame.*

This means that $\mathrm{H}(\mathbb{X}_{\text{sub}})$ can behave badly only at $(-\infty, +\infty)$. It is easy to construct examples which are definitely not q-tame. The simplest example is $X = \mathbf{Z}$ (the integers), with $f(n) = n$. The 0-homology of any sublevelset is infinite dimensional, and all inclusions have infinite rank.

Proof (h-tameness) Let $b < c < d$. We must show that

$$\langle \bullet_b \!-\!\!-\! \bullet_c \!-\!\!-\! \circ_d \mid \mathrm{H}(\mathbb{X}_{\text{sub}}) \rangle < \infty. \tag{h-*}$$

Begin with a triangulation of X. Only finitely many simplices meet the compact set $f^{-1}(b)$, so again after a finite number of subdivisions no simplex meets both $f^{-1}(b)$ and $f^{-1}(c)$.

Now let Y be the union of the closed simplices which meet X^b, and let Z be the union of the closed simplices which meet X^d. This gives a diagram of inclusions

$$X^b \subseteq Y \subseteq X^c \subseteq X^d \subseteq Z.$$

Note that the polyhedron Z differs from its subpolyhedron Y by the addition of only finitely many simplices, since each such simplex must meet the compact set $f^{-1}[b, d]$. Thus the relative homology $\mathrm{H}(Z, Y)$ is finite-dimensional.

We now work with the induced homology diagram

$$\mathrm{H}(X^b) \longrightarrow \mathrm{H}(Y) \longrightarrow \mathrm{H}(X^c) \longrightarrow \mathrm{H}(X^d) \longrightarrow \mathrm{H}(Z).$$

In the obvious notation,

$$\begin{aligned}\langle \bullet_b\text{———}\bullet_c\text{—}\circ_d\text{——}\rangle &= \langle \bullet_b\text{—}\bullet_y\text{—}\bullet_c\text{—}\circ_d\text{—}\circ_z\rangle\\ &\le \langle\text{——}\bullet_y\text{————}\circ_z\rangle\\ &= \dim[\ker(\mathrm{H}(Y) \to \mathrm{H}(Z))].\end{aligned}$$

By the homology long exact sequence for the pair (Z, Y), we have

$$\ker(\mathrm{H}(Y) \longrightarrow \mathrm{H}(Z)) = \mathrm{im}(\mathrm{H}(Z, Y) \longrightarrow \mathrm{H}(Y))$$

which is finite-dimensional. This confirms (h-*).

v-tameness. Let $a < b < c$. We must show that

$$\langle \circ_a\text{—}\bullet_b\text{—}\bullet_c \mid \mathrm{H}(\mathbb{X}_{\mathrm{sub}})\rangle < \infty \tag{v-*}$$

Using a similar argument to the above, we construct a diagram of inclusions

$$Y \subseteq X^a \subseteq X^b \subseteq Z \subseteq X^c$$

where Y, Z are polyhedra with $\mathrm{H}(Z, Y)$ finite-dimensional. Working with the homology diagram

$$\mathrm{H}(Y) \longrightarrow \mathrm{H}(X^a) \longrightarrow \mathrm{H}(X^b) \longrightarrow \mathrm{H}(Z) \longrightarrow \mathrm{H}(X^c),$$

we estimate

$$\begin{aligned}\langle\text{——}\circ_a\text{—}\bullet_b\text{———}\bullet_c\rangle &= \langle \circ_y\text{—}\circ_a\text{—}\bullet_b\text{—}\bullet_z\text{—}\bullet_c\rangle\\ &\le \langle \circ_y\text{————}\bullet_z\text{——}\rangle\\ &= \dim[\mathrm{coker}(\mathrm{H}(Y) \to \mathrm{H}(Z))]\end{aligned}$$

By the homology long exact sequence of the pair (Z, Y), we have

$$\mathrm{coker}(\mathrm{H}(Y) \to \mathrm{H}(Z)) \cong \mathrm{im}(\mathrm{H}(Z) \to \mathrm{H}(Z, Y))$$

which is finite-dimensional. This confirms (v-*). □

There are many other examples of naturally occurring q-tame modules. It is shown in [16] that the Vietoris–Rips and Čech complexes of a compact metric space have q-tame persistent homology. This is a situation where persistence is manifestly necessary, because such complexes can behave very badly at individual parameter values. For instance, J.-M. Droz [28] has constructed a compact metric space whose Vietoris–Rips complex has homology of uncountable dimension at uncountably many parameter values (indeed, over an entire interval). The construction is not at all pathological in appearance; see [16] for additional examples.

3.10 Vanishing Lemmas

Here are some easy lemmas that guarantee the vanishing of the persistence diagram in certain parts of the plane. These lemmas simplify the task of computing Dgm, often reducing it to a few specific quiver calculations.

Lemma 3.38 *Let $\mathbb{V}$ be a persistence module, and let $s < t$.*

(1) The map v_t^s is surjective iff $\mathsf{Dgm}(\mathbb{V})$ is empty in the rectangle $[s, t] \times [t, +\infty]$.
(2) The map v_t^s is injective iff $\mathsf{Dgm}(\mathbb{V})$ is empty in the rectangle $[-\infty, s] \times [s, t]$.

'Empty' means that the r-interior of the rectangle contains no points or singular points of $\mathsf{Dgm}(\mathbb{V})$. See Fig. 3.3.

Metaphorically, the surjectivity of v_t^s tells us that every feature that survives to time t already existed at time s. Injectivity tells us that every feature that survives to time s remains alive time t. The small triangle between the two rectangles is the 'wiggle-room': new features may appear in the time interval (s, t) but must also disappear in the same time interval, and vice versa. The lemma makes these claims precise in terms of the persistence diagram.

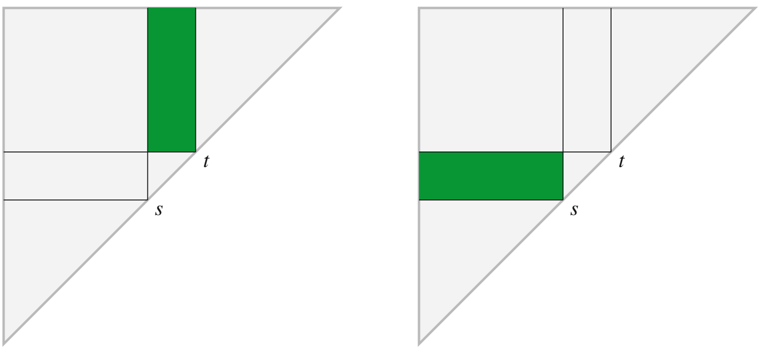


Fig. 3.3 The *shaded green rectangle* indicates the region where the persistence diagram is guaranteed to be empty: (*left*) when v_t^s is surjective; (*right*) when v_t^s is injective

Proof Indeed,

$$\mu_{\mathbb{V}}([s,t] \times [t,+\infty]) = \langle \circ_s\!\!-\!\!\bullet_t \mid \mathbb{V}\rangle = \operatorname{conullity}(v_t^s),$$

the right-hand side of which is zero iff v_t^s is surjective; and

$$\mu_{\mathbb{V}}([-\infty,s] \times [s,t]) = \langle \bullet_s\!\!-\!\!\circ_t \mid \mathbb{V}\rangle = \operatorname{nullity}(v_t^s),$$

the right-hand side of which is zero iff v_t^s is injective. □

In many situations the maps v_t^s are surjective or injective everywhere in some interval. The following lemma is stated carefully to give a sharp statement for all possible interval types (open, closed, half-open, infinite).

Lemma 3.39 *Let $\mathbb{V}$ be a persistence module and let $J = (p^*, q^*)$ be an interval.*

(1) The maps $(v_t^s \mid s,t \in J \text{ with } s < t)$ are all surjective iff $\mathsf{Dgm}(\mathbb{V})$ *is empty in the vertical band*

$$\{(x^*, y^*) \in \overline{\mathcal{H}} \mid x^* \in J\}.$$

(2) The maps $(v_t^s \mid s,t \in J \text{ with } s < t)$ are all injective iff $\mathsf{Dgm}(\mathbb{V})$ *is empty in the horizontal band*

$$\{(x^*, y^*) \in \overline{\mathcal{H}} \mid y^* \in J\}.$$

(We recall from Eq. (2.3) that $x^* \in J$ means $p^* < x^* < q^*$.) See Fig. 3.4.

Proof This follows from Lemma 3.38, since the vertical (resp. horizontal) band is the union, over $s,t \in J$ with $s < t$, of the r-interiors of the vertical (resp. horizontal) rectangles of the lemma. □

We now calculate the persistence diagram for the example of Webb given earlier.

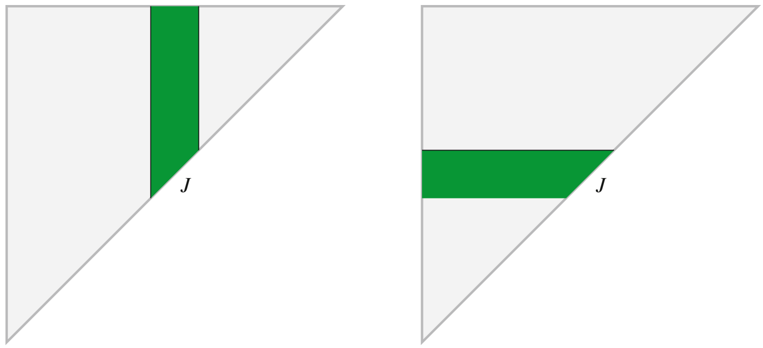


Fig. 3.4 The *shaded green band* indicates the region where the persistence diagram is guaranteed to be empty when the maps (v_t^s) over the interval J are: (*left*) surjective; (*right*) injective. Along each of the parallel boundary edges, vanishing of the diagram for points with inward ticks is guaranteed if J contains its corresponding endpoint

Example 3.40 (*Continuation of Example* 3.31) Recall that $\mathbb{W}$ was defined by setting

$$\begin{aligned} W_t &= 0 && \text{for } t > 0 \\ W_0 &= \{\text{sequences } (x_1, x_2, x_3, \dots) \text{ of real numbers}\} \\ W_t &= \{\text{sequences with } x_n = 0 \text{ for all } n \leq |t|\} && \text{for } t < 0 \end{aligned}$$

and taking (w_t^s) to be the canonical inclusion maps or zero maps. We also take $W_{-\infty} = W_{+\infty} = 0$ to allow uniform treatment of finite and infinite rectangles. Then the maps (w_t^s) are surjective over the intervals

$$(-1, +\infty] \quad \text{and} \quad (-2, -1], \quad (-3, -2], \quad \dots, \quad (-n-1, -n], \quad \dots$$

and injective over the intervals

$$[-\infty, 0] \quad \text{and} \quad (0, +\infty]$$

so Lemma 3.39 allows only the possibilities

$$x^* \in \{-1^+, -2^+, -3^+, \dots, -\infty^+\},$$

for (x^*, y^*) that are points or singular points of $\mathsf{Dgm}(\mathbb{W})$. We can determine the multiplicity of $(-n^+, 0^+)$ by enclosing it in a rectangle that contains no other candidate points: say $[-n, -n+1] \times [0, +\infty]$. Let $\hat{\mathbb{W}}$ be the restriction

$$W_{-n} \longrightarrow W_{-n+1} \longrightarrow W_0$$

of $\mathbb{W}$ to the index set $\{-n, -n+1, 0\}$. Then

$$\begin{aligned} \mathrm{m}_{\mathbb{W}}((-n^+, 0^+)) &= \mu_{\mathbb{W}}([-n, -n+1] \times [0, +\infty]) \\ &= \langle \circ \!-\!\!-\! \bullet \!-\!\!-\! \bullet \ | \hat{\mathbb{W}} \rangle \\ &= \langle \circ \!-\!\!-\! \bullet \!-\!\!-\!\!-\!\!-\! \ | \hat{\mathbb{W}} \rangle - \langle \circ \!-\!\!-\! \bullet \!-\!\!-\! \circ \ | \hat{\mathbb{W}} \rangle \\ &= \langle \circ \!-\!\!-\! \bullet \!-\!\!-\!\!-\!\!-\! \ | \hat{\mathbb{W}} \rangle \\ &= \mathrm{conullity}(W_{-n} \to W_{-n+1}) = 1. \end{aligned}$$

The fourth equality follows from the injectivity of $W_{-n+1} \to W_0$, which implies that $\langle \circ \!-\!\!-\! \bullet \!-\!\!-\! \circ \ | \hat{\mathbb{W}} \rangle$ vanishes.

The conclusion is that each $(-n^+, 0^+)$ occurs exactly once in the persistence diagram. Finally, we have already seen that $(-\infty^+, 0^+)$ is a singular point of $\mu_{\mathbb{W}}$.[4] All decorated points in $\overline{\mathcal{H}}$ have been accounted for. See Fig. 3.5.

---

[4] In retrospect it's even clearer: any rectangle containing $(-\infty^+, 0^+)$ must also contain infinitely many of the points $(-n^+, 0^+)$ that we have calculated to have multiplicity 1, and therefore must have infinite measure.

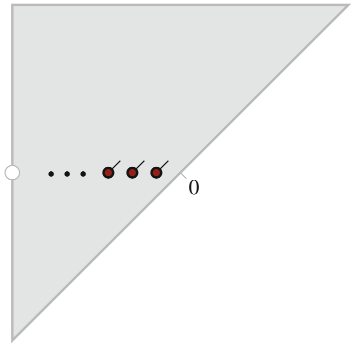


Fig. 3.5 The persistence diagram $\mathsf{Dgm}(\mathbb{W}) = \mathsf{Dgm}(\mu_{\mathbb{W}})$ for the example of Webb. This is defined everywhere in the extended half-plane except at the singular point $(-\infty^+, 0^+)$

3.11 Vanishing Lemmas (Continued)

In this optional section, we provide a set of four infinitesimal vanishing lemmas to accompany the results of the previous section. These lemmas are expressed in terms of direct limits and inverse limits. The reader unfamiliar with these concepts from category theory may wish to consult a standard textbook, such as [43], for additional details and context.

We begin with the observation that any persistence module over the real line has a canonical extension to the totally ordered set

$$\mathbf{R} \cup \mathbf{R}^* = \{t, t^-, t^+ \mid t \in \mathbf{R}\}$$

defined using direct limits and inverse limits in the category of vector spaces. Here are the constructions.

Definition 3.41 (*extension to t^-*) Let $\mathbb{V}$ be a persistence module and let $t \in \mathbf{R}$. We define

$$V_{t^-} = \varinjlim\, (V_a \mid a < t).$$

This direct limit can be defined explicitly as the quotient of the direct sum vector space

$$\bigoplus_{a<t} V_a \tag{3.7}$$

by the subspace generated by all vectors of the form $x_a \oplus -v_b^a(x_a) \in V_a \oplus V_b$ where $a < b < t$.

Definition 3.42 (*extension to t^+*) Let $\mathbb{V}$ be a persistence module and let $t \in \mathbf{R}$. We define

$$V_{t^+} = \varprojlim\, (V_a \mid a > t).$$

This inverse limit can be defined explicitly as the subspace of the product vector space

$$\prod_{a>t} V_a \tag{3.8}$$

that comprises those product vectors $(x_a \mid a > t)$ satisfying the constraint $x_b = v_b^a(x_a)$ whenever $t < a < b$.

Note that for $a < t < b$ there are canonical maps

$$V_a \xrightarrow{v_{t^-}^a} V_{t^-} \quad \text{and} \quad V_{t^+} \xrightarrow{v_b^{t^+}} V_b$$

induced, respectively, from the inclusion of V_a into the direct sum of (3.7) and from the projection of the product of (3.8) onto V_b. We observe that

$$V_{t^-} = \bigcup_{a<t} \mathrm{im}(v_{t^-}^a) \quad \text{and} \quad 0 = \bigcap_{b>t} \ker(v_b^{t^+}) \tag{3.9}$$

where the union is a nested union and the intersection is a nested intersection. In the other direction, there are canonical maps

$$V_{t^-} \xrightarrow{v_t^{t^-}} V_t \xrightarrow{v_{t^+}^t} V_{t^+} \tag{3.10}$$

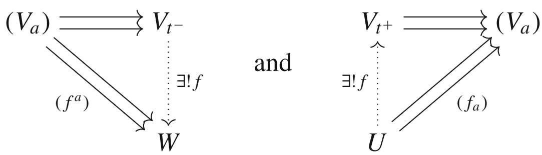
resulting from the universal properties of direct and inverse limits stated below. It is not difficult to check that these maps, and their various composites with each other and with the maps v_t^s, define an extension of the persistence module $\mathbb{V}$ to the index set $\mathbf{R} \cup \mathbf{R}^*$. The details are left to the reader.

Here are the universal properties that characterise these direct and inverse limits:

Proposition 3.43 (universal property of V_{t^-}) *Given a vector space W and a family of linear maps $(f^a : V_a \to W \mid a < t)$ such that $f^a = f^b v_b^a$ whenever $a < b < t$, there is a unique linear map $f : V_{t^-} \to W$ such that $f_a = f v_{t^-}^a$ whenever $a < t$.* □

Proposition 3.44 (universal property of V_{t^+}) *Given a vector space U and a family of linear maps $(f_a : U \to V_a \mid a > t)$ such that $f_b = v_b^a f_a$ whenever $t < a < b$, there is a unique linear map $f : U \to V_{t^+}$ such that $f_a = v_a^{t^+} f$ whenever $a > t$.* □

These universal properties are easily verified from the explicit definitions above. The following diagrams represent these properties schematically:



Double arrows indicate a system of maps that commute with the (v_b^a).

Here is the main result of this section, stated in terms of the maps in (3.10).

Theorem 3.45 (infinitesimal vanishing lemmas) *Let $\mathbb{V}$ be a q-tame persistence module, and let $t \in \mathbf{R}$. In the following, restrict to $p < t < q$.*

(1) If $v_t^{t^-}$ is surjective then $\mathsf{Dgm}(\mathbb{V})$ contains no points of the form (t^-, q^*).
(2) If $v_t^{t^-}$ is injective then $\mathsf{Dgm}(\mathbb{V})$ contains no points of the form (p^*, t^-).
(3) If $v_{t^+}^t$ is surjective then $\mathsf{Dgm}(\mathbb{V})$ contains no points of the form (t^+, q^*).
(4) If $v_{t^+}^t$ is injective then $\mathsf{Dgm}(\mathbb{V})$ contains no points of the form (p^*, t^+).

The q-tameness assumption is needed for elementary numerical reasons in all four cases, and also for deeper structural reasons in the case of (3).

Proof (1) Suppose $t < c < q$. We will find $a < t$ such that the rectangle $[a, t] \times [c, +\infty]$ has measure zero. This will imply that $\mathsf{Dgm}(\mathbb{V})$ does not contain (t^-, q^*).

Using the sequence

$$V_a \longrightarrow V_{t^-} \longrightarrow V_t \longrightarrow V_c$$

and the surjectivity hypothesis, $\langle \circ_{t^-} \!—\! \bullet_t \rangle = 0$, we calculate

$$\begin{aligned}
\mu_{\mathbb{V}}([a, t] \times [c, +\infty]) &= \langle \circ_a \text{———} \bullet_t — \bullet_c \rangle \\
&= \langle \circ_a — \bullet_{t^-} — \bullet_t — \bullet_c \rangle + \langle \circ_a — \circ_{t^-} — \bullet_t — \bullet_c \rangle \\
&= \langle \circ_a — \bullet_{t^-} — \bullet_t — \bullet_c \rangle \\
&= \langle \circ_a — \bullet_{t^-} \text{———} \bullet_c \rangle \\
&= \dim \left[\frac{\operatorname{im}(V_{t^-} \longrightarrow V_c)}{\operatorname{im}(V_a \longrightarrow V_c)} \right].
\end{aligned}$$

It follows from the first assertion of (3.9) that

$$\operatorname{im}(V_{t^-} \longrightarrow V_c) = \bigcup_{a<t} \operatorname{im}(V_a \longrightarrow V_c) \tag{3.11}$$

where the right-hand side is a nested union. Since the left-hand side is finite-dimensional, $\mathbb{V}$ being q-tame, it follows that there exists $a < t$ such that

$$\operatorname{im}(V_{t^-} \longrightarrow V_c) = \operatorname{im}(V_a \longrightarrow V_c)$$

and for this a we have $\mu_{\mathbb{V}}([a, t] \times [c, +\infty]) = 0$ as required.

(2) Suppose $p < b < t$. We will find c with $b < c < t$ such that the rectangle $[-\infty, b] \times [c, t]$ has measure zero. This will imply that $\mathsf{Dgm}(\mathbb{V})$ does not contain $(p^*, t-)$. We use the sequence

$$V_b \longrightarrow V_c \longrightarrow V_{t^-} \longrightarrow V_t$$

and the injectivity hypothesis, $\langle \bullet_{t^-} \text{—} \circ_t \rangle = 0$, to calculate

$$\mu_{\mathbb{V}}([-\infty, b] \times [c, t]) = \langle \bullet_b \text{—} \bullet_c \text{———} \circ_t \rangle$$
$$= \langle \bullet_b \text{—} \bullet_c \text{—} \circ_{t^-} \text{——} \rangle = \dim \left[\frac{\ker(V_b \longrightarrow V_{t^-})}{\ker(V_b \longrightarrow V_c)} \right]$$

analogously to the computation in (1). Note that this quantity is finite, since it is bounded by $\langle \bullet_b \text{—} \bullet_c \rangle = \mathrm{r}_c^b$. This time we use

$$\ker(V_b \longrightarrow V_{t^-}) = \bigcup_{c<t} \ker(V_b \longrightarrow V_c) \tag{3.12}$$

which follows from the fact that any element of V_c that maps to zero in V_{t^-} must map to zero in some $V_{c'}$ with $c < c' < t$. Now the right-hand side is a nested increasing union of subspaces of finite codimension in the left-hand side. It follows that there exists $c < t$ such that

$$\ker(V_b \longrightarrow V_{t^-}) = \ker(V_b \longrightarrow V_c)$$

therefore for this c we have $\mu_{\mathbb{V}}([-\infty, b] \times [c, t])$ as required.

(3) Suppose $t < c < q$. We will find b with $t < b < c$ such that the rectangle $[t, b] \times [c, +\infty]$ has measure zero. This will imply that $\mathsf{Dgm}(\mathbb{V})$ does not contain (t^+, q^*). We use the sequence

$$V_t \longrightarrow V_{t^+} \longrightarrow V_b \longrightarrow V_c$$

and the surjectivity hypothesis $\langle \circ_t \text{—} \bullet_{t^+} \rangle$ to calculate

$$\mu_{\mathbb{V}}([t, b] \times [c, +\infty]) = \langle \circ_t \text{———} \bullet_b \text{—} \bullet_c \rangle$$
$$= \langle \text{——} \circ_{t^+} \text{—} \bullet_b \text{—} \bullet_c \rangle = \dim \left[\frac{\mathrm{im}(V_b \longrightarrow V_c)}{\mathrm{im}(V_{t^+} \longrightarrow V_c)} \right]$$

in the usual way. We now proceed to deduce

$$\mathrm{im}(V_{t^+} \longrightarrow V_c) = \bigcap_{b>t} \mathrm{im}(V_b \longrightarrow V_c) \tag{3.13}$$

from the q-tameness of $\mathbb{V}$. Certainly the left-hand side is contained in the right-hand side. Conversely, suppose that $x_c \in V_c$ lies in the image of every V_b. We have to find a consistent family $(x_b \mid t < b \leq c)$ of vectors $x_b \in V_b$ which map to x_c. It suffices to consider an arbitrary decreasing sequence (b_i) that converges to t, and define a consistent family (x_{b_i}). We begin with $b_0 = c$, and then recursively select $x_{b_i} \in V_{b_i}$ so that x_{b_i} maps to $x_{b_{i-1}}$ and is contained in the intersection of the subspaces $\mathrm{im}(v_{b_i}^b)$ where $t < b < b_i$. The crucial property is that each such nested decreasing family of images is eventually constant; this is the *Mittag-Leffler condition*, which is satisfied by q-tame modules because each image has finite dimension. Using this property, we set $x_{b_i} = v_{b_i}^{b}(x_b)$ for some x_b mapping to $x_{b_{i-1}}$ where b is sufficiently small that the nested images in V_{b_i} have stabilised.

With (3.13) established, the eventual constancy of the images on the right-hand side implies that there exists $b > t$ such that

$$\mathrm{im}(V_{t^+} \longrightarrow V_c) = \mathrm{im}(V_b \longrightarrow V_c)$$

and for this b we have $\mu_{\mathbb{V}}([t, b] \times [c, +\infty]) = 0$ as required.

(4) Suppose $p < b < t$. We will find $d > t$ such that the rectangle $[-\infty, b] \times [t, d]$ has measure zero. This will imply that $\mathsf{Dgm}(\mathbb{V})$ does not contain the point (p^*, t^+). We use the sequence

$$V_b \longrightarrow V_t \longrightarrow V_{t^+} \longrightarrow V_d$$

and the injectivity hypothesis, $\langle \bullet_t \!-\!\!-\! \circ_{t^+} \rangle = 0$, to calculate

$$\begin{aligned} \mu_{\mathbb{V}}([-\infty, b] \times [t, d]) &= \langle \bullet_b \!-\!\!-\! \bullet_t \!-\!\!-\!\!-\!\!-\!\!-\! \circ_d \rangle \\ &= \langle \bullet_b \!-\!\!-\!\!-\!\!-\!\!-\! \bullet_{t^+} \!-\!\!-\! \circ_d \rangle = \dim \left[\frac{\ker(V_b \longrightarrow V_d)}{\ker(V_b \longrightarrow V_{t^+})} \right] \end{aligned}$$

as usual. This quantity is finite, being bounded by $\langle \bullet_b \!-\!\!-\! \bullet_t \rangle = \mathrm{r}_t^b$. Now we use

$$\ker(V_b \longrightarrow V_{t^+}) = \bigcap_{d>t} \ker(V_b \longrightarrow V_d) \tag{3.14}$$

which follows from the second assertion of (3.9). Since the codimension of the left-hand side is finite in any of the kernels on the right-hand side, this nested decreased family of kernels must eventually be constant. Thus there exists $d > t$ such that

$$\ker(V_b \longrightarrow V_{t^+}) = \ker(V_b \longrightarrow V_d)$$

and for this d we have $\mu_{\mathbb{V}}([-\infty, b] \times [t, d]) = 0$ as required. □

Remark 3.46 In the language of abelian categories, identities (3.11)–(3.14) result from the exactness properties of direct and inverse limits of vector spaces. Indeed, (3.11) and (3.12) follow from the right- and left-exactness of the direct-limit functor, and (3.14) follows from the left-exactness of the inverse-limit functor. The difficulty with (3.13) is that the inverse-limit functor is not right-exact. When $\mathbb{V}$ satisfies the Mittag-Leffler condition, however, the derived functor $\varprojlim^1$ evaluates to zero on $\mathbb{V}|_{(t,+\infty)}$ and restores right-exactness precisely where we need it.

We finish this section with two corollaries. The first of these originally appeared in Cerri et al. [12] as the assertion that certain ‘size functions’ are right-continuous.

Corollary 3.47 ([12] Proposition 2.9) *Let* $\mathbb{X}_{\text{sub}}$ *be the sublevelset filtration of a pair* (X, f) *where* X *is a compact polyhedron, and let* $\check{\mathrm{H}} = \check{\mathrm{H}}_k(-; \mathbf{k})$ *be a Čech homology functor.*[5] *Then the only points of* $\mathsf{Dgm}(\check{\mathrm{H}}(\mathbb{X}_{\text{sub}}))$ *away from the diagonal are of the form* (p^-, q^-).

Proof The q-tameness of $\check{\mathrm{H}}(\mathbb{X}_{\text{sub}})$ follows the proof of Theorem 3.33, since Čech homology agrees with simplicial homology on the intermediate polyhedron Y.

Chapter X, Theorem 3.1 of [32] implies that the natural map

$$v^t_{t+} : \check{\mathrm{H}}(X^t) \longrightarrow \varprojlim \left(\check{\mathrm{H}}(X^a) \mid a > t\right)$$

is an isomorphism for every t. Parts (3) and (4) of Theorem 3.45 now constrain the off-diagonal points in the persistence diagram to negative decorations only. □

Remark 3.48 In the compact ‘Morselike’ situations of Example 3.51, a stronger conclusion holds for simpler reasons and in any homology theory.

Corollary 3.49 *Let* $\mathbb{X}^{\circ}_{\text{sub}}$ *be the open sublevelset filtration*[6] *of a pair* (X, f) *where* X *is a compact polyhedron, and let* $\mathrm{H} = \mathrm{H}_k(-, \mathbf{k})$ *be a singular homology functor. Then the only points of* $\mathsf{Dgm}(\check{\mathrm{H}}(\mathbb{X}_{\text{sub}}))$ *away from the diagonal are of the form* (p^+, q^+).

Proof The q-tameness of $\mathrm{H}(\mathbb{X}^{\circ}_{\text{sub}})$ follows the proof of Theorem 3.33, with the compact polyhedron Y equally well serving as an intermediate space between open sublevelsets.

Since singular simplices are compactly supported, the natural map

$$v^{t^-}_t : \varinjlim \left(\mathrm{H}(X^a_{\circ}) \mid a < t\right) \longrightarrow \mathrm{H}(X^t_{\circ})$$

is an isomorphism for every t. Parts (1) and (2) of Theorem 3.45 now constrain the off-diagonal points in the persistence diagram to positive decorations only. □

3.12 Finite Approximations

We finish this chapter by relating our measure-theoretic persistence diagrams to the diagrams constructed more traditionally, perhaps by computer, in situations of finite information.

---

[5] The theory of Čech homology is described in detail by Eilenberg and Steenrod [32].

[6] The spaces are the open sublevelsets $X^t_{\circ} = f^{-1}(-\infty, t)$ and the maps are their inclusions.

We begin by noting that, away from its finite r-interior, a persistence measure gives only a limited view of the structure of its persistence module. For example:

- It is not possible to distinguish between the many non-isomorphic persistence modules $\mathbb{V}$ for which $\mu_{\mathbb{V}}$ is infinite on every rectangle.
- If the persistence diagram of $\mathbb{V}$ contains a sequence of points (p_n^*, q_n^*) with p_n converging to r from below and q_n converging to r from above, then there is no way to determine the multiplicity of (r^-, r^+) from the measure alone.

On the other hand, from $\mu_{\mathbb{V}}$ we do recover all information obtainable by restricting $\mathbb{V}$ to finite subsets $\mathbf{T} \subset \mathbf{R}$. We may call this the 'finitely observable' part of $\mathbb{V}$. Specifically, for any finite index set

$$\mathbf{T}: \quad a_1 < a_2 < \cdots < a_n$$

we can determine the interval decomposition of $\mathbb{V}_{\mathbf{T}}$ (the restriction of $\mathbb{V}$ to the index set $\mathbf{T}$) in terms of the measure $\mu_{\mathbb{V}}$. There are four plausible naming conventions for intervals in $\mathbf{T}$:

$$\{a_i, a_{i+1}, \ldots, a_j\} = [a_i, a_j] = [a_i, a_{j+1}) = (a_{i-1}, a_j] = (a_{i-1}, a_{j+1})$$

Here a_0, a_{n+1} are to be interpreted as $-\infty, +\infty$ respectively. It is conventional in this setting, for reasons that we will come to shortly, to agree to adopt the second convention. Then we have

$$\begin{aligned}
\langle\, [a_i, a_{j+1}) \,|\, \mathbb{V}_{\mathbf{T}} \rangle &= \langle \circ_{a_1} \!-\! \ldots \!-\! \circ_{a_{i-1}} \!-\! \bullet_{a_i} \!-\! \ldots \!-\! \bullet_{a_j} \!-\! \circ_{a_{j+1}} \!-\! \ldots \!-\! \circ_{a_n} \,|\, \mathbb{V} \rangle \\
&= \langle\, \circ_{a_{i-1}} \!-\! \bullet_{a_i} \!-\! \bullet_{a_j} \!-\! \circ_{a_{j+1}} | \,\mathbb{V} \rangle \\
&= \mu_{\mathbb{V}}([a_{i-1}, a_i] \times [a_j, a_{j+1}]).
\end{aligned}$$

If we now draw the interval decomposition of $\mathbb{V}_{\mathbf{T}}$ as a persistence diagram, identifying the half-open intervals $[a_i, a_{j+1})$ with decorated points (a_i^-, a_{j+1}^-) in the usual way, we find that $\mathsf{Dgm}(\mathbb{V}_{\mathbf{T}})$ is obtained by 'snapping' each decorated point of $\mathsf{Dgm}(\mathbb{V})$ upwards and rightwards to the grid determined by $\mathbf{T}$. Figure 3.6 illustrates this for the case $n = 3$. This is the 'snapping principle' of [14, 15].

There are some well known situations where the entire structure of $\mathbb{V}$ determined by its behaviour on a particular finite index set.

Definition 3.50 We say that $\mathbb{V}$ is **Morselike** if there exists a finite set of indices $\mathbf{T} = \{a_1, \ldots, a_n\}$ such that v_t^s is an isomorphism whenever $s < t$ belong to an interval $[a_i, a_{i+1})$ for some $1 \leq i \leq n$, and also $V_t = 0$ for $t < a_1$.

When $\mathbb{V}$ is Morselike, it follows from Lemma 3.39 that if (x^*, y^*) is a point or singular point of $\mathsf{Dgm}(\mathbb{V})$ then

$$x^*, y^* \in \{-\infty^+, a_1^-, a_2^-, \ldots, a_n^-, +\infty^-\}.$$

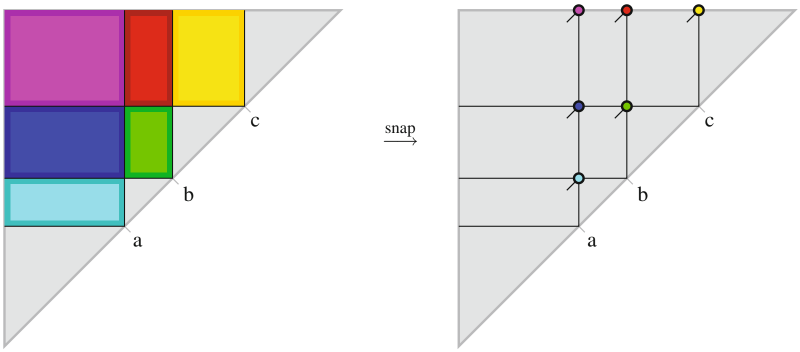


Fig. 3.6 A persistence module $\mathbb{V}$ discretised at $\mathbf{T} = \{a, b, c\}$. The persistence diagram $\mathsf{Dgm}(\mathbb{V}_{\mathbf{T}})$ is localised at six grid vertices, corresponding to the six possible interval summands of a 3-index persistence module. The multiplicity of each vertex of $\mathsf{Dgm}(\mathbb{V}_{\mathbf{T}})$ is equal to the number of decorated points of $\mathsf{Dgm}(\mathbb{V})$ in the rectangle immediately below and to the left of it, and may be computed by evaluating $\mu_{\mathbb{V}}$ on that rectangle. The tick directions indicate lower-closed half-open intervals. Decorated points of $\mathsf{Dgm}(\mathbb{V})$ in the remaining triangular regions do not show up in $\mathsf{Dgm}(\mathbb{V}_{\mathbf{T}})$

The possibility that $x^* = -\infty^+$ is ruled out by the vanishing of $\mathbb{V}$ below a_1. The remaining candidates are the points (a_i^-, a_{j+1}^-) with $1 \leq i \leq j \leq n$. Similarly to Example 3.40, we compute the multiplicity of each candidate by finding a rectangle that contains that candidate alone:

$$\mathrm{m}_{\mathbb{V}}(a_i^-, a_{j+1}^-) = \mu_{\mathbb{V}}([a_{i-1}, a_i] \times [a_j, a_{j+1}]).$$

Thus $\mathsf{Dgm}(\mathbb{V}) = \mathsf{Dgm}(\mathbb{V}_{\mathbf{T}})$ exactly, provided we use the half-open convention for intervals in $\mathbf{T}$ that we agreed on earlier.

Example 3.51 The sublevelset persistent homology $\mathrm{H}(\mathbb{X}_{\mathrm{sub}})$ of a pair (X, f) is Morselike if

- X is a compact manifold and f is a Morse function; or
- X is a compact polyhedron and f is piecewise linear.

Indeed, let $\mathbf{T}$ be the set of critical points of the Morse function, or the set of vertex-values of the piecewise-linear function. Then the inclusion $X^{a_i} \subseteq X^t$ is a homotopy equivalence whenever $t \in [a_i, a_{i+1})$, so $\mathrm{H}(X^s) \to \mathrm{H}(X^t)$ is an isomorphism whenever $s < t$ belong to the same half-open interval $[a_i, a_{i+1})$.

Remark 3.52 Compare Example 3.51 with Corollary 3.47.

Example 3.53 Let S be a finite simplicial complex, and let $f : S \to \mathbf{R}$ be a function on its simplices such that $f(\sigma) \leq f(\tau)$ whenever $\sigma \leq \tau$. This defines a nested family $\mathbb{S}$ of simplicial complexes

$$S^t := \{\sigma \in S \mid f(\sigma) \leq t\}$$

and their inclusions. Then $\mathrm{H}(\mathbb{S})$ is Morselike with respect to the index set $\mathbf{T} = f(S)$, since S^t is constant over each half-open interval $[a_i, a_{i+1})$. This class of examples occurs frequently in topological data analysis; the Vietoris–Rips filtration of a finite metric space is perhaps the most commonly used. The classical algorithms [30, 50] take (S, f) as their input and return the summands of $\mathrm{H}(\mathbb{S})$ as a list of half-open intervals.

Chapter 4
Interleaving

Interleaving is a way of comparing two persistence modules. As with any category, two persistence modules $\mathbb{U}, \mathbb{V}$ are said to be isomorphic if there are maps

$$\Phi \in \mathrm{Hom}(\mathbb{U}, \mathbb{V}), \quad \Psi \in \mathrm{Hom}(\mathbb{V}, \mathbb{U}),$$

such that

$$\Psi\Phi = 1_{\mathbb{U}}, \quad \Phi\Psi = 1_{\mathbb{V}}.$$

This relation is too strong in situations where the persistence modules have been constructed from noisy or uncertain data. The natural response is to consider a weaker relation, **δ-interleaving**, where $\delta \geq 0$ quantifies the uncertainty.

In this section, we define the interleaving relation and study its elementary properties. We prove the nontrivial result (from [14]) that if two persistence modules are δ-interleaved, then they are connected in the space of persistence modules by a path of length δ. This 'interpolation lemma' is a crucial step in the proof of the stability theorem in Chap. 5.

4.1 Shifted Homomorphisms

The first step is to consider homomorphisms which shift the value of the persistence index. Let $\mathbb{U}, \mathbb{V}$ be persistence modules over $\mathbf{R}$, and let δ be any real number. A **homomorphism of degree** δ is a collection Φ of linear maps

$$\phi_t : U_t \to V_{t+\delta}$$

for all $t \in \mathbf{R}$, such that the diagram

F. Chazal et al., *The Structure and Stability of Persistence Modules*,
SpringerBriefs in Mathematics, DOI 10.1007/978-3-319-42545-0_4

$$\begin{array}{ccc} U_s & \xrightarrow{u_t^s} & U_t \\ \downarrow{\scriptstyle \phi_s} & & \downarrow{\scriptstyle \phi_t} \\ V_{s+\delta} & \xrightarrow{v_{t+\delta}^{s+\delta}} & V_{t+\delta} \end{array}$$

commutes whenever $s \le t$.

We write

$$\mathrm{Hom}^{\delta}(\mathbb{U}, \mathbb{V}) = \{\text{homomorphisms } \mathbb{U} \to \mathbb{V} \text{ of degree } \delta\},$$
$$\mathrm{End}^{\delta}(\mathbb{V}) = \{\text{homomorphisms } \mathbb{V} \to \mathbb{V} \text{ tof degree } \delta\}.$$

Composition gives a map

$$\mathrm{Hom}^{\delta_2}(\mathbb{V}, \mathbb{W}) \times \mathrm{Hom}^{\delta_1}(\mathbb{U}, \mathbb{V}) \to \mathrm{Hom}^{\delta_1+\delta_2}(\mathbb{U}, \mathbb{W}).$$

For $\delta \ge 0$, the most important degree-δ endomorphism is the shift map

$$1_{\mathbb{V}}^{\delta} \in \mathrm{End}^{\delta}(\mathbb{V}),$$

which is the collection of maps $(v_{t+\delta}^{t})$ from the persistence structure on $\mathbb{V}$. If Φ is a homomorphism $\mathbb{U} \to \mathbb{V}$ of any degree, then by definition $\Phi 1_{\mathbb{U}}^{\delta} = 1_{\mathbb{V}}^{\delta}\Phi$ for all $\delta \ge 0$.

Remark 4.1 Here is another way to think of morphisms of non-zero degree. For any persistence module $\mathbb{V}$, and $\delta \in \mathbf{R}$, we let $\mathbb{V}[\delta]$ denote the **shifted module**

$$(V[\delta])_t = V_{t+\delta}, \quad (v[\delta])_t^s = v_{t+\delta}^{s+\delta}.$$

In other words, $\mathbb{V}[\delta]$ is obtained from $\mathbb{V}$ by shifting all the information downwards by δ. Then there are obvious identifications

$$\mathrm{Hom}^{\delta}(\mathbb{U}, \mathbb{V}) = \mathrm{Hom}(\mathbb{U}, \mathbb{V}[\delta]) = \mathrm{Hom}(\mathbb{U}[a], \mathbb{V}[a+\delta])$$

for all $a \in \mathbf{R}$. To avoid excessive notation, we will use the same symbol for

$$\Phi = (\phi_t) : \mathbb{U} \to \mathbb{V}[\delta]$$

as for its shifted equivalent

$$\Phi = (\phi_{t+a}) : \mathbb{U}[a] \to \mathbb{V}[a+\delta]$$

since the constituent maps are the same.

4.2 Interleaving

Let $\delta \geq 0$. Two persistence modules $\mathbb{U}, \mathbb{V}$ are said to be δ **-interleaved** if there are maps

$$\Phi \in \mathrm{Hom}^{\delta}(\mathbb{U}, \mathbb{V}), \quad \Psi \in \mathrm{Hom}^{\delta}(\mathbb{V}, \mathbb{U})$$

such that

$$\Psi\Phi = 1_{\mathbb{U}}^{2\delta}, \quad \Phi\Psi = 1_{\mathbb{V}}^{2\delta}.$$

More expansively (with many more indices written out), there are maps

$$\phi_t : U_t \to V_{t+\delta} \quad \text{and} \quad \psi_t : V_t \to U_{t+\delta}$$

defined for all t, such that the following diagrams

$$
\begin{array}{ccc}
U_s & \xrightarrow{u_t^s} & U_t \\
{\scriptstyle \phi_s}\downarrow & & \downarrow{\scriptstyle \phi_t} \\
V_{s+\delta} & \xrightarrow{v_{t+\delta}^{s+\delta}} & V_{t+\delta}
\end{array}
\qquad
\begin{array}{ccc}
U_{s-\delta} & \xrightarrow{u_{s+\delta}^{s-\delta}} & U_{s+\delta} \\
{\scriptstyle \phi_{s-\delta}}\searrow & & \nearrow{\scriptstyle \psi_s} \\
 & V_s &
\end{array}
$$

$$
\begin{array}{ccc}
V_s & \xrightarrow{v_t^s} & V_t \\
{\scriptstyle \psi_s}\downarrow & & \downarrow{\scriptstyle \psi_t} \\
U_{s+\delta} & \xrightarrow{u_{t+\delta}^{s+\delta}} & U_{t+\delta}
\end{array}
\qquad
\begin{array}{ccc}
V_{s-\delta} & \xrightarrow{v_{s+\delta}^{s-\delta}} & V_{s+\delta} \\
{\scriptstyle \psi_{s-\delta}}\searrow & & \nearrow{\scriptstyle \phi_s} \\
 & U_s &
\end{array}
\tag{4.1}
$$

commute for all eligible parameter values; that is, for all $s \leq t$.

Remark 4.2 Where possible, we will be concise rather than expansive.

Example 4.3 Let X be a topological space and let $f, g : X \to \mathbf{R}$. Suppose $\|f - g\|_\infty < \delta$. Then the persistence modules $\mathrm{H}(\mathbb{X}_{\mathrm{sub}}^f), \mathrm{H}(\mathbb{X}_{\mathrm{sub}}^g)$ are δ-interleaved. Indeed, there are inclusions

$$(X, f)^t \subseteq (X, g)^{t+\delta}$$
$$(X, g)^t \subseteq (X, f)^{t+\delta}$$

for all t, which induce maps

$$\Phi : \mathrm{H}(\mathbb{X}_{\mathrm{sub}}^f) \to \mathrm{H}(\mathbb{X}_{\mathrm{sub}}^g)$$
$$\Psi : \mathrm{H}(\mathbb{X}_{\mathrm{sub}}^g) \to \mathrm{H}(\mathbb{X}_{\mathrm{sub}}^f)$$

of degree δ. Since all the maps are induced functorially from inclusion maps, the interleaving relations are automatically satisfied.

This is the situation for which the stability theorem of Cohen-Steiner, Edelsbrunner and Harer [19] was originally stated: if two functions f, g are close then the diagrams for their sublevelset persistent homology are close. Subsequently, stability has been formulated as a theorem about the diagrams of interleaved persistence modules [14, 15]. In the present work, we will come to view stability as a theorem about r-measures.

4.3 Interleaving (Continued)

An interleaving between two persistence modules can be thought of as a persistence module over a certain partially ordered set (poset). We develop this idea next.

Consider the standard partial order on the plane:

$$(p_1, q_1) \leq (p_2, q_2) \quad \Leftrightarrow \quad p_1 \leq p_2 \text{ and } q_1 \leq q_2.$$

For any real number x, define the corresponding shifted diagonal in the plane:

$$\Delta_x = \{(p, q) \mid q - p = 2x\} = \{(t - x, t + x) \mid t \in \mathbf{R}\}$$

As a poset, this is isomorphic to the real line. We will use the specific isomorphism by which $t \in \mathbf{R}$ corresponds to $(t - x, t + x) \in \Delta_x$. This gives a canonical identification between persistence modules over $\mathbf{R}$ and persistence modules over Δ_x.

Proposition 4.4 *Let x, y be real numbers. Persistence modules $\mathbb{U}, \mathbb{V}$ are $|y - x|$-interleaved if and only if there is a persistence module $\mathbb{W}$ over $\Delta_x \cup \Delta_y$ such that $\mathbb{W}|_{\Delta_x} = \mathbb{U}$ and $\mathbb{W}|_{\Delta_y} = \mathbb{V}$. Here $\Delta_x \cup \Delta_y$ is regarded as a subposet of $\mathbf{R}^2$.*

Proof Assume $x < y$ without loss of generality. We claim that (i) the extra information carried by $(y - x)$-interleaving maps Φ, Ψ is equivalent to (ii) the extra information carried by $\mathbb{W}$. Let us describe both, more carefully:

(i) In addition to $\mathbb{U}, \mathbb{V}$ we have a system of maps $\Phi = (\phi_t)$, where

$$\phi_t : U_t \to V_{t+y-x},$$

and a system of maps $\Psi = (\psi_t)$, where

$$\psi_t : V_t \to U_{t+y-x}.$$

These are constrained by the relations (for all $\eta \geq 0$).

$$\varPhi 1_{\mathbb{U}}^{\eta} = 1_{\mathbb{V}}^{\eta}\varPhi, \quad \varPsi 1_{\mathbb{V}}^{\eta} = 1_{\mathbb{U}}^{\eta}\varPsi, \quad \varPsi\varPhi = 1_{\mathbb{U}}^{2y-2x}, \quad \varPhi\varPsi = 1_{\mathbb{V}}^{2y-2x}. \tag{4.2}$$

There are no other constraints.

(ii) In addition to $\mathbb{U}$, $\mathbb{V}$ the persistence module $\mathbb{W}$ carries maps between the two components $\varDelta_x$, $\varDelta_y$. These maps are constrained by the composition law

$$w_T^R = w_T^S \circ w_S^R$$

for all $R, S, T \in \varDelta_x \cup \varDelta_y$ with $R \leq S \leq T$.

First, observe that we recover the maps ϕ_t, ψ_t as vertical maps from $\varDelta_x$ to $\varDelta_y$, and horizontal maps from $\varDelta_y$ to $\varDelta_x$, respectively (see Fig. 4.1):

$$\begin{aligned} U_t &= W_{(t-x,t+x)} \to W_{(t-x,t+2y-x)} = V_{t+y-x} \\ V_t &= W_{(t-y,t+y)} \to W_{(t+y-2x,t+y)} = U_{t+y-x} \end{aligned}$$

Next, observe that the composition law implies all of the relations (4.2).

Finally, there is no additional information in $\mathbb{W}$, beyond the interleaving maps and relations. Indeed, all remaining maps w_T^S, where $S \leq T$, can all be factored in the form:

$$\begin{aligned} w_T^S &= v_t^{s+y-x} \circ \phi_s \quad &\text{if } S \in \varDelta_x \text{ and } T \in \varDelta_y, \\ w_T^S &= u_t^{s+y-x} \circ \psi_s \quad &\text{if } S \in \varDelta_y \text{ and } T \in \varDelta_x. \end{aligned}$$

Thus each map in $\mathbb{W}$ is an instance of one of

$$\begin{aligned} &1_{\mathbb{U}}^{\eta} && \text{from } \varDelta_x \text{ to } \varDelta_x, \\ &1_{\mathbb{V}}^{\eta} && \text{from } \varDelta_y \text{ to } \varDelta_y, \\ &1_{\mathbb{V}}^{\eta}\varPhi && \text{from } \varDelta_x \text{ to } \varDelta_y, \\ &1_{\mathbb{U}}^{\eta}\varPsi && \text{from } \varDelta_y \text{ to } \varDelta_x. \end{aligned}$$

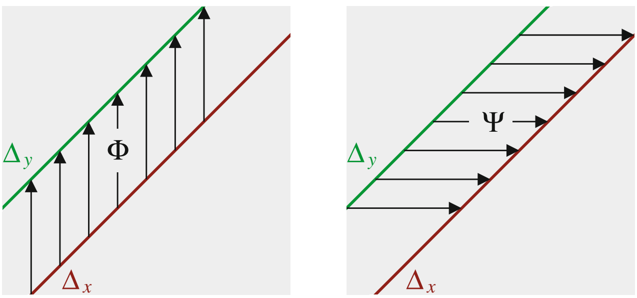


Fig. 4.1 The maps $\varPhi$, $\varPsi$ recovered from the module $\mathbb{W}$ over $\varDelta_x \cup \varDelta_y$

It is a simple matter to verify that the composition law is satisfied for each composable pair of maps. For instance:

$$(1_{\mathbb{V}}^{\eta}\Phi)(1_{\mathbb{U}}^{\zeta}\Psi) = 1_{\mathbb{V}}^{\eta}\Phi 1_{\mathbb{U}}^{\zeta}\Psi = 1_{\mathbb{V}}^{\eta}1_{\mathbb{V}}^{\zeta}\Phi\Psi = 1_{\mathbb{V}}^{\eta+\zeta}1_{\mathbb{V}}^{2y-2x} = 1_{\mathbb{V}}^{\eta+\zeta+2y-2x}$$

This can be done using only the known relations, so there are no further constraints on the w_T^S. □

Remark 4.5 This characterisation makes it clear (or, in another view, depends on the fact) that all composable combinations of the maps u, v, ϕ, ψ from a given domain to a given codomain must be equal: indeed, they must agree with the appropriate map w_T^S of $\mathbb{W}$.

4.4 The Interpolation Lemma

In this section we prove a crucial result from [14]:

Lemma 4.6 (interpolation lemma) *Suppose $\mathbb{U}$, $\mathbb{V}$ are a δ-interleaved pair of persistence modules. Then there exists a 1-parameter family of persistence modules $(\mathbb{U}_x \mid x \in [0, \delta])$ such that $\mathbb{U}_0, \mathbb{U}_\delta$ are equal to $\mathbb{U}, \mathbb{V}$ respectively, and $\mathbb{U}_x, \mathbb{U}_y$ are $|y - x|$-interleaved for all $x, y \in [0, \delta]$. Moreover, if $\mathbb{U}, \mathbb{V}$ are q-tame then the $(\mathbb{U}_x)$ may be assumed q-tame also.*

We prove something sharper: given a specific pair of interleaving maps

$$\Phi \in \mathrm{Hom}^{\delta}(\mathbb{U}, \mathbb{V}) \qquad \Psi \in \mathrm{Hom}^{\delta}(\mathbb{V}, \mathbb{U})$$

we explicitly provide, for each $x < y$ in $[0, \delta]$, a pair of interleaving maps

$$\Phi_y^x \in \mathrm{Hom}^{y-x}(\mathbb{U}_x, \mathbb{U}_y) \qquad \Psi_x^y \in \mathrm{Hom}^{y-x}(\mathbb{U}_y, \mathbb{U}_x)$$

such that $\Phi_\delta^0 = \Phi$ and $\Psi_0^\delta = \Psi$, and moreover

$$\Phi_z^y \Phi_y^x = \Phi_z^x \qquad \Phi_x^y \Phi_y^z = \Phi_x^z$$

for all $x < y < z$. In view of Proposition 4.4, this sharp form of the interpolation lemma can be restated as follows.

Theorem 4.7 (interpolation lemma, version 2) *Any persistence module $\mathbb{W}$ over $\Delta_0 \cup \Delta_\delta$ extends to a persistence module $\overline{\mathbb{W}}$ over the diagonal strip*

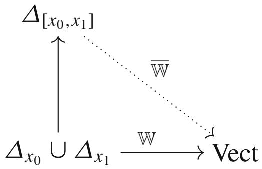
$$\Delta_{[0,\delta]} = \{(p, q) \mid 0 \le q - p \le 2\delta\} \subset \mathbf{R}^2.$$

If $\mathbb{W}|_{\Delta_0}, \mathbb{W}|_{\Delta_\delta}$ are q-tame, then the extension may be chosen so that each $\overline{\mathbb{W}}|_{\Delta_x}$ is q-tame.

Remark 4.8 The extension is by no means unique.

Let us clarify how Theorem 4.7 implies Lemma 4.6. If $\mathbb{U}$, $\mathbb{V}$ are δ-interleaved, then there exists a persistence module $\mathbb{W}$ over $\Delta_0 \cup \Delta_\delta$ such that $\mathbb{W}|_{\Delta_0} = \mathbb{U}$ and $\mathbb{W}|_{\Delta_\delta} = \mathbb{V}$. By Theorem 4.7, this extends to $\overline{\mathbb{W}}$ over the strip $\Delta_{[0,\delta]}$. If we define a 1-parameter family $\mathbb{U}_x = \overline{\mathbb{W}}|_{\Delta_x}$, then $\mathbb{U}_x, \mathbb{U}_y$ are $|x - y|$-interleaved for all $x, y \in [0, \delta]$.

For readers familiar with Kan extensions in category theory [43], here is a very short proof of the theorem. Let us regard the posets $\Delta_0 \cup \Delta_\delta$ and $\Delta_{[0,\delta]}$ as categories (see Remark 2.1); then persistence modules over these posets are the same as functors to the category of vector spaces. The theorem asserts the existence of an extension $\overline{\mathbb{W}}$



for any functor $\mathbb{W}$. Peter Bubenik has pointed out to us that the Kan extension theorem immediately provides two such extensions, since the category Vect is both complete (which yields the 'right-extension') and co-complete (which yields the 'left-extension').

We proceed now to a detailed proof, for those who would still like one.

Proof (Theorem 4.7) In order to express the proof more symmetrically, it is convenient to replace the interval $[0, \delta]$ by the interval $[-1, 1]$. This can be done by rescaling and translating the plane. Accordingly, suppose we are given a persistence module $\mathbb{W}$ over $\Delta_{-1} \cup \Delta_1$.

Our strategy is to construct two persistence modules over the strip $\Delta_{[-1,1]}$ and a module map between them. The image (or coimage) of this map is itself a persistence module over the strip, and will be the required extension.

By Proposition 4.4, $\mathbb{W}$ provides $\mathbb{U} = \mathbb{W}|_{\Delta_{-1}}$ and $\mathbb{V} = \mathbb{W}|_{\Delta_1}$, which we can view as persistence modules over $\mathbf{R}$ using the canonical parametrisation:

$$U_t = W_{(t+1,t-1)} \quad \text{and} \quad V_t = W_{(t-1,t+1)}$$

and corresponding linear maps u_t^s, v_t^s. The module $\mathbb{W}$ also provides interleaving maps $\Phi \in \mathrm{Hom}^2(\mathbb{U}, \mathbb{V})$ and $\Psi \in \mathrm{Hom}^2(\mathbb{V}, \mathbb{U})$ of degree 2:

$$\phi_t = w_{(t+1,t+3)}^{(t+1,t-1)} : U_t \to V_{t+2}, \quad \psi_t = w_{(t+3,t+1)}^{(t-1,t+1)} : V_t \to U_{t+2},$$

From $\mathbb{U}$, $\mathbb{V}$ we construct four persistence modules over $\mathbf{R}^2$:

$$\begin{array}{llll}
\mathbb{A} & \text{defined by} \quad A_{(p,q)} = U_{p-1} & \text{and} & a_{(r,s)}^{(p,q)} = u_{r-1}^{p-1} \\
\mathbb{B} & \text{defined by} \quad B_{(p,q)} = V_{q-1} & \text{and} & b_{(r,s)}^{(p,q)} = v_{s-1}^{q-1} \\
\mathbb{C} & \text{defined by} \quad C_{(p,q)} = U_{q+1} & \text{and} & c_{(r,s)}^{(p,q)} = u_{s+1}^{q+1} \\
\mathbb{D} & \text{defined by} \quad D_{(p,q)} = V_{p+1} & \text{and} & d_{(r,s)}^{(p,q)} = v_{r+1}^{p+1}
\end{array}$$

Note that $\mathbb{A}$, $\mathbb{B}$, $\mathbb{C}$, $\mathbb{D}$ are the vertical and horizontal extensions of the modules $\mathbb{W}|_{\Delta_{\pm 1}}$ to the whole plane. In particular, we have

$$\mathbb{A}|_{\Delta_{-1}} = \mathbb{U}, \quad \mathbb{B}|_{\Delta_1} = \mathbb{V}, \quad \mathbb{C}|_{\Delta_{-1}} = \mathbb{U}, \quad \mathbb{D}|_{\Delta_1} = \mathbb{V},$$

with respect to our canonical parametrisations of each diagonal. Restricting each module to its 'other' diagonal, on the other hand, we have

$$\mathbb{A}|_{\Delta_1} = \mathbb{U}[-2], \quad \mathbb{B}|_{\Delta_{-1}} = \mathbb{V}[-2], \quad \mathbb{C}|_{\Delta_1} = \mathbb{U}[2], \quad \mathbb{D}|_{\Delta_{-1}} = \mathbb{V}[2],$$

using the 'shifted module' notation of Remark 4.1.

Next, we construct four module maps:

$$\begin{array}{lll}
\overline{1}_{\mathbb{U}} : \mathbb{A} \to \mathbb{C} & \text{defined at } (p,q) \text{ to be} & u_{q+1}^{p-1} : U_{p-1} \to U_{q+1} \\
\overline{\Phi} \ : \mathbb{A} \to \mathbb{D} & \text{defined at } (p,q) \text{ to be} & \phi_{p-1} : U_{p-1} \to V_{p+1} \\
\overline{\Psi} \ : \mathbb{B} \to \mathbb{C} & \text{defined at } (p,q) \text{ to be} & \psi_{q-1} : V_{q-1} \to U_{q+1} \\
\overline{1}_{\mathbb{V}} : \mathbb{B} \to \mathbb{D} & \text{defined at } (p,q) \text{ to be} & v_{p+1}^{q-1} : V_{q-1} \to V_{p+1}
\end{array}$$

The maps $\overline{\Phi}$, $\overline{\Psi}$ are defined over the whole plane, whereas $\overline{1}_{\mathbb{U}}$ is defined only where $p - 1 \leq q + 1$, and $\overline{1}_{\mathbb{V}}$ is defined only where $q - 1 \leq p + 1$. To verify that the four definitions give module maps, it is enough to observe that the required commutation relations involve composable combinations of the maps u, v, ϕ, ψ, which always agree by Remark 4.5.

Note that the intersection of the regions of definition, where all four maps are defined, is precisely the strip $\Delta_{[-1,1]}$. Henceforth, we restrict $\mathbb{A}$, $\mathbb{B}$, $\mathbb{C}$, $\mathbb{D}$ and the four maps to that strip.

Define $\Omega \in \mathrm{Hom}(\mathbb{A} \oplus \mathbb{B}, \mathbb{C} \oplus \mathbb{D})$ by the 2-by-2 matrix

$$\begin{bmatrix} \overline{1}_{\mathbb{U}} & \overline{\Psi} \\ \overline{\Phi} & \overline{1}_{\mathbb{V}} \end{bmatrix}$$

of module maps. Our claim is that $\overline{\mathbb{W}} = \mathrm{im}(\Omega)$ is the required extension. We may equivalently claim that $\overline{\mathbb{W}} = \mathrm{coim}(\Omega) = (\mathsf{A} \oplus \mathsf{B})/\ker(\Omega)$ is the required extension.[1]

Step 1. $\overline{\mathbb{W}}|_{\Delta_{-1}}$ is isomorphic to $\mathbb{U}$.

Proof On the diagonal Δ_{-1} we have

$$(\mathbb{A} \oplus \mathbb{B})|_{\Delta_{-1}} = \mathbb{U} \oplus \mathbb{V}[-2], \quad (\mathbb{C} \oplus \mathbb{D})|_{\Delta_{-1}} = \mathbb{U} \oplus \mathbb{V}[2],$$

and the homomorphism $\Omega|_{\Delta_{-1}}$ takes the matrix form

$$\begin{bmatrix} 1_{\mathbb{U}} & \Psi \\ \Phi & 1_{\mathbb{V}}^4 \end{bmatrix}.$$

Since $1_{\mathbb{V}}^4 = \Phi\Psi$, this factorises as

$$\mathbb{U} \oplus \mathbb{V}[-2] \xrightarrow{\Omega_1} \mathbb{U} \xrightarrow{\Omega_2} \mathbb{U} \oplus \mathbb{V}[2]$$

where

$$\Omega_1 = [\, 1_{\mathbb{U}} \; \Psi \,] \quad \text{and} \quad \Omega_2 = \begin{bmatrix} 1_{\mathbb{U}} \\ \Phi \end{bmatrix}$$

in matrix form. Thanks to the $1_{\mathbb{U}}$ entries, we see that Ω_1 is surjective and Ω_2 is injective. This implies that Ω_1, Ω_2 induce isomorphisms

$$\mathrm{coim}(\Omega|_{\Delta_{-1}}) \xrightarrow{\cong} \mathbb{U} \xrightarrow{\cong} \mathrm{im}(\Omega|_{\Delta_{-1}})$$

as required. □

Step 2. $\overline{\mathbb{W}}|_{\Delta_1}$ is isomorphic to $\mathbb{V}$.

Proof On the diagonal Δ_1 we have

$$(\mathbb{A} \oplus \mathbb{B})_{\Delta_1} = \mathbb{U}[-2] \oplus \mathbb{V}, \quad (\mathbb{C} \oplus \mathbb{D})_{\Delta_1} = \mathbb{U}[2] \oplus \mathbb{V},$$

and the homomorphism $\Omega|_{\Delta_1}$ takes the matrix form

$$\begin{bmatrix} 1_{\mathbb{U}}^4 & \Psi \\ \Phi & 1_{\mathbb{V}} \end{bmatrix}.$$

[1] The image and the coimage are naturally isomorphic; the difference is whether we wish to think of $\overline{\mathbb{W}}$ as a submodule of $\mathbb{C} \oplus \mathbb{D}$ or as a quotient module of $\mathbb{A} \oplus \mathbb{B}$. In the following pages, we will treat the two points of view with equal emphasis.

Since $1^4_{\mathbb{U}} = \Psi\Phi$, this factorises as

$$\mathbb{U}[-2] \oplus \mathbb{V} \xrightarrow{\Omega_3} \mathbb{V} \xrightarrow{\Omega_4} \mathbb{U}[2] \oplus \mathbb{V}$$

where

$$\Omega_3 = \begin{bmatrix} \Phi & 1_{\mathbb{V}} \end{bmatrix} \quad \text{and} \quad \Omega_4 = \begin{bmatrix} \Psi \\ 1_{\mathbb{V}} \end{bmatrix}$$

in matrix form. Thanks to the $1_{\mathbb{V}}$ entries, we see that Ω_3 is surjective and Ω_4 is injective. This implies that Ω_3, Ω_4 induce isomorphisms

$$\mathrm{coim}(\Omega|_{\Delta_1}) \xrightarrow{\cong} \mathbb{V} \xrightarrow{\cong} \mathrm{im}(\Omega|_{\Delta_1})$$

as required. □

Step 3. The cross-maps of $\overline{\mathbb{W}}$ between Δ_{-1} and Δ_1 correspond to Φ and Ψ under the isomorphisms of Steps 1 and 2.

Proof The cross maps for $\overline{\mathbb{W}}$ are induced by the cross maps for $\mathbb{A} \oplus \mathbb{B}$ (if we view $\overline{\mathbb{W}}$ as a coimage) and equally by the cross maps for $\mathbb{C} \oplus \mathbb{D}$ (if we view $\overline{\mathbb{W}}$ as an image). The vertical cross-map for $\mathbb{A} \oplus \mathbb{B}$ is a map

$$(\mathbb{A} \oplus \mathbb{B})|_{\Delta_{-1}} \longrightarrow (\mathbb{A} \oplus \mathbb{B})|_{\Delta_1}$$

of degree 2 which we can identify as

$$\mathbb{U} \oplus \mathbb{V}[-2] \xrightarrow{1_{\mathbb{U}} \oplus 1^4_{\mathbb{V}}} \mathbb{U} \oplus \mathbb{V}[2].$$

Alternatively, the vertical cross-map for $\mathbb{C} \oplus \mathbb{D}$ is a map

$$(\mathbb{C} \oplus \mathbb{D})|_{\Delta_{-1}} \longrightarrow (\mathbb{C} \oplus \mathbb{D})|_{\Delta_1}$$

of degree 2 which we can identify as

$$\mathbb{U} \oplus \mathbb{V}[2] \xrightarrow{1^4_{\mathbb{U}} \oplus 1_{\mathbb{V}}} \mathbb{U}[4] \oplus \mathbb{V}[2].$$

The following diagram shows the vertical cross-maps for $\mathbb{A} \oplus \mathbb{B}$ (on the left) and $\mathbb{C} \oplus \mathbb{D}$ (on the right), the factorisations of Steps 1 and 2, and the map Φ.

$$\begin{CD}
\mathbb{U} \oplus \mathbb{V}[-2] @>{\Omega_1}>> \mathbb{U} @>{\Omega_2}>> \mathbb{U} \oplus \mathbb{V}[2] \\
@V{1_{\mathbb{U}} \oplus 1_{\mathbb{V}}^4}VV @V{\Phi}VV @VV{1_{\mathbb{U}}^4 \oplus 1_{\mathbb{V}}}V \\
\mathbb{U} \oplus \mathbb{V}[2] @>{\Omega_3}>> \mathbb{V}[2] @>{\Omega_4}>> \mathbb{U}[4] \oplus \mathbb{V}[2]
\end{CD}$$

It is enough to show that either square commutes. And indeed

$$\Omega_3(1_{\mathbb{U}} \oplus 1_{\mathbb{V}}^4) = \begin{bmatrix} \Phi & 1_{\mathbb{V}} \end{bmatrix} \begin{bmatrix} 1_{\mathbb{U}} & 0 \\ 0 & 1_{\mathbb{V}}^4 \end{bmatrix} = [\, \Phi \;\; 1_{\mathbb{V}}^4 \,] = [\Phi][\, 1_{\mathbb{U}} \;\; \Psi \,] = \Phi \Omega_1$$

for the left square, and

$$(1_{\mathbb{U}}^4 \oplus 1_{\mathbb{V}})\Omega_2 = \begin{bmatrix} 1_{\mathbb{U}}^4 & 0 \\ 0 & 1_{\mathbb{V}} \end{bmatrix} \begin{bmatrix} 1_{\mathbb{U}} \\ \Phi \end{bmatrix} = \begin{bmatrix} 1_{\mathbb{U}}^4 \\ \Phi \end{bmatrix} = \begin{bmatrix} \Psi \\ 1_{\mathbb{V}} \end{bmatrix} [\Phi] = \Omega_4 \Phi$$

for the right square. Thus the induced vertical cross-map corresponds to Φ.

A similar argument using the diagram

$$\begin{CD}
\mathbb{U}[2] \oplus \mathbb{V} @>{\Omega_3}>> \mathbb{V} @>{\Omega_4}>> \mathbb{U}[2] \oplus \mathbb{V} \\
@V{1_{\mathbb{U}}^4 \oplus 1_{\mathbb{V}}}VV @V{\Psi}VV @VV{1_{\mathbb{U}} \oplus 1_{\mathbb{V}}^4}V \\
\mathbb{U}[2] \oplus \mathbb{V} @>{\Omega_1}>> \mathbb{U}[2] @>{\Omega_2}>> \mathbb{U}[2] \oplus \mathbb{V}[4]
\end{CD}$$

shows that the induced horizontal cross-map corresponds to Ψ. □

This completes the construction of the extension $\overline{\mathbb{W}}$. Now we verify the last assertion of theorem. Suppose that $\mathbb{U}$, $\mathbb{V}$ are q-tame, meaning that their non-identity structure maps have finite rank. On any diagonal Δ_x, the restricted modules $\mathbb{A}|_{\Delta_x}$, $\mathbb{B}|_{\Delta_x}$ are shifted copies of $\mathbb{U}$, $\mathbb{V}$ so they are q-tame. It follows that the direct sum $\mathbb{A} \oplus \mathbb{B}|_{\Delta_x}$ and its homomorphic image $\mathbb{U}_x = \Omega(\mathbb{A} \oplus \mathbb{B})|_{\Delta_x}$ are q-tame.

This completes the proof of Theorem 4.7. □

We point out that Step 3 isn't necessary to deduce the interpolation Lemma 4.6. It is sufficient to show that $\mathbb{U} = \mathbb{U}_{-1}$ and $\mathbb{V} = \mathbb{U}_1$ for some 1-parameter family of persistence modules such that each pair $\mathbb{U}_x$, $\mathbb{U}_y$ admits an $|x - y|$-interleaving. This already follows from Steps 1 and 2. We do not need to know that the 2-interleaving of $\mathbb{U}$, $\mathbb{V}$ induced from $\overline{\mathbb{W}}$ is equal to the original 2-interleaving.

4.5 The Interpolation Lemma (Continued)

In this optional section, we study the interpolation lemma in greater depth. The results are not used elsewhere. Given two modules $\mathbb{U}$, $\mathbb{V}$ and a δ-interleaving between them, there are at least three natural ways to construct an interpolation. We describe these constructions and some relationships between them.

As in the proof of Theorem 4.7, we may suppose that $\delta = 2$ and that $\mathbb{U}$, $\mathbb{V}$ and their interleaving are represented as a module over $\Delta_{-1} \cup \Delta_1$ in the plane, which we wish to extend to a module over $\Delta_{[-1,1]}$.

It will be helpful to introduce some temporary notation. Let $\mathbb{V}$ be a persistence module over $\mathbf{R}$. Then $\mathbb{V}^{\mathrm{P}}$, $\mathbb{V}^{\mathrm{Q}}$ are the persistence modules over $\mathbf{R}^2$ defined by

$$(V^{\mathrm{P}})_{(p,q)} = V_p, \quad (V^{\mathrm{Q}})_{(p,q)} = V_q,$$

and the canonical linear maps.

Now consider the sequence

$$\begin{matrix} \mathbb{U}[-3]^{\mathrm{Q}} & & \mathbb{U}[-1]^{\mathrm{P}} & & \mathbb{U}[1]^{\mathrm{Q}} & & \mathbb{U}[3]^{\mathrm{P}} \\ \oplus & \xrightarrow{\Omega'} & \oplus & \xrightarrow{\Omega} & \oplus & \xrightarrow{\Omega''} & \oplus \\ \mathbb{V}[-3]^{\mathrm{P}} & & \mathbb{V}[-1]^{\mathrm{Q}} & & \mathbb{V}[1]^{\mathrm{P}} & & \mathbb{V}[3]^{\mathrm{Q}} \end{matrix} \tag{4.3}$$

of modules over $\Delta_{[-1,1]}$ with maps

$$\Omega' = \begin{bmatrix} \overline{1_{\mathbb{U}}} & -\overline{\Psi} \\ -\overline{\Phi} & \overline{1_{\mathbb{V}}} \end{bmatrix}, \quad \Omega = \begin{bmatrix} \overline{1_{\mathbb{U}}} & \overline{\Psi} \\ \overline{\Phi} & \overline{1_{\mathbb{V}}} \end{bmatrix}, \quad \Omega'' = \begin{bmatrix} \overline{1_{\mathbb{U}}} & -\overline{\Psi} \\ -\overline{\Phi} & \overline{1_{\mathbb{V}}} \end{bmatrix}$$

defined analogously to Ω from the proof of Theorem 4.7.

Notice that Ω, Ω' and Ω'' are essentially the same map. Certainly Ω', Ω'' are formally identical, up to a translation τ of the strip. In fact, each of the modules in the sequence is related to the next by an isomorphism σ which changes the sign of the $\mathbb{V}$-term and transforms indices by $(p, q) \mapsto (q + 2, p + 2)$. We have $\tau = \sigma^2$, and conjugacies $\Omega = \sigma \Omega' \sigma^{-1}$ and $\Omega'' = \sigma \Omega \sigma^{-1}$.

Proposition 4.9 *Each of the three modules*

$$\mathrm{coker}(\Omega'), \quad \mathrm{coim}(\Omega) = \mathrm{im}(\Omega), \quad \ker(\Omega'')$$

over $\Delta_{[-1,1]}$ defines an interpolating family between $\mathbb{U}$, $\mathbb{V}$.

Proof We already know this for $\mathrm{coim}(\Omega) = \mathrm{im}(\Omega)$ from the proof of Theorem 4.7. Now we outline the proof that $\mathrm{coker}(\Omega')$ and $\ker(\Omega'')$ restrict on Δ_{-1} to modules isomorphic to $\mathbb{U}$.

On the diagonal Δ_{-1} the sequence (4.3) restricts to:

$$\begin{matrix} \mathbb{U}[-4] \\ \oplus \\ \mathbb{V}[-2] \end{matrix} \xrightarrow{\Omega'} \begin{matrix} \mathbb{U} \\ \oplus \\ \mathbb{V}[-2] \end{matrix} \xrightarrow{\Omega} \begin{matrix} \mathbb{U} \\ \oplus \\ \mathbb{V}[2] \end{matrix} \xrightarrow{\Omega''} \begin{matrix} \mathbb{U}[4] \\ \oplus \\ \mathbb{V}[2] \end{matrix}$$

and we have factorisations

$$\Omega' \text{ or } \Omega'' = \begin{bmatrix} 1_{\mathbb{U}}^4 & -\Psi \\ -\Phi & 1_{\mathbb{V}} \end{bmatrix} = \begin{bmatrix} -\Psi \\ 1_{\mathbb{V}} \end{bmatrix} \begin{bmatrix} -\Phi & 1_{\mathbb{V}} \end{bmatrix} = \Omega'_1\Omega'_2 \text{ or } \Omega''_1\Omega''_2.$$

These reveal that $\mathrm{im}(\Omega') = \mathrm{im}(\Omega'_1)$ is a complementary submodule to $\mathbb{U} \oplus 0$ in $\mathbb{U} \oplus \mathbb{V}[t-2]$, and that $\ker(\Omega'') = \ker(\Omega''_2)$ is a complementary submodule to $0 \oplus \mathbb{V}[t+2]$ in $\mathbb{U} \oplus \mathbb{V}[t+2]$. It follows that $\mathrm{coker}(\Omega')$ and $\ker(\Omega'')$ are each isomorphic to $\mathbb{U}$.

By a symmetric argument, the restriction of each module to Δ_1 is isomorphic to $\mathbb{V}$. This completes the proof that $\mathrm{coker}(\Omega')$ and $\ker(\Omega'')$ interpolate between $\mathbb{U}$ and $\mathbb{V}$. □

Which of the three constructions should we prefer? It turns out that $\mathrm{coker}(\Omega')$ and $\ker(\Omega'')$ are respectively isomorphic to the Kan left- and right-extensions, so these are natural from the category theoretic point of view. Now observe that $\Omega\Omega' = 0$ and $\Omega''\Omega = 0$, meaning that (4.3) is a chain complex. It follows that there is a natural projection and a natural inclusion

$$\mathrm{coker}(\Omega') \twoheadrightarrow \mathrm{coim}(\Omega) = \mathrm{im}(\Omega) \hookrightarrow \ker(\Omega'')$$

by which we see that $\mathrm{coim}(\Omega) = \mathrm{im}(\Omega)$ is isomorphic to the image of the composite map $\mathrm{coker}(\Omega') \to \ker(\Omega'')$. In this sense, it is intermediate between the left- and right-extensions; and structurally it is the 'smallest' of the three, being a quotient of one and a subobject of the other.

The surplus information carried by the two Kan extensions may be measured as the kernel of the projection and the cokernel of the inclusion. These are precisely the homology at the second and third terms of (4.3). It follows from the conjugacies described above that the two homology modules are isomorphic upon translating the strip by 2 and interchanging p and q (i.e. reversing the interpolation parameter).

We can use the 'vineyard' technique of [23] to visualise the 1-parameter family of persistence modules produced by each of the three constructions. We obtained the vineyards by sketching the supports of the eight module summands in (4.3) and using the sketches to partition the interpolation parameter range $[-1, 1]$ into suitable intervals for case splitting. It is perhaps easier done than described, so we invite readers to conduct their own calculations and confirm that our vineyards are correct. As further corroboration, one verifies that the homology modules are isomorphic in the sense described above.

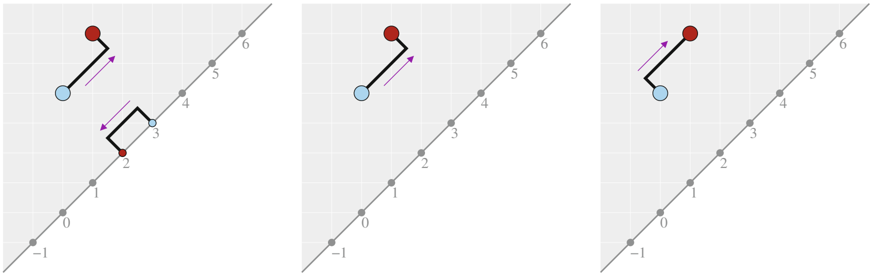


Fig. 4.2 Vineyards of the cokernel (*left*), image (*middle*), and kernel (*right*) interpolations for the 2-interleaving between $\mathbf{k}[0, 4)$ and $\mathbf{k}[1, 6)$

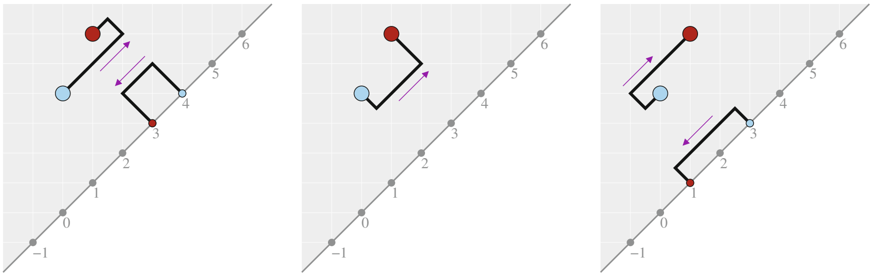


Fig. 4.3 Vineyards of the cokernel (*left*), image (*middle*), and kernel (*right*) interpolations for the 3-interleaving between $\mathbf{k}[0, 4)$ and $\mathbf{k}[1, 6)$

In Fig. 4.2, we consider the canonical 2-interleaving between interval modules $\mathbf{k}[0, 4)$ and $\mathbf{k}[1, 6)$. The thick black lines show how the points of the persistence diagram travel in the plane as we proceed along the interpolating family, for each of the three constructions. Each point travels with speed 1 and traverses a path of length 2 (in the d^∞-metric). The cokernel interpolation has an extra 'ghost' summand which emerges from the diagonal at (3, 3) at the beginning of the interpolation, and is reabsorbed by the diagonal at (2, 2) at the end.

In Fig. 4.3 we repeat the exercise using the canonical 3-interleaving between $\mathbf{k}[0, 4)$ and $\mathbf{k}[1, 6)$. The thick black paths now have length three, and the kernel and cokernel interpolations both produce 'ghosts' at the diagonal.

Chapter 5
The Isometry Theorem

In this section we discuss the metric relationship between persistence modules and their persistence diagrams. As in Chap. 4, all persistence modules are indexed by **R** unless explicitly stated otherwise.

The principal result is the famous stability theorem of Cohen-Steiner, Edelsbrunner and Harer [19], in the generality established by [15]. The main difference is that we emphasise persistence measures, deriving the standard theorem from a more general statement about measures. The structure of the proof remains the same as in [19].

The secondary result is the converse inequality, which together with the stability theorem implies that the space of q-tame persistence modules is isometric with the space of locally finite persistence diagrams. This isometry theorem appeared originally in the work of Lesnick [42] for modules which satisfy $\dim(V_t) < \infty$ for all t, and independently in the work of Bubenik and Scott [5] for modules of finite type.

5.1 The Interleaving Distance

In this section we define the interleaving distance between persistence modules. This was introduced in [15].

The first observation is that if $\mathbb{U}$ and $\mathbb{V}$ are δ-interleaved, then they are $(\delta + \varepsilon)$-interleaved for every $\varepsilon > 0$. Indeed, the maps

$$\Phi' = \Phi 1_{\mathbb{U}}^{\varepsilon} = 1_{\mathbb{V}}^{\varepsilon}\Phi$$
$$\Psi' = \Psi 1_{\mathbb{V}}^{\varepsilon} = 1_{\mathbb{U}}^{\varepsilon}\Psi$$

provide the required interleaving.

F. Chazal et al., *The Structure and Stability of Persistence Modules*,
SpringerBriefs in Mathematics, DOI 10.1007/978-3-319-42545-0_5

The challenge, then, if two persistence modules are interleaved, is to make the interleaving parameter as small as possible. The minimum is not necessarily attained, so we introduce some additional terminology: we say that two persistence modules $\mathbb{U}, \mathbb{V}$ are δ^+**-interleaved** if they are $(\delta + \varepsilon)$-interleaved for all $\varepsilon > 0$. This does not imply that $\mathbb{U}, \mathbb{V}$ are δ-interleaved, as we see now:

Example 5.1 Two persistence modules are 0-interleaved if and only if they are isomorphic.

Example 5.2 A persistence module $\mathbb{V}$ is **ephemeral** if $v_t^s = 0$ for all $s < t$. An ephemeral module may be constructed by selecting an arbitrary family of spaces (V_t) and setting every v_t^s to be zero. Let $\mathbb{U}$ and $\mathbb{V}$ be a pair of non-isomorphic ephemeral modules. Then $\mathbb{U}, \mathbb{V}$ are 0^+-interleaved but not 0-interleaved. Indeed, $1_{\mathbb{U}}^{2\varepsilon} = 0$ and $1_{\mathbb{V}}^{2\varepsilon} = 0$ for all $\varepsilon > 0$, so the zero maps

$$\begin{aligned} \Phi &= 0 \in \mathrm{Hom}^{\varepsilon}(\mathbb{U}, \mathbb{V}) \\ \Psi &= 0 \in \mathrm{Hom}^{\varepsilon}(\mathbb{V}, \mathbb{U}) \end{aligned}$$

constitute an ε-interleaving.

The **interleaving distance** between two persistence modules is defined:

$$\begin{aligned} \mathrm{d_i}(\mathbb{U}, \mathbb{V}) &= \inf\{\delta \mid \mathbb{U}, \mathbb{V} \text{ are } \delta \text{ -interleaved}\} \\ &= \min\{\delta \mid \mathbb{U}, \mathbb{V} \text{ are } \delta^+ \text{ -interleaved}\} \end{aligned}$$

If there is no δ-interleaving between $\mathbb{U}, \mathbb{V}$ for any value of δ, then $\mathrm{d_i}(\mathbb{U}, \mathbb{V}) = \infty$.

Proposition 5.3 *The interleaving distance satisfies the triangle inequality:*

$$\mathrm{d_i}(\mathbb{U}, \mathbb{W}) \leq \mathrm{d_i}(\mathbb{U}, \mathbb{V}) + \mathrm{d_i}(\mathbb{V}, \mathbb{W})$$

for any three persistence modules $\mathbb{U}, \mathbb{V}, \mathbb{W}$.

Proof Given a δ_1-interleaving between $\mathbb{U}, \mathbb{V}$ and a δ_2-interleaving between $\mathbb{V}, \mathbb{W}$ one can construct a $\delta = (\delta_1 + \delta_2)$-interleaving between $\mathbb{U}, \mathbb{W}$ by composing the interleaving maps:

$$\mathbb{U} \xrightarrow{\Phi_1} \mathbb{V} \xrightarrow{\Phi_2} \mathbb{W}$$
$$\mathbb{U} \xleftarrow{\Psi_1} \mathbb{V} \xleftarrow{\Psi_2} \mathbb{W}$$

One easily verifies that $\Phi = \Phi_2\Phi_1$ and $\Psi = \Psi_1\Psi_2$ are interleaving maps. Explicitly:

$$\begin{aligned} \Psi\Phi &= \Psi_1\Psi_2\Phi_2\Phi_1 = \Psi_1 1_{\mathbb{V}}^{2\delta_2}\Phi_1 = \Psi_1\Phi_1 1_{\mathbb{U}}^{2\delta_2} = 1_{\mathbb{U}}^{2\delta_1} 1_{\mathbb{U}}^{2\delta_2} = 1_{\mathbb{U}}^{2\delta} \\ \Phi\Psi &= \Phi_2\Phi_1\Psi_1\Psi_2 = \Phi_2 1_{\mathbb{V}}^{2\delta_1}\Psi_2 = \Phi_2\Psi_2 1_{\mathbb{W}}^{2\delta_1} = 1_{\mathbb{W}}^{2\delta_2} 1_{\mathbb{W}}^{2\delta_2} = 1_{\mathbb{W}}^{2\delta} \end{aligned}$$

Now take the infimum over δ_1, δ_2. □

The proposition tells us that $\mathrm{d_i}$ is an extended pseudometric. It is not a true metric because $\mathrm{d_i}(\mathbb{U}, \mathbb{V}) = 0$ does not imply $\mathbb{U} \cong \mathbb{V}$, as we saw above. In fact, two q-tame persistence modules have interleaving distance 0 if and only if their undecorated persistence diagrams are the same. This is a consequence of the isometry theorem.

Here is the simplest instance. The straightforward proof is left to the reader (or see Proposition 5.6).

Example 5.4 The four interval modules

$$\mathbf{k}[p,q], \ \mathbf{k}[p,q), \ \mathbf{k}(p,q], \ \mathbf{k}(p,q)$$

are 0^+-interleaved but not isomorphic.

The following property of interleaving distance will be useful later.

Proposition 5.5 *Let* $\mathbb{U}_1, \mathbb{U}_2, \mathbb{V}_1, \mathbb{V}_2$ *be persistence modules. Then*

$$\mathrm{d_i}(\mathbb{U}_1 \oplus \mathbb{U}_2, \mathbb{V}_1 \oplus \mathbb{V}_2) \leq \max\left(\mathrm{d_i}(\mathbb{U}_1, \mathbb{V}_1), \mathrm{d_i}(\mathbb{U}_2, \mathbb{V}_2)\right)$$

More generally, let $(\mathbb{U}_\ell \mid \ell \in L)$ *and* $(\mathbb{V}_\ell \mid \ell \in L)$ *be families of persistence modules indexed by the same set* L*, and let*

$$\mathbb{U} = \bigoplus_{\ell \in L} \mathbb{U}_\ell, \quad \mathbb{V} = \bigoplus_{\ell \in L} \mathbb{V}_\ell.$$

Then

$$\mathrm{d_i}(\mathbb{U}, \mathbb{V}) \leq \sup\left(\mathrm{d_i}(\mathbb{U}_\ell, \mathbb{V}_\ell) \mid \ell \in L\right).$$

Proof Given δ-interleavings Φ_ℓ, Ψ_ℓ for each pair $\mathbb{U}_\ell, \mathbb{V}_\ell$, the direct sum maps $\Phi = \bigoplus \Phi_\ell$, $\Psi = \bigoplus \Psi_\ell$ constitute a δ-interleaving of $\mathbb{U}, \mathbb{V}$. Thus any upper bound on the $\mathrm{d_i}(\mathbb{U}_\ell, \mathbb{V}_\ell)$ is an upper bound for $\mathrm{d_i}(\mathbb{U}, \mathbb{V})$. In particular, this is true for the least upper bound, or sup. □

5.2 The Bottleneck Distance

Now we define the metric on the other side of the isometry theorem, namely the bottleneck distance between undecorated persistence diagrams. For a q-tame persistence module $\mathbb{V}$, every rectangle not touching the diagonal has finite $\mu_{\mathbb{V}}$-measure. This implies that the undecorated diagram

$$\mathsf{dgm}(\mathbb{V}) = \mathsf{dgm}(\mu_{\mathbb{V}})$$

is a multiset in the extended open half-plane

$$\overline{\mathcal{H}}^{\circ} = \{(p, q) \mid -\infty \leq p < q \leq +\infty\}.$$

In order to define the bottleneck distance, we need to specify the distance between any pair of points in $\overline{\mathcal{H}}^{\circ}$, as well as the distance between any point and the diagonal (the boundary of the half-plane). These distance functions are not arbitrary; they are defined as they are because of the interleaving properties of interval modules.

(point to point): The first idea is that two undecorated diagrams are close if there is a bijection between them which doesn't move any point too far. We use the ℓ^{∞}-metric in the plane:

$$\mathrm{d}^{\infty}((p, q), (r, s)) = \max\left(|p - r|, |q - s|\right)$$

Points at infinity are compared in the expected way:

$$\mathrm{d}^{\infty}((-\infty, q), (-\infty, s)) = |q - s|,$$
$$\mathrm{d}^{\infty}((p, +\infty), (r, +\infty)) = |p - r|,$$

and

$$\mathrm{d}^{\infty}((-\infty, +\infty), (-\infty, +\infty)) = 0.$$

Distances between points in different strata (e.g. between (p, q) and $(-\infty, s)$) are infinite.

The next result gives a relationship between the ℓ^{∞}-metric and the interleaving of interval modules.

Proposition 5.6 *Let (p^*, q^*) and (r^*, s^*) be intervals (possibly infinite), and let*

$$\mathbb{U} = \mathbf{k}(p^*, q^*) \quad \textit{and} \quad \mathbb{V} = \mathbf{k}(r^*, s^*)$$

be the corresponding interval modules. Then

$$\mathrm{d_i}(\mathbb{U}, \mathbb{V}) \leq \mathrm{d}^{\infty}((p, q), (r, s)).$$

The proof is postponed to the end of the section. We remark that equality holds provided that the intervals overlap sufficiently (the closure of each interval must meet the midpoint of the other), so the proposition is tight in that sense.

(point to diagonal): The next idea is that points which are close to the diagonal are allowed to be swallowed up by the diagonal. Again we use the ℓ^{∞}-metric:

$$\mathrm{d}^{\infty}((p, q), \Delta) = \tfrac{1}{2}(q - p)$$

Again this is related to the behaviour of interval modules.

Proposition 5.7 *Let* (p^*, q^*) *be an interval, let*

$$\mathbb{U} = \mathbf{k}(p^*, q^*),$$

be the corresponding interval module, and let 0 *denote the zero persistence module. Then*

$$\mathrm{d_i}(\mathbb{U}, 0) = \tfrac{1}{2}(q - p).$$

(This is infinite if the interval is infinite.)

Proof Let $\delta \geq 0$. When is there a δ-interleaving? The interleaving maps must be zero (no other maps exist to or from the module 0), so the only condition that needs checking is $\Psi\Phi = 1_{\mathbb{U}}^{2\delta}$, which is really $0 = 1_{\mathbb{U}}^{2\delta}$. This holds when $\delta > \frac{1}{2}(q - p)$ and fails when $\delta < \frac{1}{2}(q - p)$. □

We now use these two concepts to define the bottleneck distance between two multisets A, B in the extended half-plane.

It is easier to work with sets rather than multisets. One way to do this is to attach labels to distinguish multiple instances of each repeated point. For instance, α with multiplicity k becomes $\alpha_1, \ldots, \alpha_k$. Henceforth we will do this implicitly, without comment.

A **partial matching** between A and B is a collection of pairs

$$\mathsf{M} \subset \mathsf{A} \times \mathsf{B}$$

such that:

- for every $\alpha \in \mathsf{A}$ there is at most one $\beta \in \mathsf{B}$ such that $(\alpha, \beta) \in \mathsf{M}$;
- for every $\beta \in \mathsf{B}$ there is at most one $\alpha \in \mathsf{A}$ such that $(\alpha, \beta) \in \mathsf{M}$.

We say that a partial matching M is a δ **-matching** if all of the following are true:

- if $(\alpha, \beta) \in \mathsf{M}$ then $\mathrm{d}^\infty(\alpha, \beta) \leq \delta$;
- if $\alpha \in \mathsf{A}$ is unmatched then $\mathrm{d}^\infty(\alpha, \Delta) \leq \delta$;
- if $\beta \in \mathsf{B}$ is unmatched then $\mathrm{d}^\infty(\beta, \Delta) \leq \delta$.

The **bottleneck distance** between two multisets A, B in the extended half-plane is

$$\mathrm{d_b}(\mathsf{A}, \mathsf{B}) = \inf\left(\delta \mid \text{ there exists a } \delta\text{-matching between } \mathsf{A} \text{ and } \mathsf{B}\right).$$

In Sect. 5.3, we will show that 'inf' can be replaced by 'min' if A, B are locally finite.

Remark 5.8 In order for $\mathrm{d_b}(\mathsf{A}, \mathsf{B}) < \infty$, it is necessary that the cardinalities of A, B agree over each of the three strata at infinity:

$$\begin{aligned} \mathrm{card}(\mathsf{A}|_{\{-\infty\}\times\mathbf{R}}) &= \mathrm{card}(\mathsf{B}|_{\{-\infty\}\times\mathbf{R}}) \\ \mathrm{card}(\mathsf{A}|_{\mathbf{R}\times\{+\infty\}}) &= \mathrm{card}(\mathsf{B}|_{\mathbf{R}\times\{+\infty\}}) \\ \mathrm{card}(\mathsf{A}|_{\{-\infty\}\times\{+\infty\}}) &= \mathrm{card}(\mathsf{B}|_{\{-\infty\}\times\{+\infty\}}) \end{aligned}$$

Indeed, these points have infinite distance from the diagonal and from points in the other strata, and therefore they must be bijectively matched within each stratum.

Proposition 5.9 *The bottleneck distance satisfies the triangle inequality:*

$$\mathrm{d_b}(\mathsf{A}, \mathsf{C}) \leq \mathrm{d_b}(\mathsf{A}, \mathsf{B}) + \mathrm{d_b}(\mathsf{B}, \mathsf{C})$$

for any three multisets $\mathsf{A}, \mathsf{B}, \mathsf{C}$.

Proof Suppose M_1 is a δ_1-matching between A, B, and M_2 is a δ_2-matching between B, C. Let $\delta = \delta_1 + \delta_2$. We must show that there is a δ-matching between A, C.

Define the **composition** of $\mathsf{M}_1, \mathsf{M}_2$ to be

$$\mathsf{M} = ((\alpha, \gamma) \mid \text{there exists } \beta \in \mathsf{B} \text{ such that } (\alpha, \beta) \in \mathsf{M}_1 \text{ and } (\beta, \gamma) \in \mathsf{M}_2)\,.$$

It is a partial matching because $\mathsf{M}_1, \mathsf{M}_2$ are partial matchings. We verify that M is the required δ-matching:

- If $(\alpha, \gamma) \in \mathsf{M}$ then

  $$\mathrm{d}^\infty(\alpha, \gamma) \leq \mathrm{d}^\infty(\alpha, \beta) + \mathrm{d}^\infty(\beta, \gamma) \leq \delta_1 + \delta_2 = \delta$$

  where $\beta \in \mathsf{B}$ is the point linking α to γ.
- If α is unmatched in M then there are two possibilities. Either α is unmatched in M_1, in which case

  $$\mathrm{d}^\infty(\alpha, \Delta) \leq \delta_1 \leq \delta.$$

  Or α is matched in M_1, let's say $(\alpha, \beta) \in \mathsf{M}_1$. Then β must be unmatched in M_2, so

  $$\mathrm{d}^\infty(\alpha, \Delta) \leq \mathrm{d}^\infty(\alpha, \beta) + \mathrm{d}^\infty(\beta, \Delta) \leq \delta_1 + \delta_2 = \delta.$$

- If γ is unmatched in M, then a similar argument shows that

  $$\mathrm{d}^\infty(\gamma, \Delta) \leq \delta.$$

This completes the proof. □

Remark 5.10 Because $\mathsf{A}, \mathsf{B}, \mathsf{C}$ are in truth multisets rather than sets, the composition operation between matchings is not uniquely defined, but depends on how the matchings are realised when labels are added. Figure 5.1 illustrates what can happen when B has points of multiplicity greater than 1. Since we are concerned only with the existence of the composite matching, this ambiguity does not trouble us.

Here is the first substantial-looking result comparing the interleaving and bottleneck distances.

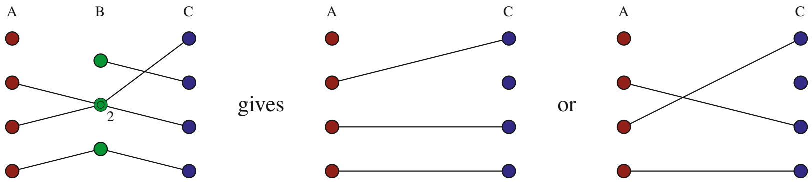


Fig. 5.1 The partial matchings between A, B and B, C (*left*) compose to give a partial matching between A, C in two different ways (*middle, right*). The non-uniqueness arises from the point of multiplicity 2 in B

Theorem 5.11 *Let* $\mathbb{U}, \mathbb{V}$ *be decomposable persistence modules (i.e. direct sums of interval modules). Then*

$$\mathrm{d_i}(\mathbb{U}, \mathbb{V}) \leq \mathrm{d_b}(\mathsf{dgm}(\mathbb{U}), \mathsf{dgm}(\mathbb{V})).$$

(We remind the reader that tameness is not required to define Dgm and dgm in this case: see Sect. 2.6.)

Proof We show that whenever there exists a δ-matching between $\mathsf{dgm}(\mathbb{U})$ and $\mathsf{dgm}(\mathbb{V})$, we have $\mathrm{d_i}(\mathbb{U}, \mathbb{V}) \leq \delta$. The result follows by taking the infimum over all such δ.

Let M be a δ-matching between the two diagrams. Since the points in each diagram correspond to the interval summands of the module, we can construct from M a partial matching between the interval summands of $\mathbb{U}$ and $\mathbb{V}$.

Rewrite $\mathbb{U}$ and $\mathbb{V}$ in the form

$$\mathbb{U} = \bigoplus_{\ell \in L} \mathbb{U}_\ell, \quad \mathbb{V} = \bigoplus_{\ell \in L} \mathbb{V}_\ell$$

so that each pair $(\mathbb{U}_\ell, \mathbb{V}_\ell)$ is one of the following:

- a pair of matched intervals;
- $\mathbb{U}_\ell$ is an unmatched interval, $\mathbb{V}_\ell = 0$;
- $\mathbb{V}_\ell$ is an unmatched interval, $\mathbb{U}_\ell = 0$.

In each case, by Propositions 5.6 and 5.7, we have $\mathrm{d_i}(\mathbb{U}_\ell, \mathbb{V}_\ell) \leq \delta$. It follows from Proposition 5.5 that $\mathrm{d_i}(\mathbb{U}, \mathbb{V}) \leq \delta$. □

We complete this section with the postponed proof.

Proof (*Proposition* 5.6) We treat the case where p, q, r, s are all finite. We must show that if

$$\delta > \max(|p - r|, |q - s|)$$

then $\mathbb{U}$, $\mathbb{V}$ are δ-interleaved. We define systems of linear maps

$$\Phi = (\phi_t : U_t \to V_{t+\delta})$$
$$\Psi = (\psi_t : V_t \to U_{t+\delta})$$

and then show that the interleaving relations

$$\Phi 1_{\mathbb{U}}^{\eta} = 1_{\mathbb{V}}^{\eta} \Phi, \quad \Psi 1_{\mathbb{V}}^{\eta} = 1_{\mathbb{U}}^{\eta} \Psi, \quad \Psi \Phi = 1_{\mathbb{U}}^{2\delta}, \quad \Phi \Psi = 1_{\mathbb{V}}^{2\delta}$$

hold.

The definition of the maps ϕ_t, ψ_t is straightforward. Each vector space in $\mathbb{U}$, $\mathbb{V}$ is equal to zero or to the field $\mathbf{k}$. If the domain and codomain equal $\mathbf{k}$, then the map is defined to be the identity $1 = 1_{\mathbf{k}}$. Otherwise, the map is necessarily 0.

The first step is to show that the systems of maps $\Phi = (\phi_t)$, $\Psi = (\psi_t)$ are module homomorphisms. For Φ this entails verifying that the diagram

$$\begin{array}{ccccc} U_t & \longrightarrow & U_{t+\eta} & & \\ & \searrow & & \searrow & \\ & & V_{t+\delta} & \longrightarrow & V_{t+\eta+\delta} \end{array}$$

commutes for all t and for all $\eta > 0$. Because of the special form of the vector spaces and maps, it is enough to show that the situation is not one of the following:

$$\begin{array}{ccccc} \bullet & \longrightarrow & \bullet & & \\ & \searrow & & \searrow & \\ & & \circ & \longrightarrow & \bullet \end{array} \qquad \text{or} \qquad \begin{array}{ccccc} \bullet & \longrightarrow & \circ & & \\ & \searrow & & \searrow & \\ & & \bullet & \longrightarrow & \bullet \end{array}$$

Here a filled circle $\bullet$ indicates that the vector space is $\mathbf{k}$, and an open circle $\circ$ denotes that it is zero. For the first situation to occur, one must have

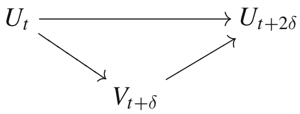
$$p \leq t \quad \text{and} \quad t + \delta \leq r$$

which would imply $\delta \leq r - t \leq r - p$. This contradicts the hypothesis $\delta > r - p$. For the second situation to occur, one must have

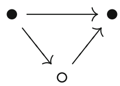
$$q \leq t + \eta \quad \text{and} \quad t + \eta + \delta \leq s$$

which would imply $\delta \leq s - t - \eta \leq s - q$. This contradicts the hypothesis $\delta > s - q$. It follows that Φ is a module homomorphism. By symmetry, so is Ψ.

The second step is to show that $\Psi\Phi = 1_{\mathbb{U}}^{2\delta}$ and $\Phi\Psi = 1_{\mathbb{V}}^{2\delta}$. For the first of these, we must verify that the diagram



commutes for all t. This time the unique forbidden configuration is



and if this occurs then the top row implies

$$p \leq t \quad \text{and} \quad t + 2\delta \leq q.$$

Since $\delta > r - p$ and $\delta > q - s$ we infer that

$$r < p + \delta \leq t + \delta \leq q - \delta < s$$

which implies that the circle on the bottom row is filled after all. It follows that $\Psi\Phi = 1_{\mathbb{U}}^{2\delta}$. By symmetry, $\Phi\Psi = 1_{\mathbb{V}}^{2\delta}$.

This finishes the proof when p, q, r, s are finite. The infinite cases are similar. □

5.3 The Bottleneck Distance (Continued)

If A, B are locally finite, it turns out that the 'inf' is attained in the definition

$$\mathrm{d_b}(\mathsf{A}, \mathsf{B}) = \inf\left(\delta \mid \text{ there exists a } \delta \text{ -matching between } \mathsf{A} \text{ and } \mathsf{B}\right),$$

and can be replaced by 'min'. This will allow us to make a tighter statement of the stability theorem (5.14′) for q-tame modules. See Theorem 5.23.

Theorem 5.12 *Let* A, B *be locally finite multisets in the extended open half-plane* $\overline{\mathcal{H}}^{\circ}$. *Suppose for every* $\eta > \delta$ *there exists an* η*-matching between* A, B. *Then there exists a* δ*-matching between* A, B.

The assertion is obvious if A, B are finite. The general case is proved using a compactness argument, as follows.

Proof As usual we treat A, B as sets rather than multisets.

For every integer $n \geq 1$, let M_n be a $(\delta + \frac{1}{n})$-matching between A, B. The plan is to construct a δ-matching M from the sequence (M_n). In practice, we work with the indicator functions

$$\chi : \mathsf{A} \times \mathsf{B} \to \{0, 1\}$$
$$\chi_n : \mathsf{A} \times \mathsf{B} \to \{0, 1\}$$

of the partial matchings M, M_n.

The first step is to construct χ as a limit of the sequence (χ_n). Take a fixed enumeration

$$((\alpha_\ell, \beta_\ell) \mid \ell \geq 1)$$

of the countable set $\mathsf{A} \times \mathsf{B}$. We will inductively construct a descending sequence

$$\mathbf{N} = \mathbf{N}_0 \supseteq \mathbf{N}_1 \supseteq \cdots \supseteq \mathbf{N}_\ell \supseteq \cdots$$

of infinite subsets of the natural numbers, with the property that $\chi_n(\alpha_\ell, \beta_\ell)$ takes the same value for all $n \in \mathbf{N}_\ell$. Having done so, we define $\chi(\alpha_\ell, \beta_\ell)$ to be this common value.

The construction of $\mathbf{N}_\ell$ is straightforward: once $\mathbf{N}_{\ell-1}$ is defined, at least one of the two sets

$$\{n \in \mathbf{N}_{\ell-1} \mid \chi_n(\alpha_\ell, \beta_\ell) = 0\} \quad \text{and} \quad \{n \in \mathbf{N}_{\ell-1} \mid \chi_n(\alpha_\ell, \beta_\ell) = 1\}$$

has infinite cardinality, and that will be our $\mathbf{N}_\ell$. (If both, then either will do.) Repeat.

Lemma 5.13 *If* F *is any finite subset of* $\mathsf{A} \times \mathsf{B}$, *then there exists* $\ell \geq 1$ *such that*

$$\chi(\alpha, \beta) = \chi_n(\alpha, \beta)$$

for all $(\alpha, \beta) \in \mathsf{F}$ *and for all* $n \in \mathbf{N}_\ell$.

Proof Indeed, select ℓ such that $(\alpha_1, \beta_1), \ldots, (\alpha_\ell, \beta_\ell)$ include all of F. □

The second step is to verify that χ is the indicator function of a δ-matching. There are several items to check.

- For $\alpha \in \mathsf{A}$ there is at most one $\beta \in \mathsf{B}$ such that $\chi(\alpha, \beta) = 1$.
  Proof Suppose $\chi(\alpha, \beta) = \chi(\alpha, \beta') = 1$ for distinct elements $\beta, \beta' \in \mathsf{B}$. By the lemma, there exists n such that $\chi_n(\alpha, \beta) = \chi_n(\alpha, \beta') = 1$, which contradicts the fact that M_n is a partial matching. □
- For $\alpha \in \mathsf{A}$ with $\mathrm{d}^\infty(\alpha, \Delta) > \delta$, there is at least one $\beta \in \mathsf{B}$ such that $\chi(\alpha, \beta) = 1$.

  Proof Select N such that $\mathrm{d}^\infty(\alpha, \Delta) > \delta + \frac{1}{N}$. Then the set

$$\mathsf{F}_\alpha = \left\{\beta \in \mathsf{B} \mid \mathrm{d}^\infty(\alpha, \beta) \leq \delta + \tfrac{1}{N}\right\}$$

is finite, since B is locally finite and these points lie in a square bounded away from the diagonal. By Lemma 5.13, there exists ℓ such that

$$\chi(\alpha, \beta) = \chi_n(\alpha, \beta)$$

for all $\beta \in \mathsf{F}_\alpha$ and for all $n \in \mathbf{N}_\ell$. On the other hand, if $n \geq N$, then M_n matches α with some $\beta \in \mathsf{F}_\alpha$. Combining these observations,

$$\chi(\alpha, \beta) = \chi_n(\alpha, \beta) = 1$$

for sufficiently large $n \in \mathbf{N}_\ell$ and for some $\beta \in \mathsf{F}_\alpha$. □

By symmetry we have:

- For $\beta \in \mathsf{B}$ there is at most one $\alpha \in \mathsf{A}$ such that $\chi(\alpha, \beta) = 1$.
- For $\beta \in \mathsf{B}$ with $\mathrm{d}^\infty(\beta, \Delta) > \delta$, there is at least one $\alpha \in \mathsf{A}$ such that $\chi(\alpha, \beta) = 1$.

Finally:

- If $\chi(\alpha, \beta) = 1$ then $\mathrm{d}^\infty(\alpha, \beta) \leq \delta$.

Proof By Lemma 5.13, there are infinitely many n for which $\chi_n(\alpha, \beta) = 1$. Then

$$\mathrm{d}^\infty(\alpha, \beta) \leq \delta + \tfrac{1}{n}$$

for these n. Since n may be arbitrarily large, the result follows. □

These five bullet points confirm that M, defined by its indicator function χ, is a δ-matching between A, B. □

Although we have chosen to spell out a direct argument, in fact Theorem 5.12 is an instance of the compactness theorem in first-order logic. The set of constraints that must be satisfied by an η-matching can be formulated as a theory T_η on a collection of binary-valued variables $x_{\alpha\beta}$. An η-matching is precisely a model for that theory. The theory T_δ is seen to be logically equivalent to the union of the theories $(T_\eta \mid \eta > \delta)$. If each T_η has a model, then any finite subtheory of this union has a model, therefore by compactness T_δ has a model. The details are left to the interested reader.

5.4 The Isometry Theorem

Having defined the interleaving distance and the bottleneck distance, we can now state the main theorem.

Theorem 5.14 *Let $\mathbb{U}, \mathbb{V}$ be q-tame persistence modules. Then*

$$\mathrm{d_i}(\mathbb{U}, \mathbb{V}) = \mathrm{d_b}(\mathsf{dgm}(\mathbb{U}), \mathsf{dgm}(\mathbb{V}))$$

(Recall that dgm *denotes the undecorated persistence diagram).*

The result naturally falls into two parts: the ‘stability theorem’ [19], [14]

$$\mathrm{d_i}(\mathbb{U}, \mathbb{V}) \geq \mathrm{d_b}(\mathsf{dgm}(\mathbb{U}), \mathsf{dgm}(\mathbb{V})), \tag{5.14$'$}$$

and the ‘converse stability theorem’ [42]

$$\mathrm{d_i}(\mathbb{U}, \mathbb{V}) \leq \mathrm{d_b}(\mathsf{dgm}(\mathbb{U}), \mathsf{dgm}(\mathbb{V})). \tag{5.14$''$}$$

The proof of the converse stability theorem (5.14″) occupies Sect. 5.5. We have already seen the result for decomposable modules, in Theorem 5.11, so it is a matter of extending the result to q-tame modules that are not known to be decomposable. The proof of the stability theorem (5.14′) is given in Sects. 5.6 and 5.7.

5.5 The Converse Stability Theorem

In this section we deduce the converse stability inequality (5.14″) for q-tame modules from Theorem 5.11. A similar argument was given by Lesnick [42], to whom the theorem is due. The key idea is that persistence modules can be approximated by better-behaved persistence modules, using a procedure we call ‘smoothing’.

Definition 5.15 Let $\mathbb{V}$ be a persistence module, and let $\varepsilon > 0$. The ε **-smoothing** of $\mathbb{V}$ is the persistence module $\mathbb{V}^\varepsilon$ defined to be the image of the map

$$1_{\mathbb{V}}^{2\varepsilon} : \mathbb{V}[-\varepsilon] \to \mathbb{V}[\varepsilon]$$

(using the ‘shift’ notation from Remark 4.1). Thus $(V^\varepsilon)_t$ is the image of the map

$$v_{t+\varepsilon}^{t-\varepsilon} : V_{t-\varepsilon} \to V_{t+\varepsilon},$$

and $(v^\varepsilon)_t^s$ is the restriction of $v_{t+\varepsilon}^{s+\varepsilon}$.

Then we have a factorisation of $1_{\mathbb{V}}^{2\varepsilon}$

$$\mathbb{V}[-\varepsilon] \longrightarrow \mathbb{V}^\varepsilon \longrightarrow \mathbb{V}[\varepsilon] \tag{5.1}$$

where the first map is surjective and the second map is injective. At a given index t this is the sequence:

$$V_{t-\varepsilon} \xrightarrow{v_{t-\varepsilon}^{t+\varepsilon}} V_t^\varepsilon \xrightarrow{1} V_{t+\varepsilon}$$

Proposition 5.16 *Let* $\mathbb{V}$ *be a persistence module. Then* $\mathrm{d_i}(\mathbb{V}, \mathbb{V}^\varepsilon) \leq \varepsilon$.

Proof One checks that the maps in (5.1) constitute an ε-interleaving. □

Smoothing changes the persistence diagram in a predictable way. Here is the atomic example (which the reader can easily verify):

Example 5.17 Let $\mathbb{V} = \mathbf{k}(p^*, q^*)$. Then:

$$\mathbb{V}^\varepsilon = \begin{cases} \mathbf{k}((p+\varepsilon)^*, (q-\varepsilon)^*) & \text{if } (p+\varepsilon)^* < (q-\varepsilon)^* \\ 0 & \text{otherwise} \end{cases}$$

In other words, ε-smoothing shrinks the interval by ε at both ends. □

Proposition 5.18 *The persistence diagram of* $\mathbb{V}^\varepsilon$ *is obtained from the persistence diagram of* $\mathbb{V}$ *by applying the translation* $T_\varepsilon : (p, q) \mapsto (p + \varepsilon, q - \varepsilon)$ *to the part of the extended half-plane that lies above the line* $\Delta_\varepsilon = \{(t - \varepsilon, t + \varepsilon) \mid t \in \mathbf{R}\}$.

In other words, the entire diagram is pushed towards the diagonal by the translation vector $(\varepsilon, -\varepsilon)$. Information in $\mathsf{Dgm}(\mathbb{V})$ that lies below the line Δ_ε is lost and cannot be retrieved from $\mathsf{Dgm}(\mathbb{V}^\varepsilon)$.

In the case where Dgm is not everywhere defined, the proposition is understood to include the assertion that the finite r-interior of the persistence measure, and hence the region where Dgm is defined, is shifted by T_ε.

Proof We consider three different cases. Case (ii) is subsumed by case (iii), but we include it because it makes the proof easier to digest.

(i) $\mathbb{V}$ **is decomposable.** The image of a direct sum of maps is the direct sum of the images of the maps; therefore ε-smoothing commutes with direct sums:

$$\left[\bigoplus_{\ell \in L} \mathbb{V}_\ell\right]^\varepsilon = \bigoplus_{\ell \in L} \mathbb{V}_\ell^\varepsilon$$

By Example 5.17, the proposition is true for interval modules. It is therefore true for direct sums of interval modules.

(ii) $\mathbb{V}$ **is q-tame.** It is enough to show that the rank function of $\mathbb{V}^\varepsilon$ is equal to the rank function of $\mathbb{V}$ shifted by T_ε, since this determines the persistence measure and hence the persistence diagram. Specifically, for all $b < c$ we require:

$$\text{rank}[(V^\varepsilon)_b \to (V^\varepsilon)_c] = \text{rank}[V_{b-\varepsilon} \to V_{c+\varepsilon}]$$

In fact, these maps are related by the sequence

$$V_{b-\varepsilon} \longrightarrow (V^\varepsilon)_b \longrightarrow (V^\varepsilon)_c \longrightarrow V_{c+\varepsilon}$$

where the first map is surjective and the third map is injective. Since the rank of a linear map is unchanged by pre-composing with a surjective map, or post-composing with an injective map, it follows that the rank of the middle map is equal to the rank of the composite. This is what we wished to prove.

(iii) **general case.** We show that the persistence measure of $\mathbb{V}^\varepsilon$ is equal to the persistence measure of $\mathbb{V}$ shifted by T_ε. Writing

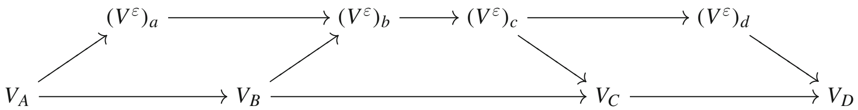
$$A = a - \varepsilon, \quad B = b - \varepsilon, \quad C = c + \varepsilon, \quad D = d + \varepsilon$$

this means showing that

$$\langle \circ_A \text{—} \bullet_B \text{—} \bullet_C \text{—} \circ_D \mid \mathbb{V} \rangle = \langle \circ_a \text{—} \bullet_b \text{—} \bullet_c \text{—} \circ_d \mid \mathbb{V}^\varepsilon \rangle$$

for all $a < b \leq c < d$.

The proof is based on the following commutative diagram



in which the maps $\nearrow$ are surjective and the maps $\searrow$ are injective. The diagram can be thought of as a persistence module over an 8-element poset, the partial order being implied by the arrows. We will carry out quiver calculations by restricting to various totally-ordered subsets of this poset.

To begin with, surjectivity of the maps $\nearrow$ means that

$$\langle \circ_A \text{—} \bullet_a \rangle = 0 \quad \text{and} \quad \langle \circ_B \text{—} \bullet_b \rangle = 0,$$

and injectivity of the maps $\searrow$ means that

$$\langle \bullet_c \text{—} \circ_C \rangle = 0 \quad \text{and} \quad \langle \bullet_d \text{—} \circ_D \rangle = 0.$$

Moreover, by the restriction principle, interval types containing any of these 'forbidden' configurations occur with multiplicity zero.

Then

$$\begin{aligned}
\langle \circ_A \text{———} \bullet_b \text{—} \bullet_c \text{———} \circ_D \rangle &= \langle \circ_A \text{—} \bullet_B \text{—} \bullet_b \text{—} \bullet_c \text{—} \bullet_C \text{—} \circ_D \rangle \\
&\quad + \text{three other terms} \\
&= \langle \circ_A \text{—} \bullet_B \text{—} \bullet_b \text{—} \bullet_c \text{—} \bullet_C \text{—} \circ_D \rangle \\
&= \langle \circ_A \text{—} \bullet_B \text{————} \bullet_C \text{—} \circ_D \rangle \\
&= \langle \circ_A \text{—} \bullet_B \text{—} \bullet_C \text{—} \circ_D \mid \mathbb{V} \rangle
\end{aligned}$$

and at the same time

$$
\begin{aligned}
\langle \circ_A \text{———} \bullet_b \text{—} \bullet_c \text{———} \circ_D \rangle &= \langle \circ_A \text{—} \circ_a \text{—} \bullet_b \text{—} \bullet_c \text{—} \circ_d \text{—} \circ_D \rangle \\
&\quad + \text{three other terms} \\
&= \langle \circ_A \text{—} \circ_a \text{—} \bullet_b \text{—} \bullet_c \text{—} \circ_d \text{—} \circ_D \rangle \\
&= \langle \text{——} \circ_a \text{—} \bullet_b \text{—} \bullet_c \text{—} \circ_d \text{——} \rangle \\
&= \langle \circ_a \text{—} \bullet_b \text{—} \bullet_c \text{—} \circ_d \mid \mathbb{V}^\varepsilon \rangle
\end{aligned}
$$

so we get the required equality. The six 'other terms' are all zero because they contain forbidden configurations. □

Corollary 5.19 *Let* $\mathbb{V}$ *be q-tame. Then* $\mathrm{d_b}(\mathsf{dgm}(\mathbb{V}), \mathsf{dgm}(\mathbb{V}^\varepsilon)) \leq \varepsilon$.

Proof Indeed, an ε-matching is defined as follows:

$$(p, q) \in \mathsf{dgm}(\mathbb{V}^\varepsilon) \quad \leftrightarrow \quad (p - \varepsilon, q + \varepsilon) \in \mathsf{dgm}(\mathbb{V})$$

This is bijective except for the unmatched points of $\mathsf{dgm}(\mathbb{V})$, which lie on or below the line Δ_ε, and therefore have distance at most ε from the diagonal. □

Proposition 5.20 *If* $\mathbb{V}$ *is q-tame then* $\mathbb{V}^\varepsilon$ *is locally finite.*

Proof Since $\dim((V^\varepsilon)_t) = \mathrm{rank}[V_{t-\varepsilon} \to V_{t+\varepsilon}] < \infty$, it follows from Theorem 2.8 (2) that $\mathbb{V}^\varepsilon$ is decomposable into interval modules. We claim that the collection of intervals is locally finite. Specifically, for any $t \in \mathbf{R}$, we estimate

$$
\begin{aligned}
&\#\{\text{intervals which meet } [t - \tfrac{1}{2}\varepsilon, t + \tfrac{1}{2}\varepsilon]\} \\
&\quad = \#\{\text{points of } \mathsf{Dgm}(\mathbb{V}^\varepsilon) \text{ in the upper-left quadrant at } (t + \tfrac{1}{2}\varepsilon, t - \tfrac{1}{2}\varepsilon)\} \\
&\quad \leq \#\{\text{points of } \mathsf{Dgm}(\mathbb{V}) \text{ in the upper-left quadrant at } (t - \tfrac{1}{2}\varepsilon, t + \tfrac{1}{2}\varepsilon)\} \\
&\quad = \mathrm{rank}[V_{t-\frac{1}{2}\varepsilon} \to V_{t+\frac{1}{2}\varepsilon}]
\end{aligned}
$$

which is finite. The '$\leq$' in the third line is a consequence of Proposition 5.18. □

We are now ready to prove the converse stability theorem for q-tame persistence modules, using the triangle inequalities for $\mathrm{d_i}$, $\mathrm{d_b}$ and our results on ε-smoothing.

Proof (5.14″) Let $\mathbb{U}, \mathbb{V}$ be q-tame persistence modules. For any $\varepsilon > 0$, the ε-smoothings $\mathbb{U}^\varepsilon, \mathbb{V}^\varepsilon$ are decomposable, so the converse stability theorem applies to them. Then:

$$
\begin{aligned}
\mathrm{d_i}(\mathbb{U}, \mathbb{V}) &\leq \mathrm{d_i}(\mathbb{U}^\varepsilon, \mathbb{V}^\varepsilon) + 2\varepsilon && \text{by Proposition 5.16} \\
&\leq \mathrm{d_b}(\mathsf{dgm}(\mathbb{U}^\varepsilon), \mathsf{dgm}(\mathbb{V}^\varepsilon)) + 2\varepsilon && \text{by Theorem 5.11} \\
&\leq \mathrm{d_b}(\mathsf{dgm}(\mathbb{U}), \mathsf{dgm}(\mathbb{V})) + 4\varepsilon && \text{by Corollary 5.19}
\end{aligned}
$$

Since this is true for all $\varepsilon > 0$, we deduce that

$$\mathrm{d_i}(\mathbb{U}, \mathbb{V}) \le \mathrm{d_b}(\mathsf{dgm}(\mathbb{U}), \mathsf{dgm}(\mathbb{V})).$$

The converse stability theorem for q-tame modules is proved. □

We finish this section with a characterisation of q-tame modules.

Theorem 5.21 *A persistence module $\mathbb{V}$ is q-tame if and only if it can be approximated, in the interleaving distance, by locally finite modules.*

Proof If $\mathbb{V}$ is q-tame then it is approximated by the modules $\mathbb{V}^\varepsilon$, which are locally finite by Proposition 5.20. Conversely, suppose $\mathbb{V}$ is approximated by locally finite modules. Suppose $b < c$ is given. Let $\mathbb{W}$ be a locally finite module which is ε-interleaved with $\mathbb{V}$, for some $\varepsilon < (c - b)/2$. Then

$$\mathrm{r}_b^c = \mathrm{rank}[V_b \to V_c] = \mathrm{rank}[V_b \to W_{b+\varepsilon} \to W_{c-\varepsilon} \to V_c] \le \dim(W_{b+\varepsilon})$$

which is finite. It follows that $\mathbb{V}$ is q-tame. □

Example 5.22 It is easy to see that there are q-tame modules which are not locally finite. For instance:

$$\bigoplus_{n=1}^{\infty} \mathbf{k}[0, \tfrac{1}{n}] \quad \text{and} \quad \prod_{n=1}^{\infty} \mathbf{k}[0, \tfrac{1}{n}]$$

The latter is the example of Crawley–Boevey [24] with no interval decomposition discussed in Remark 2.9. Incidentally, one can verify directly that the two modules are 0-interleaved; and also that their persistence measures, and hence their persistence diagrams, are equal away from the unique singular point $(0^-, 0^+)$.

5.6 The Stability Theorem

The inequality (5.14′) can be expressed in the following form:

Theorem 5.23 *Let $\mathbb{U}, \mathbb{V}$ be q-tame persistence modules which are δ^+-interleaved. Then there exists a δ-matching between the multisets* $\mathsf{dgm}(\mathbb{U})$, $\mathsf{dgm}(\mathbb{V})$.

It is easier to prove the following. (Notice the missing $^+$.)

Theorem 5.24 *Let $\mathbb{U}, \mathbb{V}$ be q-tame persistence modules which are δ-interleaved. Then there exists a δ-matching between the multisets* $\mathsf{dgm}(\mathbb{U})$, $\mathsf{dgm}(\mathbb{V})$.

Theorem 5.12 allows us to deduce Theorem 5.23 from Theorem 5.24: if $\mathbb{U}, \mathbb{V}$ are δ^+-interleaved then there is an η-matching between their diagrams for every $\eta > \delta$, hence there is a δ-matching. The proof of Theorem 5.24 depends on two main ingredients:

- The interpolation Lemma 4.6, which embeds $\mathbb{U}$, $\mathbb{V}$ within a 1-parameter family.
- The box inequalities 5.26, which relate the persistence measures of $\mathbb{U}$, $\mathbb{V}$ locally.

Once these ingredients are in place, the theorem can be proved using the continuity method of [19]. Our persistence diagrams may have infinite cardinality, so we will need an additional compactness argument to finish off the proof.

Definition 5.25 Let $R = [a, b] \times [c, d]$ be a rectangle in $\overline{\mathbf{R}}^2$. The δ**-thickening** of R is the rectangle

$$R^\delta = [a - \delta, b + \delta] \times [c - \delta, d + \delta].$$

For convenience we will write

$$A = a - \delta, \quad B = b + \delta, \quad C = c - \delta, \quad D = d + \delta$$

in this situation.

For infinite rectangles, we use $-\infty - \delta = -\infty$ and $+\infty + \delta = +\infty$. We can also thicken an individual point: if $\alpha = (p, q)$ then

$$\alpha^\delta = [p - \delta, p + \delta] \times [q - \delta, q + \delta]$$

for $\delta > 0$.

Lemma 5.26 (Box inequalities [19]) *Let* $\mathbb{U}$, $\mathbb{V}$ *be a* δ*-interleaved pair of persistence modules. Let* R *be a rectangle whose* δ*-thickening* R^δ *lies above the diagonal. Then* $\mu_{\mathbb{U}}(R) \leq \mu_{\mathbb{V}}(R^\delta)$ *and* $\mu_{\mathbb{V}}(R) \leq \mu_{\mathbb{U}}(R^\delta)$.

If we use the region extension convention (Remark 3.22) we can state the lemma without the requirement that R^δ lies above the diagonal, since the convention gives $\mu(R^\delta) = \infty$ if it doesn't.

Proof Write $R = [a, b] \times [c, d]$ and $R^\delta = [A, B] \times [C, D]$ as above. Thanks to the interleaving, the finite modules

$$\mathbb{U}_{a,b,c,d} \ : \ U_a \to U_b \to U_c \to U_d$$

and

$$\mathbb{V}_{A,B,C,D} \ : \ V_A \to V_B \to V_C \to V_D$$

are restrictions of the following 8-term module

$$\mathbb{W} \ : \ V_A \xrightarrow{\Psi} U_a \longrightarrow U_b \xrightarrow{\Phi} V_B \longrightarrow V_C \xrightarrow{\Psi} U_c \longrightarrow U_d \xrightarrow{\Phi} V_D$$

where Φ, Ψ are the interleaving maps.

Using the restriction principle, we calculate:

$$\begin{aligned}
\mu_{\mathbb{V}}([A,B]\times[C,D]) &= \langle\, \circ_A \text{————————} \bullet_B \text{—} \bullet_C \text{————————} \circ_D \mid \mathbb{V}\,\rangle \\
&= \langle\, \circ_A \text{————————} \bullet_B \text{—} \bullet_C \text{————————} \circ_D \mid \mathbb{W}\,\rangle \\
&= \langle\, \circ_A \text{—} \circ_a \text{—} \bullet_b \text{—} \bullet_B \text{—} \bullet_C \text{—} \bullet_c \text{—} \circ_d \text{—} \circ_D \mid \mathbb{W}\,\rangle \\
&\qquad + \text{eight other terms} \\
&\geq \langle\, \circ_A \text{—} \circ_a \text{—} \bullet_b \text{—} \bullet_B \text{—} \bullet_C \text{—} \bullet_c \text{—} \circ_d \text{—} \circ_D \mid \mathbb{W}\,\rangle \\
&= \langle \text{———} \circ_a \text{—} \bullet_b \text{————————} \bullet_c \text{—} \circ_d \text{———} \mid \mathbb{W}\,\rangle \\
&= \langle \text{———} \circ_a \text{—} \bullet_b \text{————————} \bullet_c \text{—} \circ_d \text{———} \mid \mathbb{U}\,\rangle \\
&= \mu_{\mathbb{U}}([a,b]\times[c,d])
\end{aligned}$$

This proves $\mu_{\mathbb{U}}(R) \leq \mu_{\mathbb{V}}(R^{\delta})$. Then $\mu_{\mathbb{V}}(R) \leq \mu_{\mathbb{U}}(R^{\delta})$ follows by symmetry. □

Recall the measures at infinity defined in Sect. 3.6. By considering the appropriate limits, we immediately have:

Proposition 5.27 (Box inequalities at infinity) *Let μ, ν be r-measures on $\overline{\mathbf{R}}^2$ which satisfy a one-sided box inequality with parameter δ*

$$\mu(R) \leq \nu(R^{\delta})$$

for all rectangles $R \in \mathrm{Rect}(\overline{\mathbf{R}}^2)$. Then

$$\begin{aligned}
\mu([a,b],-\infty) &\leq \nu([A,B],-\infty), &\qquad \mu(-\infty,[c,d]) &\leq \nu(-\infty,[C,D]), \\
\mu([a,b],+\infty) &\leq \nu([A,B],+\infty), &\qquad \mu(+\infty,[c,d]) &\leq \nu(+\infty,[C,D]),
\end{aligned}$$

for all $a < b$ and $c < d$; and

$$\begin{aligned}
\mu(-\infty,-\infty) &\leq \nu(-\infty,-\infty), &\qquad \mu(+\infty,-\infty) &\leq \nu(+\infty,-\infty), \\
\mu(-\infty,+\infty) &\leq \nu(-\infty,+\infty), &\qquad \mu(+\infty,+\infty) &\leq \nu(+\infty,+\infty).
\end{aligned}$$

Here $A = a-\delta$, $B = b+\delta$, $C = c-\delta$, $D = d+\delta$. □

Consequently, if $\mathbb{U}, \mathbb{V}$ are δ-interleaved persistence modules then $\mu_{\mathbb{U}}, \mu_{\mathbb{V}}$ satisfy (two-sided) box inequalities on $(-\infty, \mathbf{R})$ and $(\mathbf{R}, +\infty)$ as well as the equality $\mu_{\mathbb{U}}(-\infty,+\infty) = \mu_{\mathbb{V}}(-\infty,+\infty)$.

5.7 The Measure Stability Theorem

We now embed Theorem 5.24 as a special case of a stability theorem for the diagrams of abstract r-measures. The more general statement is no more difficult[1] to prove, and seems to be the natural home for the result.

Let $\mathcal{D}$ be an open subset of $\overline{\mathbf{R}}^2$. For $\alpha \in \mathcal{D}$, define the **exit distance** of α to be

$$\mathrm{ex}^\infty(\alpha, \mathcal{D}) = \mathrm{d}^\infty(\alpha, \overline{\mathbf{R}}^2 - \mathcal{D}) = \min\left(\mathrm{d}^\infty(\alpha, x) \mid x \in \overline{\mathbf{R}}^2 - \mathcal{D}\right). \tag{5.2}$$

For instance, for the extended half-plane we have $\mathrm{ex}^\infty(\alpha, \overline{\mathcal{H}}) = \mathrm{d}^\infty(\alpha, \Delta)$.

Let A, B be multisets in $\mathcal{D}$. A δ**-matching** between A, B is a partial matching $\mathsf{M} \subset \mathsf{A} \times \mathsf{B}$ such that

$$\begin{aligned} \mathrm{d}^\infty(\alpha, \beta) &\le \delta && \text{if } \alpha, \beta \text{ are matched},\\ \mathrm{ex}^\infty(\alpha, \mathcal{D}) &\le \delta && \text{if } \alpha \in \mathsf{A} \text{ is unmatched},\\ \mathrm{ex}^\infty(\beta, \mathcal{D}) &\le \delta && \text{if } \beta \in \mathsf{B} \text{ is unmatched}. \end{aligned}$$

If $\mathcal{D}$ is not clear from the context, we refer to M as a 'δ-matching between $(\mathsf{A}, \mathcal{D})$ and $(\mathsf{B}, \mathcal{D})$'.

With the same proof as Proposition 5.9, we have:

Proposition 5.28 (triangle inequality) *If* $\mathsf{A}, \mathsf{B}, \mathsf{C}$ *are multisets in* $\mathcal{D}$ *and there exist a* δ_1*-matching between* $(\mathsf{A}, \mathcal{D}), (\mathsf{B}, \mathcal{D})$ *and a* δ_2*-matching between* $(\mathsf{B}, \mathcal{D}), (\mathsf{C}, \mathcal{D})$, *then there exists a* $(\delta_1 + \delta_2)$*-matching between* $(\mathsf{A}, \mathcal{D}), (\mathsf{C}, \mathcal{D})$. □

Now for the main theorem.

Theorem 5.29 (stability for finite measures) *Suppose* $(\mu_x \mid x \in [0, \delta])$ *is a* 1-*parameter family of finite r-measures on an open set* $\mathcal{D} \subseteq \overline{\mathbf{R}}^2$. *Suppose for all* $x, y \in [0, \delta]$ *the box inequality*

$$\mu_x(R) \le \mu_y(R^{|y-x|})$$

holds for all rectangles R *whose* $|y - x|$*-thickening* $R^{|y-x|}$ *belongs to* $\mathrm{Rect}(\mathcal{D})$. *Then there exists a* δ*-matching between the undecorated diagrams* $(\mathsf{dgm}(\mu_0), \mathcal{D})$ *and* $(\mathsf{dgm}(\mu_\delta), \mathcal{D})$.

In view of the interpolation Lemma 4.6, this implies Theorem 5.24 (take $\mu_x = \mu(\mathbb{U}_x)$ and $\mathcal{D} = \overline{\mathcal{H}^\circ}$) and therefore the stability theorem (5.14′) for q-tame modules.

Example 5.30 The existence of a 1-parameter family interpolating between μ_0 and μ_δ may seem unnecessarily strong. It is natural to hope that two measures μ, ν which satisfy the (two-sided) box inequality with parameter δ will have diagrams

[1]In fact it's a little easier to prove, because the compactness argument for diagrams with infinitely many points can be made more cleanly in this generality.

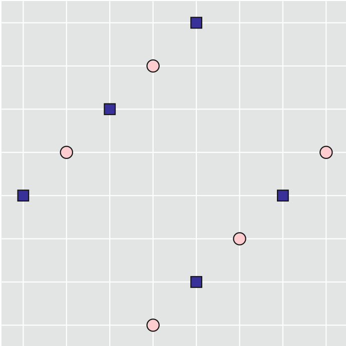


Fig. 5.2 The box inequalities do not control the bottleneck distance. The two diagrams (5 *dark blue squares*; 5 *light pink circles*) have box distance 1 and bottleneck distance 3. Generalising this example, one can exhibit a pair of diagrams with $4k + 1$ points each, whose box distance is 1 and whose bottleneck distance is $2k + 1$. By 'box distance 1' we mean that every rectangle R covers at most as many pink points as its 1-thickening R^1 covers blue points, and vice versa

$\mathsf{dgm}(\mu)$, $\mathsf{dgm}(\nu)$ which are δ-matched. This is not true, and in fact there is no universal bound on the bottleneck distance between the two diagrams. See Fig. 5.2.

Our goal for the rest of this section is to prove Theorem 5.29. Parts 1 and 2 closely follow the method of Cohen-Steiner, Edelsbrunner and Harer [19]. Afterwards, in Sect. 5.8, we generalise the theorem to r-measures that are not finite.

Initial remark. Because the metric d^∞ separates $\overline{\mathbf{R}}^2$ into nine strata (the standard plane, the four lines at infinity, and the four points at infinity), we seek separate δ-matchings for each stratum that meets $\mathcal{D}$. We begin with the points in the standard plane.

Temporary hypothesis. Suppose initially that $\mathcal{D} \subseteq \mathbf{R}^2$.

Part 1. *The Hausdorff distance between* $(\mathsf{dgm}(\mu_x), \mathcal{D})$, *and* $(\mathsf{dgm}(\mu_y), \mathcal{D})$ *is at most* $|y - x|$.

Write $\mathsf{A} = \mathsf{dgm}(\mu_x)$, $\mathsf{B} = \mathsf{dgm}(\mu_y)$, and $\eta = |y - x|$. The assertion is understood to mean:

- If $\alpha \in \mathsf{A}$ and $\mathrm{ex}^\infty(\alpha, \mathcal{D}) > \eta$, then there exists $\beta \in \mathsf{B}$ with $\mathrm{d}^\infty(\alpha, \beta) \leq \eta$.
- If $\beta \in \mathsf{B}$ and $\mathrm{ex}^\infty(\beta, \mathcal{D}) > \eta$, then there exists $\alpha \in \mathsf{A}$ with $\mathrm{d}^\infty(\alpha, \beta) \leq \eta$.

Proof By symmetry, it is enough to prove the first statement. Given α, let $\varepsilon > 0$ be small enough that $\eta + \varepsilon < \mathrm{ex}^\infty(\alpha, \mathcal{D})$. Then the box inequality gives

$$1 \leq \mu_x(\alpha^\varepsilon) \leq \mu_y(\alpha^{\eta+\varepsilon})$$

so there is at least one point of B in the square $\alpha^{\eta+\varepsilon}$. This is true for all sufficiently small $\varepsilon > 0$, and moreover B is locally finite. Therefore there is at least one point of B in α^η. □

Henceforth, we will write $\mathsf{A}_x = \mathsf{dgm}(\mu_x)$ for all x.

Part 2. *The theorem is true if* A_x *has finite cardinality for all* x.

Proof (i) The triangle inequality for matchings includes the implication

$$\left.\begin{array}{l}\mathsf{A}_0, \mathsf{A}_x \text{ are } x\text{-matched}\\ \mathsf{A}_x, \mathsf{A}_y \text{ are } (y-x)\text{-matched}\end{array}\right\} \Rightarrow \mathsf{A}_0, \mathsf{A}_y \text{ are } y\text{-matched}$$

whenever $0 < x < y$.

(ii) We claim that for every $x \in [0, \delta]$ there exists $\rho(x) > 0$ such that $\mathsf{A}_x, \mathsf{A}_y$ are $|y - x|$-matched whenever $y \in [0, \delta]$ with $|y - x| < \rho(x)$.

Suppose $\alpha_1, \dots, \alpha_k$ is an enumeration of the distinct points of A_x, with respective multiplicities $n_1, \dots, n_k$. Let $\rho(x)$ be chosen to satisfy the following finite set of constraints:

$$0 < \rho(x) \leq \begin{cases} \frac{1}{2}\mathrm{ex}^\infty(\alpha_i, \mathcal{D}) & \text{all } i \\ \frac{1}{2}\mathrm{d}^\infty(\alpha_i, \alpha_j) & \text{all } i,j \text{ distinct} \end{cases}$$

We must show that if $|y - x| < \rho(x)$ then $\mathsf{A}_x, \mathsf{A}_y$ are $|y - x|$-matched. Write $\eta = |y - x|$ and let

$$(\mathbf{R}^2 - \mathcal{D})^\eta = \left\{\alpha \in \mathcal{D} \mid \mathrm{ex}^\infty(\alpha, \mathcal{D}) \leq \eta\right\}.$$

It follows from Part 1 that A_y is contained entirely in the closed set

$$(\mathbf{R}^2 - \mathcal{D})^\eta \cup \alpha_1^\eta \cup \dots \cup \alpha_k^\eta$$

and it follows from the definition of $\rho(x)$ that the terms in the union are disjoint. It is easy to count the points of A_y in each square α_i^η. Let $\varepsilon > 0$ be small enough that $2\eta + \varepsilon < 2\rho(x)$. Then the box inequality gives

$$n_i = \mu_x(\alpha_i^\varepsilon) \leq \mu_y(\alpha_i^{\eta+\varepsilon}) \leq \mu_x(\alpha_i^{2\eta+\varepsilon}) = n_i.$$

Thus $\mu_y(\alpha_i^{\eta+\varepsilon}) = n_i$ for all small $\varepsilon > 0$. We conclude that the square α_i^η contains precisely n_i points of A_y.

This completes the proof of (ii), because we can match the n_i copies of α_i with the n_i points of A_y in the square α_i^η, for each i, to define an η-matching between $(\mathsf{A}_x, \mathcal{D})$, $(\mathsf{A}_y, \mathcal{D})$. All points of A_x are matched, and the only unmatched points of A_y lie in $\mathbf{R}^2 - \mathcal{D}$ and do not need to be matched.

Items (i) and (ii) formally imply that $\mathsf{A}_0, \mathsf{A}_\delta$ are δ-matched, using a standard Heine–Borel argument. Indeed, let

$$m = \sup(x \in [0, \delta] \mid \mathsf{A}_0 \text{ and } \mathsf{A}_x \text{ are } x\text{-matched}).$$

First, m is positive; specifically $m \geq \rho(0)$. Applying (i) to $0 < m' < m$, where $\mathsf{A}_0, \mathsf{A}_{m'}$ are m'-matched and $m - m' < \rho(m)$, we deduce that $\mathsf{A}_0, \mathsf{A}_m$ are m-matched. Suppose

$m < \delta$. Applying (i) to $0 < m < m''$, where $m'' - m < \rho(m)$, we deduce that $\mathsf{A}_0, \mathsf{A}_{m''}$ are m''-matched. This contradicts the definition of m. Therefore $m = \delta$, and $\mathsf{A}_0, \mathsf{A}_\delta$ are δ-matched. □

Part 3. *The theorem is true without assuming finite cardinality.*

Proof Let $(\mathcal{D}_n)$ be an increasing sequence of open subsets of $\mathcal{D}$ whose union equals $\mathcal{D}$ and such that each $\mathcal{D}_n$ has compact closure. Because A_x is locally finite, it follows that $\mathsf{A}_x \cap \mathcal{D}_n$ is finite for all x, n. We can therefore restrict the family of measures to each $\mathcal{D}_n$ in turn, and apply Part 2 to get a δ-matching M_n between $(\mathsf{A}_0 \cap \mathcal{D}_n, \mathcal{D}_n)$ and $(\mathsf{A}_\delta \cap \mathcal{D}_n, \mathcal{D}_n)$.

We now take a limit M of the partial matchings M_n, using the construction in the proof of Theorem 5.12. This works because $\mathsf{A}_0, \mathsf{A}_\delta$ are locally finite and therefore countable. Let χ, χ_n denote the indicator functions of M, M_n. Recall Lemma 5.13: for any finite subset $\mathsf{F} \subset \mathsf{A}_0 \times \mathsf{A}_\delta$, there are infinitely many $n \in \mathbf{N}$ for which

$$\chi(\alpha, \beta) = \chi_n(\alpha, \beta)$$

for all $(\alpha, \beta) \in \mathsf{F}$.

We must show that M is a δ-matching between $(\mathsf{A}_0, \mathcal{D})$ and $(\mathsf{A}_\delta, \mathcal{D})$. It is immediate that each matched pair is separated by at most δ, since this is true for every M_n. The argument that each α is matched with at most one β, and vice versa, is the same as in Lemma 5.13.

Finally, suppose $\alpha \in \mathsf{A}_0$ with $\mathrm{ex}^\infty(\alpha, \mathcal{D}) > \delta$. The square α^δ is contained in $\mathcal{D}$ and is compact, and therefore is contained in $\mathcal{D}_n$ for sufficiently large n. This means that $\mathrm{ex}^\infty(\alpha, \mathcal{D}_n) > \delta$ and hence α is matched in M_n for sufficiently large n. Now α has only finitely many δ-neighbours $\beta_1, \dots, \beta_k$ in the locally finite set A_δ, so by Lemma 5.13 there are infinitely many n such that $\chi(\alpha, \beta_i) = \chi_n(\alpha, \beta_i)$ for all i. By taking a sufficiently large such n, we conclude that

$$\chi(\alpha, \beta_i) = \chi_n(\alpha, \beta_i) = 1$$

for some i. Thus α is matched.

By symmetry, any $\beta \in \mathsf{A}_\delta$ with $\mathrm{ex}^\infty(\beta, \Delta) > \delta$ is matched in M to some $\alpha \in \mathsf{A}_0$. It follows that M is the required δ-matching. □

The theorem at infinity. Now suppose $\mathcal{D} \subseteq \overline{\mathbf{R}}^2$ meets any of the strata at infinity. For each of the four lines at infinity, the 3-part proof given above works almost verbatim, if we replace $\mathcal{D}$ with its intersection with the chosen line, and each r-measure μ_x with the corresponding measure at infinity. The other change is to replace the word 'square' with the word 'interval'. The necessary box inequality at infinity is found in Proposition 5.27.

For the four corners $(\pm\infty, \pm\infty)$, it is easier still: the box inequality at each corner implies that μ_0, μ_δ have the same multiplicity there. The interpolating measures are not needed. □

This completes the proof of the stability theorem for finite measures on an open region $\mathcal{D}$, and hence the stability theorem for q-tame persistence modules, and hence the isometry theorem for q-tame persistence modules.

5.8 The Measure Stability Theorem (Continued)

The stability theorem generalises to measures that are not necessarily finite. By the region extension convention (Remark 3.22), we may suppose that the measures are defined on $\overline{\mathbf{R}}^2$ (rather than just a subset of $\overline{\mathbf{R}}^2$). Given a 1-parameter family $(\mu_x \mid x \in [0, \delta])$, the finite interiors

$$\mathcal{F}_x = \mathcal{F}^{\circ}(\mu_x)$$

now depend on x; whereas previously we had $\mathcal{F}_x = \mathcal{D}$ for all x.

For $\mathcal{F} \subset \overline{\mathbf{R}}^2$ an open set and $\delta \geq 0$, the 'reverse offset' is the open set

$$\mathcal{F}^{-\delta} = \left\{\alpha \in \mathcal{F} \mid \mathrm{ex}^{\infty}(\alpha, \mathcal{F}) > \delta\right\} = \left\{\alpha \in \mathcal{F} \mid \alpha^{\delta} \subset \mathcal{F}\right\}.$$

Intuitively, this shrinks $\mathcal{F}$ by δ at the boundary. Clearly $\mathcal{F} \supseteq \mathcal{G}$ implies $\mathcal{F}^{-\delta} \supseteq \mathcal{G}^{-\delta}$, and $(\mathcal{F}^{-\delta_1})^{-\delta_2} = \mathcal{F}^{-(\delta_1+\delta_2)}$. Note also that $(\mathcal{F} \cap \mathcal{G})^{-\delta} = \mathcal{F}^{-\delta} \cap \mathcal{G}^{-\delta}$. This is easiest to see from the second characterisation.

Remark 5.31 The operation $[\cdot]^{-\delta}$ has no effect on the corners at infinity, and acts independently on the standard plane and on the four lines at infinity.

We define δ-matchings for multisets in unequal regions. Let $\mathcal{F}, \mathcal{G}$ be open subsets of $\overline{\mathbf{R}}^2$, let A, B be multisets in $\mathcal{F}, \mathcal{G}$ respectively, and let $\delta > 0$. A δ **-matching between** $(\mathsf{A}, \mathcal{F}), (\mathsf{B}, \mathcal{G})$ is a partial matching M between A, B such that the following four conditions hold:

- $\mathcal{F} \supseteq \mathcal{G}^{-\delta}$ and $\mathcal{G} \supseteq \mathcal{F}^{-\delta}$,
- if $(\alpha, \beta) \in \mathsf{M}$ then $\mathrm{d}^{\infty}(\alpha, \beta) \leq \delta$,
- every $\alpha \in \mathsf{A} \cap \mathcal{G}^{-\delta}$ is matched with some $\beta \in \mathsf{B}$,
- every $\beta \in \mathsf{B} \cap \mathcal{F}^{-\delta}$ is matched with some $\alpha \in \mathsf{A}$.

The first of these is a compatibility condition between the regions: they cannot be too unequal. This is automatic if $\mathcal{F} = \mathcal{G}$, which is why we haven't seen it before. Notice the cross-over in the last two conditions: a point in A is allowed to be unmatched only if it is close to the boundary of B's region $\mathcal{G}$, and vice versa.

Proposition 5.32 (triangle inequality) *If* $\mathsf{A}, \mathsf{B}, \mathsf{C}$ *are multisets in* $\mathcal{F}, \mathcal{G}, \mathcal{H}$ *respectively, and there exist a* δ_1*-matching between* $(\mathsf{A}, \mathcal{F}), (\mathsf{B}, \mathcal{G})$ *and a* δ_2*-matching between* $(\mathsf{B}, \mathcal{G}), (\mathsf{C}, \mathcal{H})$*, then there exists a* $(\delta_1 + \delta_2)$*-matching between* $(\mathsf{A}, \mathcal{F})$, $(\mathsf{C}, \mathcal{H})$.

Proof As usual, compose the two partial matchings to get a partial matching M between A, C. Writing $\delta = \delta_1 + \delta_2$, we must check that this is a δ-matching between $(\mathsf{A}, \mathcal{F})$, $(\mathsf{C}, \mathcal{H})$. For the first condition we see that

$$\mathcal{F} \supseteq \mathcal{G}^{-\delta_1} \supseteq (\mathcal{H}^{-\delta_2})^{-\delta_1} = \mathcal{H}^{-\delta} \quad \text{and} \quad \mathcal{H} \supseteq \mathcal{G}^{-\delta_2} \supseteq (\mathcal{F}^{-\delta_1})^{-\delta_2} = \mathcal{F}^{-\delta}.$$

The second condition follows from the triangle inequality for d^∞. For the third condition, if $\alpha \in \mathsf{A}$ lies in $\mathcal{H}^{-\delta}$ then by the inclusion above it lies in $\mathcal{G}^{-\delta_1}$. Therefore α is matched with $\beta \in \mathsf{B}$. Moreover β must then lie in $\mathcal{H}^{-(\delta-\delta_1)} = \mathcal{H}^{-\delta_2}$ and so is matched with $\gamma \in \mathsf{C}$. The fourth condition follows by symmetry. □

Remark 5.33 There is no triangle inequality if the compatibility condition between the regions is dropped in the definition of δ-matching.

Here is the main theorem of this section and the last new theorem of the monograph. Again we use the abbreviation $\mathcal{F}_x = \mathcal{F}^\circ(\mu_x)$ for the finite interiors.

Theorem 5.34 (stability for measures) *Suppose* $(\mu_x \mid x \in [0, \delta])$ *is a* 1*-parameter family of r-measures on* $\overline{\mathbf{R}}^2$. *Suppose for all* $x, y \in [0, \delta]$ *the box inequality*

$$\mu_x(R) \le \mu_y(R^{|y-x|})$$

holds for all rectangles $R \in \mathrm{Rect}(\overline{\mathbf{R}}^2)$. *Then there exists a* δ*-matching between the undecorated diagrams* $(\mathsf{dgm}(\mu_0), \mathcal{F}_0)$ *and* $(\mathsf{dgm}(\mu_\delta), \mathcal{F}_\delta)$.

Remark 5.35 This version of measure stability allows us to dispense with the final assertion in Theorem 4.7 when deducing stability (5.14′) for q-tame modules: we no longer need to assume that the interpolating modules are themselves q-tame. The reader may wish to consider why this works. It results from careful management of the boundary.

An easy first step is to verify the condition on the regions $\mathcal{F}_0, \mathcal{F}_\delta$:

Proposition 5.36 *Under the hypotheses of Theorem 5.34, we have inclusions*

$$\mathcal{F}_x \supseteq \mathcal{F}_y^{-|y-x|}$$

for all $x, y \in [0, \delta]$.

Proof Suppose $\alpha \in \mathcal{F}_y^{-|y-x|}$, then equivalently $\alpha^{|y-x|} \subset \mathcal{F}_y$. Since the square $\alpha^{|y-x|}$ is compact and $\mathcal{F}_y$ is open, there exists $\varepsilon > 0$ such that $\alpha^{|y-x|+\varepsilon} \subset \mathcal{F}_y$. The box inequality gives

$$\mu_x(\alpha^\varepsilon) \le \mu_y(\alpha^{|y-x|+\varepsilon})$$

and the right-hand side is finite by Proposition 3.18. Thus $\alpha \in \mathcal{F}^\circ(\mu_x) = \mathcal{F}_x$. □

Proof (*Theorem* 5.34) The argument closely follows the proof of the stability theorem for finite measure, so we will confine ourselves to indicating the necessary modifications. We use the abbreviation $\mathsf{A}_x = \mathsf{dgm}(\mu_x)$.

Initial remark. Recall that the proof is carried out separately for each of the nine strata. The four corners at infinity are handled easily (each corner belongs to both $\mathcal{F}_0$ and $\mathcal{F}_\delta$, or to neither; in the former case the μ_0, μ_δ multiplicities agree). The proof is described for the points in the standard plane. The same proof applies to each of the four lines at infinity, replacing each μ_x with the corresponding measure at infinity.

Part 1. *The Hausdorff distance between* $(\mathsf{A}_x, \mathcal{F}_x)$, *and* $(\mathsf{A}_y, \mathcal{F}_y)$ *is at most* $\eta = |y - x|$.

The assertion is understood to mean:

- If $\alpha \in \mathsf{A}_x$ and $\mathrm{ex}^\infty(\alpha, \mathcal{F}_y) > \eta$, then there exists $\beta \in \mathsf{A}_y$ with $\mathrm{d}^\infty(\alpha, \beta) \leq \eta$.
- If $\beta \in \mathsf{A}_y$ and $\mathrm{ex}^\infty(\beta, \mathcal{F}_x) > \eta$, then there exists $\alpha \in \mathsf{A}_x$ with $\mathrm{d}^\infty(\alpha, \beta) \leq \eta$.

Proof By symmetry it is enough to prove the first statement. Let $\alpha \in \mathsf{A}_x$. For all $\varepsilon > 0$ with $\eta + \varepsilon < \mathrm{ex}^\infty(\alpha, \mathcal{F}_y)$, we have $1 \leq \mu_x(\alpha^\varepsilon) \leq \mu_y(\alpha^{\eta+\varepsilon})$ so there is at least one point of A_y in α^η. □

Part 2. *The theorem is true if* A_x *has finite cardinality for all* x.

Proof Item (i) is given by the triangle inequality (Proposition 5.32).

Item (ii) uses the same strategy as before. Let (α_i) be a finite enumeration of the distinct points of A_x, with respective multiplicities (n_i). Then $\rho(x)$ is chosen to satisfy

$$0 < \rho(x) \leq \begin{cases} \frac{1}{2}\mathrm{ex}^\infty(\alpha_i, \mathcal{F}_x) & \text{all } i \\ \frac{1}{2}\mathrm{d}^\infty(\alpha_i, \alpha_j) & \text{all } i, j \text{ distinct.} \end{cases}$$

If $\eta = |y - x| < \rho(x)$, then Part 1 implies that A_y is contained in the disjoint union

$$(\mathbf{R}^2 - \mathcal{F}_x)^\eta \cup \alpha_1^\eta \cup \cdots \cup \alpha_k^\eta.$$

The box inequality is then used to count precisely n_i points of A_y in the square α_i^η. This defines a partial matching where all points of A_x are matched and all points of $\mathsf{A}_y \cap \mathcal{F}_x^{-\eta}$ are matched.

The formal deduction of Part 2 from (i) and (ii) is unchanged, since it is a formal deduction.

Part 3. *The theorem is true without assuming finite cardinality.*

Proof The idea is to restrict each measure μ_x to a relatively compact open subset $\hat{\mathcal{F}}_x \subset \mathcal{F}_x = \mathcal{F}^\circ(\mu_x)$. The subsets satisfy the compatibility condition

$$\hat{\mathcal{F}}_x \supseteq \hat{\mathcal{F}}_y^{-|y-x|}$$

for all $x, y \in [0, \delta]$.

Specifically, for $\varepsilon > 0$ and $r > \delta$, let

$$\hat{\mathcal{F}}_x = \mathcal{F}_x^{-\varepsilon} \cap \mathcal{Q}^r$$

where $\mathcal{Q}^r = (-r, r) \times (-r, r)$ is the open d^∞-disk of radius r. Define a function on rectangles as follows:

$$\hat{\mu}_x(R) = \begin{cases} \mu_x(R) & \text{if } R \subset \hat{\mathcal{F}}_x \\ \infty & \text{otherwise} \end{cases}$$

It is easy to check that $\hat{\mu}_x$ is an r-measure (additivity still holds), that $\mathcal{F}^{\circ}(\hat{\mu}_x) = \hat{\mathcal{F}}_x$, and that $\mathsf{dgm}(\hat{\mu}_x) = \mathsf{dgm}(\mu_x) \cap \hat{\mathcal{F}}_x$.

Lemma 5.37 *The family $(\hat{\mu}_x)$ satisfies the box inequality $\hat{\mu}_x(R) \leq \hat{\mu}_y(R^{|y-x|})$ for all $x, y \in [0, \delta]$.*

Proof Since the box inequality is assumed to hold for (μ_x), it will automatically hold for $(\hat{\mu}_x)$; except possibly for rectangles R where the left-hand side of the inequality has become infinite while the right-hand side hasn't. This happens when $R \not\subset \hat{\mathcal{F}}_x$ while $R^{|y-x|} \subset \hat{\mathcal{F}}_y$, and we can prevent it by ensuring that $\hat{\mathcal{F}}_x \supseteq \hat{\mathcal{F}}_y^{-|y-x|}$. And, indeed,

$$\hat{\mathcal{F}}_y^{-|y-x|} = (\mathcal{F}_y^{-\varepsilon} \cap \mathcal{Q}^r)^{-|y-x|} = \mathcal{F}_y^{-(\varepsilon+|y-x|)} \cap \mathcal{Q}^{r-|y-x|} \subseteq \mathcal{F}_x^{-\varepsilon} \cap \mathcal{Q}^r = \hat{\mathcal{F}}_x$$

as required. □

We resume the proof of Part 3. Since $\hat{\mathcal{F}}_x$ has compact closure in $\mathcal{F}_x$, and A_x is locally finite, it follows that $\hat{\mathsf{A}}_x = \mathsf{dgm}(\hat{\mu}_x) = \mathsf{A}_x \cap \hat{\mathcal{F}}_x$ has finite cardinality. We can therefore apply Part 2 to the family $(\hat{\mu}_x)$ to get a δ-matching between $(\hat{\mathsf{A}}_0, \hat{\mathcal{F}}_0)$ and $(\hat{\mathsf{A}}_\delta, \hat{\mathcal{F}}_\delta)$. This can be interpreted as a partial δ-matching between $\mathsf{A}_0, \mathsf{A}_\delta$ where:

- $\alpha \in \mathsf{A}_0$ is matched whenever $\alpha \in (\mathcal{F}_\delta^{-\varepsilon} \cap \mathcal{Q}^r)^{-\delta} = \mathcal{F}_\delta^{-(\delta+\varepsilon)} \cap \mathcal{Q}^{r-\delta}$
- $\beta \in \mathsf{A}_\delta$ is matched whenever $\beta \in (\mathcal{F}_0^{-\varepsilon} \cap \mathcal{Q}^r)^{-\delta} = \mathcal{F}_0^{-(\delta+\varepsilon)} \cap \mathcal{Q}^{r-\delta}$

Repeat this argument for a sequence (ε_n, r_n) where $\varepsilon_n \to 0$ and $r_n \to +\infty$. This gives a sequence of δ-matchings M_n, and we can form a limit M as before.

If $\alpha \in \mathsf{A}_0 \cap \mathcal{F}_\delta^{-\delta}$ then eventually $\alpha \in \mathcal{F}_\delta^{-(\delta+\varepsilon_n)} \cap \mathcal{Q}^{r_n-\delta}$ and so α is matched by M_n for all sufficiently large n. The same is true for $\beta \in \mathsf{A}_\delta \cap \mathcal{F}_0^{-\delta}$. With this information, we can complete the usual proof that M is a δ-matching between $(\mathsf{A}_0, \mathcal{F}_0)$ and $(\mathsf{A}_\delta, \mathcal{F}_\delta)$. This completes the proof of Part 3, and hence of Theorem 5.34. □

Here is a sample consequence.

Example 5.38 (*Stability of the Webb module*) Let $\mathbb{V}$ be a persistence module which is δ-interleaved with the module $\mathbb{W}$ of Example 3.31. By interpolation (Lemma 4.6) and the box inequalities (Lemma 5.26), we can apply the measure stability theorem (Theorem 5.34): there exists a δ-matching between the undecorated diagrams $(\mathsf{dgm}(\mu_{\mathbb{V}}), \mathcal{F}^{\circ}(\mu_{\mathbb{V}}))$ and $(\mathsf{dgm}(\mu_{\mathbb{W}}), \mathcal{F}^{\circ}(\mu_{\mathbb{W}}))$. This amounts to the following.

- **In the finite part $\mathcal{H}$ of the half-plane**: Any singular points of $\mu_{\mathbb{V}}$ are confined to the diagonal strip $\Delta_{[0,\delta]}$. Each point of $\mathsf{dgm}(\mu_{\mathbb{V}})$ outside this strip is matched with some point $(-n, 0) \in \mathsf{dgm}(\mu_{\mathbb{W}})$. Conversely, the only unmatched points of $\mathsf{dgm}(\mu_{\mathbb{W}})$ must lie within distance δ of the diagonal or a singular point of $\mu_{\mathbb{V}}$. In particular, if $\delta < \frac{1}{4}$ then all points of $\mathsf{dgm}(\mu_{\mathbb{W}})$ are matched.
- **On the line** $(-\infty, \mathbf{R})$: All points and singular points of $\mu_{\mathbb{V}}$ are contained in the interval $(-\infty, [-\delta, +\delta])$. There is at least one singular point.
- **On** $(\mathbf{R}, +\infty)$ **and at** $(-\infty, +\infty)$: The measure $\mu_{\mathbb{V}}$ has no points or singular points.

Chapter 6
Variations

6.1 Partial Interleavings

In some practical data analysis situations, one considers persistence modules which are only partially interleaved. One such scenario is presented by Chazal et al. in the context of clustering by mode-seeking [18]. A filtered simplicial complex on an input point cloud is compared with the sublevelset filtration of the density function it was sampled from. In low-density regions, the sample is too sparse to expect there to be an interleaving. Nevertheless, there is interleaving when the density is sufficiently high.

This leads to the following notion of partial interleaving, adapted from [18]. Two persistence modules $\mathbb{U}$ and $\mathbb{V}$ are said to be **δ-interleaved up to time** t_0 if there are maps $\phi_t : U_t \to V_{t+\delta}$ and $\psi_t : V_t \to U_{t+\delta}$ defined for all $t \leq t_0$, such that the diagrams (4.1) commute for all values $s < t \leq t_0$; that is, for all values where the maps are defined.

We can prove a modified version of the stability theorem; see Fig. 6.1 (left).

Theorem 6.1 (from [18]) *Let $\mathbb{U}$ and $\mathbb{V}$ be two q-tame persistence modules that are δ-interleaved up to time t_0. Then, there is a partial matching $\mathsf{M} \subset \mathsf{dgm}(\mathbb{U}) \times \mathsf{dgm}(\mathbb{V})$ with the following properties:*

- *Points (p, q) in either diagram for which $\frac{1}{2}|p - q| \leq \delta$ are not required to be matched.*
- *Points (p, q) in either diagram for which $p \geq t_0 - \delta$ are not required to be matched.*

All other points must be matched. Then:

- *If α, β are matched, then the p-coordinates of α, β differ by at most δ.*
- *If α, β are matched and one of α, β lies below the line $q = t_0$, then we have $\mathrm{d}^\infty(\alpha, \beta) \leq \delta$.*

F. Chazal et al., *The Structure and Stability of Persistence Modules*,
SpringerBriefs in Mathematics, DOI 10.1007/978-3-319-42545-0_6

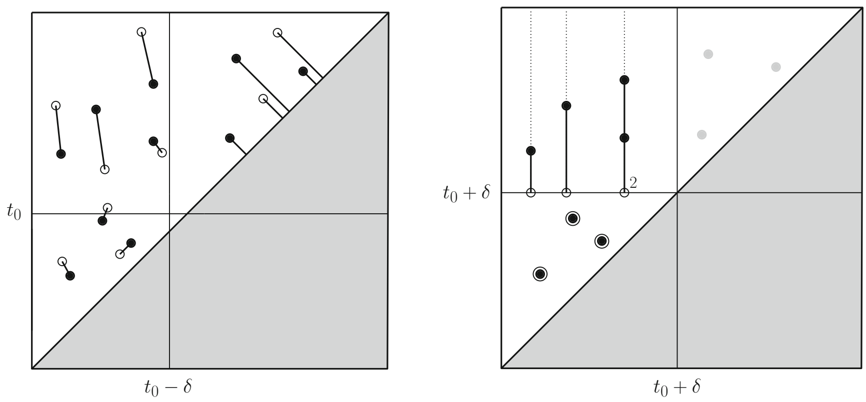


Fig. 6.1 *Left:* The partial matching of Theorem 6.1 between $\mathsf{dgm}(\mathbb{U})$ (•) and $\mathsf{dgm}(\mathbb{V})$ (∘). *Right:* The projection from $\mathsf{dgm}(\mathbb{U})$ (• and •) to $\mathsf{dgm}(\tilde{\mathbb{U}})$ (∘). The *grey dots* are the points that disappear

For the proof, we introduce two new persistence modules $\tilde{\mathbb{U}}$, $\tilde{\mathbb{V}}$.

$$\begin{array}{lll} \tilde{U}_t = U_t & \text{if } t \le t_0 + \delta \text{ and } \tilde{U}_t = 0 & \text{otherwise} \\ \tilde{V}_t = V_t & \text{if } t \le t_0 + \delta \text{ and } \tilde{V}_t = 0 & \text{otherwise} \end{array}$$

with maps

$$\begin{array}{lll} \tilde{u}_t^s = u_t^s & \text{if } t \le t_0 + \delta \text{ and } \tilde{u}_t^s = 0 & \text{otherwise} \\ \tilde{v}_t^s = v_t^s & \text{if } t \le t_0 + \delta \text{ and } \tilde{v}_t^s = 0 & \text{otherwise} \end{array}$$

for all $s \le t$. We may call $\hat{\mathbb{U}}$, $\hat{\mathbb{V}}$ the **truncations** of $\mathbb{U}$, $\mathbb{V}$ to $(-\infty, T]$, where $T = t_0 + \delta$.

Proof There are three steps.

Step 1. The decorated diagram of a persistence module $\mathbb{U}$ determines the decorated diagram of its truncation $\tilde{\mathbb{U}}$, in a straightforward way. Specifically, transform each point $(p^*, q^*) \in \mathsf{Dgm}(\mathbb{U})$ as follows:

$$(p^*, q^*) \mapsto \begin{cases} (p^*, q^*) & \text{if } q^* < T^+ \\ (p^*, T^+) & \text{if } p^* < T^+ \le q^* \\ \text{disappears} & \text{if } T^+ \le p^* \end{cases} \tag{6.1}$$

Then $\mathsf{Dgm}(\tilde{\mathbb{U}})$ is the result of this transformation. The consequent relationship between the undecorated diagrams is illustrated in Fig. 6.1 (right).

Step 2. If $\mathbb{U}$, $\mathbb{V}$ are δ-interleaved up to time t_0, then $\tilde{\mathbb{U}}$, $\tilde{\mathbb{V}}$ are δ-interleaved.

Combining the first two steps we get the third.

Step 3. The stability theorem gives a δ-matching between $\mathsf{dgm}(\tilde{\mathbb{U}})$, $\mathsf{dgm}(\tilde{\mathbb{V}})$. This lifts to a matching between $\mathsf{dgm}(\mathbb{U})$, $\mathsf{dgm}(\mathbb{V})$ which has the properties stated in the theorem.

The second and third steps are straightforward. Only the (intuitively plausible) first step requires any technical input. The framework developed in [14] leads to a 2-page argument, presented in the appendix of [18]. Here is a shorter proof. Write $\mu = \mu_{\mathbb{U}}$ and $\tilde{\mu} = \mu_{\tilde{\mathbb{U}}}$. Let A denote the multiset obtained from $\mathsf{Dgm}(\mathbb{U})$ by applying the transformation in Eq. (6.1). Consider an arbitrary rectangle $[a, b] \times [c, d] \in \mathrm{Rect}(\overline{\mathcal{H}})$. We easily see:

$$\mathrm{card}(\mathsf{A}|_{[a,b]\times[c,d]}) = \begin{cases} \mu([a, b] \times [c, d]) & \text{if } d \leq T \\ \mu([a, b] \times [c, +\infty]) & \text{if } c \leq T < d \\ 0 & \text{if } T < c \end{cases}$$

To show that we have correctly determined $\mathsf{Dgm}(\tilde{\mathbb{U}})$, it suffices to show that $\mathrm{card}(\mathsf{A}|_{[a,b]\times[c,d]}) = \tilde{\mu}([a, b] \times [c, d])$ for all rectangles. And indeed:

- If $d \leq T$, then:

$$\begin{aligned} \tilde{\mu}([a, b] \times [c, d]) &= \langle \circ_a\!\!-\!\!\bullet_b\!\!-\!\!\bullet_c\!\!-\!\!\circ_d \mid \tilde{\mathbb{U}} \rangle \\ &= \langle \circ_a\!\!-\!\!\bullet_b\!\!-\!\!\bullet_c\!\!-\!\!\circ_d \mid \mathbb{U} \rangle = \mu([a, b] \times [c, d]) \end{aligned}$$

- If $c \leq T < d$, then:

$$\begin{aligned} \tilde{\mu}([a, b] \times [c, d]) &= \langle \circ_a\!\!-\!\!\bullet_b\!\!-\!\!\bullet_c\!\!-\!\!\circ_d \mid \tilde{\mathbb{U}} \rangle \\ &= \langle \circ_a\!\!-\!\!\bullet_b\!\!-\!\!\bullet_c\!\!-\!\!- \mid \tilde{\mathbb{U}} \rangle \\ &= \langle \circ_a\!\!-\!\!\bullet_b\!\!-\!\!\bullet_c\!\!-\!\!- \mid \mathbb{U} \rangle = \mu([a, b] \times [c, +\infty]) \end{aligned}$$

  since $\tilde{U}_d = 0$.
- If $T < c$, then:

$$\tilde{\mu}([a, b] \times [c, d]) = \langle \circ_a\!\!-\!\!\bullet_b\!\!-\!\!\bullet_c\!\!-\!\!\circ_d \mid \tilde{\mathbb{U}} \rangle = 0$$

  since $\tilde{U}_c = 0$.

It follows that $\mathsf{Dgm}(\tilde{\mathbb{U}}) = \mathsf{A}$ as claimed. □

6.2 Extended Persistence

Cohen-Steiner, Edelsbrunner and Harer [20] introduced extended persistence to capture the homological information carried by a pair (X, f). Some but not all of this information is recovered by the sublevelset persistence $\mathrm{H}(\mathbb{X}_{\mathrm{sub}})$. The idea is to grow the space from the bottom up, through sublevelsets; and then to relativise the space from the top down, with superlevelsets. Extended persistence is the persistent homology of this sequence of spaces and pairs.

It is usually assumed that (X, f) has finitely many homological critical points (a_i). One applies a homology functor to the finite sequence[1]

$$X^{a_0} \to X^{a_1} \to \cdots \to X^{a_{n-1}} \to X \to (X, X_{a_n}) \to \cdots \to (X, X_{a_2}) \to (X, X_{a_1})$$

to get a quiver representation. The indecomposable summands of this representation are interpreted as features, and are drawn as points in the 'extended persistence diagram'. There are three kinds of feature:

- ordinary features (which are born and die before the central X);
- relative features (which are born and die after the central X);
- extended features (which are born before the X and die after it).

We refer to [20] for the interpretation of these three types of features. The finiteness assumption is satisfied when (X, f) is a compact manifold with a Morse function, or a compact polyhedron with a piecewise-linear map. In the former situation, there are extra symmetries (Poincaré, Lefschetz) which are explored in [20].

In practice, it is straightforward to define the extended persistence diagram under a weaker hypothesis. Suppose X is a compact polyhedron and f is a continuous real-valued function on X. Then:

- rank $\left(\mathrm{H}(X^s) \to \mathrm{H}(X^t)\right) < \infty$ whenever $s < t$; and
- rank $(\mathrm{H}(X, X_s) \to \mathrm{H}(X, X_t)) < \infty$ whenever $s > t$.

The first of these facts is Theorem 3.33. The second fact is proved similarly, by factorising the map $\mathrm{H}(X, X_s) \to \mathrm{H}(X, X_t)$ through some $\mathrm{H}(X, Y)$, where Y is a subpolyhedron of X nested between X_s, X_t. Since $\mathrm{H}(X, Y)$ is finite-dimensional the result follows.

Define the ordered set

$$\mathbf{Я} = \{\underline{t} \mid t \in \mathbf{R}\} \quad \text{ordered by} \quad \underline{s} \leq \underline{t} \;\Leftrightarrow\; s \geq t,$$

thought of as a 'backwards' copy of the real line, with bars under numbers to remind us. For extended persistence we may work with the set

$$\mathbf{R}_{\mathrm{EP}} = \mathbf{R} \cup \{+\infty\} \cup \mathbf{Я}$$

with the ordering $s < +\infty < \underline{t}$ for all $s, \underline{t}$.

[1] We write $X^t = (X, f)^t = f^{-1}(-\infty, t]$ and $X_t = (X, f)_t = f^{-1}[t, +\infty)$ for sublevelsets and superlevelsets.

The extended persistence module $\mathbb{X}_{\mathrm{EP}} = \mathbb{X}^f_{\mathrm{EP}}$ for (X, f) is defined as follows:

$$
\begin{aligned}
V_t &= \mathrm{H}(X^t) && \text{for } t \in \mathbf{R} \\
V_{+\infty} &= \mathrm{H}(X) \\
V_{\underline{t}} &= \mathrm{H}(X, X_t) && \text{for } \underline{t} \in \mathbf{Я}
\end{aligned}
$$

Note that $\mathrm{H}(X^{+\infty}) = \mathrm{H}(X) \cong \mathrm{H}(X, \emptyset) = \mathrm{H}(X, X_{+\infty})$.

Since $\mathbf{R}_{\mathrm{EP}}$ is order-isomorphic to the real line, we may interpret $\mathbb{X}_{\mathrm{EP}}$ it as a persistence module over $\mathbf{R}$. The two facts cited above imply that it is q-tame, so the decorated diagram is defined away from the diagonal.

Alternatively, we can define the extended persistence diagram in three pieces:

$$
\begin{aligned}
\mu_{\mathrm{ord}}([a,b]\times[c,d]) &= \langle \circ_a\!\!-\!\!\bullet_b\!\!-\!\!\bullet_c\!\!-\!\!\circ_d \rangle && \text{when } a < b \le c < d \\
\mu_{\mathrm{rel}}([\underline{a},\underline{b}]\times[\underline{c},\underline{d}]) &= \langle \circ_{\underline{a}}\!\!-\!\!\bullet_{\underline{b}}\!\!-\!\!\bullet_{\underline{c}}\!\!-\!\!\circ_{\underline{d}} \rangle && \text{when } a > b \ge c > d \\
\mu_{\mathrm{ext}}([a,b]\times[\underline{c},\underline{d}]) &= \langle \circ_a\!\!-\!\!\bullet_b\!\!-\!\!\bullet_{\underline{c}}\!\!-\!\!\circ_{\underline{d}} \rangle && \text{when } a < b \text{ and } c > d
\end{aligned}
$$

taking $V_{-\infty} = 0$ and $V_{\underline{-\infty}} = 0$ whenever needed.

The measures $\mu_{\mathrm{ord}}, \mu_{\mathrm{rel}}$ are defined over the half-plane $\overline{\mathcal{H}}$, whereas μ_{ext} is defined over $\overline{\mathbf{R}}^2$.

Stability for $\mathsf{dgm}_{\mathrm{ord}}$, $\mathsf{dgm}_{\mathrm{rel}}$ and $\mathsf{dgm}_{\mathrm{ext}}$ may be proved individually for each diagram. Given two functions f, g which are δ-close in the supremum norm, there are inclusions

$$
\begin{aligned}
(X, f)^t &\subseteq (X, g)^{t+\delta} & (X, f)_t &\subseteq (X, g)_{t-\delta} \\
(X, g)^t &\subseteq (X, f)^{t+\delta} & (X, g)_t &\subseteq (X, f)_{t-\delta}
\end{aligned}
$$

which imply the box lemma (Lemma 5.26) for each measure. Since linear combinations of continuous functions are continuous, we can interpolate between f and g to satisfy the hypotheses required by the measure stability theorem (Theorem 5.29).

Remark 6.1 In the spirit of Theorem 3.37, one may treat the case where X is a locally compact polyhedron and f is proper. The exercise of locating the possible singularities of the three measures is left to the persistent reader.

References

1. Auslander, M.: Representation theory of Artin algebras. II. Commun. Algebr. **1**(4), 269–310 (1974)
2. Azumaya, G.: Corrections and supplementaries to my paper concerning Krull-Remak-Schmidt's theorem. Nagoya Math. J. **1**, 117–124 (1950)
3. Ulrich, B., Lesnick, M.: Induced matchings of barcodes and the algebraic stability of persistence. In: Proceedings of the 30th Annual Symposium on Computational Geometry (SoCG), pp. 355–364, Kyoto, Japan, June (2014)
4. Bubenik, P., de Silva, V., Jonathan, S.: Metrics for generalized persistence modules. Found. Comput. Math. **15**(6), 1501–1531 (2015)
5. Bubenik, P., Scott, J.A.: Categorification of persistent homology. Discret. Comput. Geom. **51**, 600–627 (2013)
6. Cagliari, F., Landi, C.: Finiteness of rank invariants of multidimensional persistent homology groups. Appl. Math. Lett. **24**(4), 516–518 (2011)
7. Carlsson, G.: Topology and data. Bull. Am. Math. Soc. **46**(2), 255–308 (2009)
8. Carlsson, G., de Silva, V.: Zigzag persistence. Found. Comput. Math. **10**(4), 367–405 (2010)
9. Carlsson, G., Singh, G., Zomorodian, A.: Computing multidimensional persistence. J. Comput. Geom. **1**(1), 72–100 (2010)
10. Carlsson, G., Zomorodian, A.: The theory of multidimensional persistence. In: Proceedings of the 23rd Annual Symposium on Computational Geometry (SoCG), Gyeongju, South Korea, June (2007)
11. Carrière, M., Oudot, M.S., Ovsjanikov, M.: Stable topological signatures for points on 3d shapes. In: Proceedings of the Symposium on Geometry Processing (2015)
12. Cerri, A., Fabio, B. D., Ferri, M., Frosini, P., Landi, C.: Betti numbers in multidimensional persistent homology are stable functions. Mathematical Methods in the Applied Sciences **36**, 1485–1648 (2013)
13. Chazal, F., Cohen-Steiner, D., Guibas, L.J., Mémoli, F., Oudot, S.Y.: Gromov-Hausdorff stable signatures for shapes using persistence. In: Proceedings of the Computer Graphics Forum (proc. SGP 2009), pp. 1393–1403 (2009)
14. Chazal, F., Cohen-Steiner, D., Glisse, M., Guibas, L., Oudot, S.: Proximity of persistence modules and their diagrams. In: Proceedings of the 25th Annual Symposium on Computational Geometry (SoCG), pp. 237–246 (2009)
15. Chazal, F., Cohen-Steiner, D., Glisse, M., Guibas, L.J., Oudot, S.: Proximity of Persistence Modules and their Diagrams. Research Report RR-6568, INRIA (2008)
16. Chazal, F., de Silva, V., Oudot, S.: Persistence stability of geometric complexes. Geom. Dedicata **173**, 193–214 (2014)
17. Chazal, F., Glisse, M., Labruère, C., Michel, B.: Convergence rates for persistence diagram estimation in topological data analysis. In: Proceedings of the 31st International Conference on Machine Learning (ICML-14), pp. 163–171 (2014)

F. Chazal et al., *The Structure and Stability of Persistence Modules*,
SpringerBriefs in Mathematics, DOI 10.1007/978-3-319-42545-0

18. Chazal, F., Guibas, L.J., Oudot, S.Y., Skraba, P.: Persistence-based clustering in Riemannian manifolds. J. ACM, **60**(6) (2013)
19. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. Discret. Comput. Geom. **37**(1), 103–120 (2007)
20. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Extending persistence using Poincaré and Lefschetz duality. Found. Comput. Math. **9**, 79–103 (2008)
21. Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Mileyko, Y.: Lipschitz functions have Lp-stable persistence. Found. Comput. Math. **10**(2), 127–139 (2010)
22. Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Morozov, D.: Persistent homology for kernels, images, and cokernels. In: Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1011–1020 (2009)
23. Cohen-Steiner, D., Edelsbrunner, H., Morozov, D.: Vines and vineyards by updating persistence in linear time. In: Proceedings of the 22nd Annual Symposium on Computational Geometry (SoCG), pp. 119–126 (2006)
24. Crawley-Boevey, W.: Personal communication, September (2012)
25. Crawley-Boevey, W.: Decomposition of pointwise finite-dimensional persistence modules. J. Algeb. Appl. **14**(5) (2015)
26. de Silva, V., Morozov, D., Vejdemo-Johansson, M.: Dualities in persistent (co)homology. Inverse Probl. **27**, 124003 (2011)
27. Derksen, H., Weyman, J.: Quiver representations. Not. Am. Math. Soc. **52**(2), 200–206 (February 2005)
28. Droz, J.M.: A subset of Euclidean space with large Vietoris–Rips homology. arXiv:1210.4097 [math.GT], (2012)
29. Edelsbrunner, H., Harer, J.L.: Computational Topology: An Introduction. American Mathematical Society, New York (2010)
30. Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. Discret. Comput. Geom. **28**, 511–533 (2002)
31. Edelsbrunner, H., Morozov, D.: Persistent homology: Theory and practice. In: Proceedings of the European Congress of Mathematics, pp. 31–50. European Mathematical Society (2012)
32. Eilenberg, S., Steenrod, N.E.: Foundations of Algebraic Topology. Princeton University Press, Princeton (1952)
33. Ellis, G., King, S.: Persistent homology of groups. J. Group Theory **14**(4), 575–587 (2011)
34. Frosini, P.: Discrete computation of size functions. J. Comb. Inf. Syst. Sci. **17**(3), 232–250 (1992)
35. Gabriel, P.: Unzerlegbare Darstellungen I. Manuscripta Mathematica **6**, 71–103 (1972)
36. Ghrist, Robert: Barcodes: the persistent topology of data. Bull. Am. Math. Soc. **45**(1), 61–75 (2008)
37. Ghrist, R.: Elementary Applied Topology. CreateSpace Independent Publishing Platform, London (September 2014)
38. Hausmann, J.-C.: On the Vietoris-Rips complexes and a cohomology theory for metric spaces. Prospects in Topology. Annals of Mathematical Studies, vol. 138. pp. 175–188. Princeton University Press, Princeton, New Jersey (1995)
39. Kaczynski, T., Mischaikow, K., Mrozek, M.: Computational Homology. Applied Mathematical Sciences. vol. 157, Springer, Berlin (2004)
40. Latschev, J.: Vietoris-Rips complexes of metric spaces near a closed Riemannian manifold. Archiv der Mathematik **77**, 522–528 (2001)
41. Lesnick, M.: Personal communication. (July 2012)
42. Lesnick, M.: The theory of the interleaving distance on multidimensional persistence modules. Found. Comput. Math. **15**(3), 613–650 (2015)
43. Mac Lane, S.: Categories for the Working Mathematician. 2nd edn. Springer, Berlin (1998)
44. Oudot, S.Y.: Persistence Theory: from quiver representations to data analysis. Mathematical Surveys and Monographs. vol. 209, American Mathematical Society, New York (2015)
45. Pakianathan, J., Winfree, T.: Threshold complexes and connections to number theory. Turk. J. Math. **37**, 511–539 (2013)

46. Robins, V.: Towards computing homology from finite approximations. Topol. Proc. **24**, 503–532 (1999)
47. Tachikawa, H., Michael Ringel, C.: Qf-3 rings. Journal für die reine und angewandte Mathematik (Crelle's Journal). **1975**(272), 49–72 (1975)
48. Webb, Cary: Decomposition of graded modules. Proc. Am. Math. Soc. **94**(4), 565–571 (1985)
49. Zomorodian, A.: Topology for Computing. Cambridge Monographs on Applied and Computational Mathematics. vol. 16. Cambridge University Press, Cmabridge (2005)
50. Zomorodian, A., Carlsson, G.: Computing persistent homology. Discret. Comput. Geom. **33**(2), 249–274 (2005)

Index

F. Chazal et al., *The Structure and Stability of Persistence Modules*,
SpringerBriefs in Mathematics, DOI 10.1007/978-3-319-42545-0