ISBN 978-82-326-4718-7 (printed ver.)
ISBN 978-82-326-4719-4 (electronic ver.)
ISSN 1503-8181

Doctoral thesis

Håvard Bakke Bjerkevik

Doctoral theses at NTNU, 2020:185

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical
Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

NTNU

Doctoral theses at NTNU, 2020:185

Håvard Bakke Bjerkevik

Stability and Complexity in Multi-Parameter Persistence

NTNU
Norwegian University of
Science and Technology

Håvard Bakke Bjerkevik

Stability and Complexity in Multi-Parameter Persistence

Thesis for the Degree of Philosophiae Doctor

Trondheim, June 2020

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

ISBN 978-82-326-4718-7 (printed ver.)
ISBN 978-82-326-4719-4 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2020:185

Printed by NTNU Grafisk senter

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of Philosophiae Doctor (Ph.D.) at the Norwegian University of Science and Technology (NTNU) in Trondheim.

First, I would like to thank my advisor Gereon Quick and co-advisor Nils Baas for all your support. I have had a tendency to go my own way in terms of research interests, but you have never hesitated to give me advice, feedback or encouragement whenever I have needed it, and for that I am very grateful.

The TDA community has been a great place to take my first steps in the academic world. I have fond memories of Werewolf, bouldering, hikes and dinners at various conferences, and everyone seems eager to talk to young researchers and help us learn. In particular, I would like to thank Erik Rybakken and my collaborators Magnus Botnan, Michael Kerber, Ulrich Bauer, Benedikt Fluhr and Michael Lesnick. It has been great fun discussing with you, and I have learned a lot. Extra thanks go to Uli for letting me visit TU Munich for my abroad period.

I also want to thank my colleagues, friends and family. Life is good when you are surrounded by great people. Most of all, I would like to thank Barbara. It has been two fantastic years; may there be many more to come.

Contents

Chapter 1

Introduction

This thesis consists of three papers, as well as a short appendix. They all concern the theoretical properties of *persistence modules*, which are important objects in topological data analysis (TDA). In this chapter, we give a brief presentation of the key notions and results in TDA this thesis is about, before outlining the contributions of the papers.

1.1 Persistence modules

Let **vec** be the category of vector spaces over a fixed field k.

Definition 1. *Let* $\boldsymbol{P}$ *be the poset category associated to a poset* P*. A* P-persistence module *is a functor* $\boldsymbol{P} \to \boldsymbol{vec}$*.*

The associated category of a poset P is a poset which has an object for each element in P, a single morphism from a to b if $a \leq b$, and no morphism otherwise. In this thesis, 'module' means 'persistence module'.

To make the definition a little more explicit, a P-module M contains the following data: For each $p \in P$, there is a vector space M_p, and for each $p \leq q$, there is a morphism $M_{p \to q} : M_p \to M_q$ of vector spaces. This respects composition, so $M_{q \to r} \circ M_{p \to q} = M_{p \to r}$.

In the classical setting, the poset is $\mathbb{R}$, $\mathbb{Z}$ or a subset of $\mathbb{R}$ with the standard poset structure. We will call such modules *single-parameter*, as opposed to *multi-parameter* modules, for which the underlying poset is $\mathbb{R}^n$ or $\mathbb{Z}^n$ for some $n \geq 2$. In this case the poset structure is given by $(a_1, \ldots, a_n) \leq (b_1, \ldots, b_n)$ if $a_i \leq b_i$ for all i.

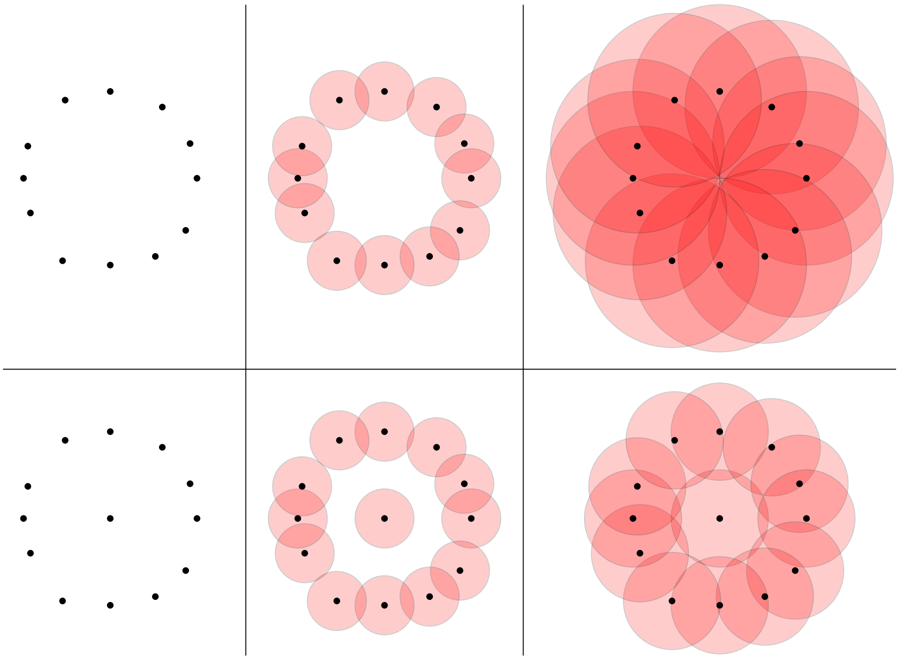


Figure 1.1: Top row, left to right: $B_0(P) = P$, $B_{0.34}(P)$ and $B_1(P)$. Bottom row, left to right: $B_0(Q) = Q$, $B_{0.34}(Q)$ and $B_{0.56}(Q)$.

1.2 Motivation from data sets

We will show a couple of examples to illustrate how persistence modules may arise from data sets, and why we sometimes need to consider multi-parameter modules.

Persistence modules can be seen as representing a data set by estimating its topological properties; the following simple example shows how this might happen in practice.

In the top left of Fig. 1.1, we have a set $P \subset \mathbb{R}^2$ of points that looks like it is sampled from a circle. One way to express this is to say that it should somehow have a generator of the first homology group H_1 (which we will take to be a vector space over some field, usually finite, throughout the thesis). Computing homology directly is useless, as it does not tell us anything more than the number of points. We solve this problem by computing H_1 of the union of balls centered at each point in P. In fact, we get a vector space

$M_r = H_1(B_r(P))$ for each radius $r \geq 0$, where

$$B_r(P) = \{x \in \mathbb{R}^2 \mid d(x, P) \leq r\},$$

$d(x, P)$ being the euclidean distance from x to the closest point in P. In addition there are morphisms $M_r \to M_s$ for $r \leq s$ induced by the inclusion $B_r(P) \hookrightarrow B_s(P)$. As we see, this is a single-parameter persistence module.

In the example above, we get a generator that is born when the balls enclose the hole in the middle at $r = 0.34$ and dies when the hole is filled in at $r = 1$. That is, we have a generator that persists over the interval $[0.34, 1)$. A common idea in TDA is that generators persisting over a long interval are significant, while short-lived generators are more likely to be the result of noise.

Now imagine that we add some outliers. For instance, let Q be as P, but with a single added point in the middle, as shown in the lower left corner of Fig. 1.1. Suddenly, our generator will persist for only a third of the time, and it will not be clear from the resulting persistence module that we have a circle with radius 1.

A way of fixing this is to introduce a second parameter (the first being the radius of the balls) that filters out points in regions with low density. For instance, at $(r, t) \in \mathbb{R}^2$, we might include only points in the plane with at least $-t$ points within a distance of r (the *multicover* bifiltration, see [11]). This way, we get a generator that survives inside a large region of $\mathbb{R}^2$, and again we are clearly detecting the circle. This simple example demonstrates how multi-parameter persistence can sometimes give us information single-parameter persistence cannot.

1.3 Stability and decomposition

A neat feature of single-parameter persistence is that all modules decompose nicely into *interval modules.*

Definition 2. *Let $J \subset \mathbb{R}$ be an interval. The interval module $\mathbb{I}^J$ is defined by $\mathbb{I}^J_a = k$ for $k \in J$ and $\mathbb{I}^J_a = 0$, otherwise, and $\mathbb{I}^J_{a \to b}$ the identity for $a \leq b \in J$.*

Theorem 3. *Any single-parameter persistence module M decomposes uniquely into a direct sum of interval modules, up to permutation and isomorphisms. That is, we can write $M = \bigoplus_{I \in B(M)} \mathbb{I}^J$ for a unique multiset $B(M)$.*

That modules decompose into interval modules is showed in [10], and that the decomposition is unique follows from Theorem 1 in [1]. We call $B(M)$

the *barcode* of M. We see that the decomposition theorem lets us pick a "persistent basis" of the vector spaces, so that we can think of the persistence module as consisting of independent generators surviving over certain intervals. In the example above where we constructed a persistence module from a point cloud P, the barcode would have a single interval $[0.34, 1)$, which fits with our intuition of H_1 having a single generator surviving over that interval.

Now we have described the whole single-parameter persistence pipeline: From the input, we construct a parametrized topological space, get a persistence module by computing homology and decompose it to obtain a barcode, which we consider the output.

As the input is usually assumed to carry some noise, we would like this process to be stable, meaning that a small change in the input does not lead to a large change in the output. To formalize this, one defines the interleaving distance d_I between persistence modules and the bottleneck distance d_B between barcodes. We refer to any of the papers in the thesis for the definition of d_I, and to either Paper I or Paper II for the definition of d_B.

In some cases, small changes in data sets lead to small changes in d_I. For instance, a function $f : \mathbb{R}^n \to \mathbb{R}$ induces a one-parameter persistence module by taking homology of sublevel sets, $F_a = H_i(f^{-1}(-\infty, a])$, and letting the morphisms $F_a \to F_b$ be the ones induced by inclusions $f^{-1}(-\infty, a] \hookrightarrow f^{-1}(-\infty, b]$ for $a \leq b$. Now it follows immediately from the definitions that $d_I(F, G) \leq ||f - g||_\infty$, where F is the module induced by f and G the one induced by g. Looking at sublevel sets of functions might seem arbitrary, but this is exactly what happens in the example above, where $B_r(P)$ is the sublevel set of the function $x \mapsto d(x, P)$ at r.

In other words, if one moves the points in P a limited distance, there will only be a limited change in the induced module, as measured by the interleaving distance. Together with the following theorem, this shows $d_B(B(F), B(G)) \leq ||f - g||_\infty$, which is a statement of stability of single-parameter persistence.

Theorem 4 ([8],[9])**.** *Let M and N be $\mathbb{R}$-modules. Then $d_I(M, N) = d_B(B(M), B(N))$.*

This is called the algebraic stability theorem (AST), and is one of the fundamental results in TDA. We sometimes abuse notation and write for instance $d_I = d_B$ meaning that $d_I(M, N) = d_B(B(M), B(N))$ holds for all M and N.

Some of this discussion carries over to the multi-parameter setting. We can decompose multi-parameter modules uniquely into indecomposables, and we can define the interleaving distance, which under sufficient conditions will be stable with respect to the input. In some cases, like in the example with the noisy circle, it will even fix some of the stability issues we have in the one-parameter setting, as it is less sensitive to outliers.

There is, however, the major problem that the path algebra of even a relatively small grid in $\mathbb{Z}^2$ is of wild representation type. This means that there is no nice way of describing all indecomposable multi-parameter modules. Because of this, the concept of barcodes and the bottleneck distance is not so useful anymore.

On the topic of stability, we should mention that in the world of applications, it can be far from obvious when data sets or persistence modules coming from data sets should be considered 'close'. We already gave one example showing this, where just one outlier changed the module drastically according to the interleaving distance. On the other hand, there might be applications where short intervals in a barcode can carry significant information about a data set, even though they are hardly noticed by the interleaving distance. Thus, the meaning of stability and the relevance of different metrics is not constant, and should be re-evaluated for each type of application. As the focus of the thesis is mainly on theoretical questions in TDA, we will not go into a long discussion about what happens when theory meets the messy world of applications, though this is of course a crucial part of a field called 'topological *data analysis*'.

1.4 Reeb graphs and zigzag modules

To explain some of the results in Paper I, we need to define *Reeb graphs* and *zigzag modules*.

Definition 5. *Let X be a finitely triangulable topological space, $g : X \to \mathbb{R}$ a continuous function, and $\sim_g$ the equivalence relation with equivalence classes $g^{-1}(\{x\})$. The Reeb graph of g is $X/\sim_g$ together with the map to $\mathbb{R}$ induced by g.*

This is the definition used in [2], though different restrictions on X and g appear in the literature. The idea is that given a space with a function to $\mathbb{R}$, the Reeb graph stores information about how level sets are connected,

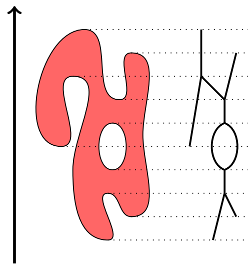


Figure 1.2: A space with a function to $\mathbb{R}$ given by the y-coordinate. Its Reeb graph is shown on the right.

hoping to simplify computation by ignoring higher dimensional structure. An illustration is given in Fig. 1.2.

Definition 6. *Let $\mathbb{ZZ}$ be the poset with $\mathbb{Z}$ as its underlying set, and with $a < b$ if a is even and $b = a+1$ or $b = a-1$. A zigzag module is a $\mathbb{ZZ}$-module.*

A zigzag module looks like an infinite string of vector spaces with morphisms in alternating directions:

$$\cdots \leftarrow M_{-2} \rightarrow M_{-1} \leftarrow M_0 \rightarrow M_1 \leftarrow M_2 \rightarrow \ldots$$

Given a Reeb graph (G, g), one can associate a vector space to any interval $[a, b]$ by taking the zeroth homology of $g^{-1}([a, b])$. We can define a 2-parameter module M by letting $M_{(-a,b)} = H_0(g^{-1}([a, b]))$ and $M_{(-a,b)\to(-c,d)}$ be the map induced by the inclusion $g^{-1}([a, b]) \subset g^{-1}([c, d])$. Strictly speaking, this gives us sets instead of vector spaces, but we simply take the sets to be bases of vector spaces, and let the linear maps be given by the functions on basis elements.

There is a similar construction for zigzag modules, but this time taking the colimit of the restriction of the module to the interval $[a, b]$, again giving a vector space for each interval. We get maps induced by the universal property of colimits. See [6] for the details.

In both cases, the 2-parameter modules we get decompose into modules of a very specific form called block modules. This allows us to talk about stability similarly to the single-parameter setting, and ask if $d_B = d_I$. In [6], $d_B \leq \frac{5}{2} d_I$ was shown, which was an improvement on results in [3]. In Paper I, we show $d_B = d_I$. This carries over to stability results for Reeb graphs and zigzag modules, and as a consequence of Paper I, these results cannot be improved.

1.5 Computation

For applications of TDA it is a fundamental question to which extent invariants like the barcode and distances like d_I are computable.

In single-parameter persistence, computing the interleaving distance can be done efficiently by decomposing the modules into interval modules and computing the bottleneck distance. In the multi-parameter case, this is not a possibility, and for years it was suspected that computing d_I is NP-hard, though no proof was known. The issue of wild representation type is often brought up when discussing invariants and distances in multi-parameter persistence, but it is not entirely clear why this should be a problem for computing d_I, especially since checking isomorphism can be done in polynomial time [7].

The question was finally settled in Paper III, where we show that not only is computing d_I NP-hard, but so is approximating it to a constant less than 3. Perhaps surprisingly, the modules appearing in the proof all decompose into indecomposables analogous to interval modules, allowing d_B to be computed efficiently, though this is of little help here, as $d_I \neq d_B$.

The upshot is that in practice, one needs other distances than d_I if one wants to compare persistence modules. Defining invariants and distances in multi-parameter persistence and finding algorithms for computing these is an active area of research in TDA.

Chapter 2

Overview of the Thesis

2.1 Paper I

The starting point for the first paper is the algebraic stability theorem. We know that $d_I(M,N) = d_B(B(M), B(N))$ for single-parameter modules, but what happens if we work with multi-parameter modules? We already made the point that general modules over $\mathbb{R}^n$ do not decompose nicely, so we have to make some strong assumptions on the modules we work with to be able to talk about stability. Still, it turns out that there are some useful consequences of this discussion, not least because some of the ideas turn out to be essential for the work in Paper II and III.

The main contribution of the paper is a proof technique to show stability. We first apply it to what we call rectangle decomposable modules, for which we show $d_B(B(M), B(N)) \leq (2n-1)d_I(M,N)$. For $n = 1$, this is AST. We also give an example to show that the bound cannot be improved for $n = 2$. This disproves a conjecture in an earlier version of [6], suggesting a generalization of AST to multi-parameter modules. This provides some more insight into the role that having just one parameter plays in AST. Not only is this what allows us to decompose modules nicely and define d_B, it is actually also needed to keep the geometry simple enough to prohibit cases where $d_B > d_I$.

We go on to use the proof technique to show a stability result for block decomposable modules, which by earlier work implies stability results for Reeb graphs and zigzag modules. This is an improvement on earlier results with weaker constants, and the constants we obtain are optimal, settling these problems. While rectangle decomposable modules do not appear naturally in the study of real data, Reeb graphs and zigzag modules do. Thus, despite the strong assumptions we need to put on the modules in order to formulate stability, it ends up being applicable to real data.

2.2 Paper II

The second paper discusses complexity bounds for deciding whether persistence modules of various types are δ-interleaved for a given $\delta > 0$, and also for deciding whether they are isomorphic, i.e. if they are 0-interleaved. The main contributions are an example of a family of persistence modules for which d_I is NP-hard to compute, and a polytime reduction from something we call CI problems to deciding whether a pair of two-parameter modules are 1-interleaved.

The example is admittedly a little artificial, as the modules are parametrized over a poset that is very different from $\mathbb{R}^n$, so the construction does not carry over to single- or multi-parameter persistence. Still, it was a first example of how to encode an NP-hard problem as an interleaving decision problem.

The latter result is effectively an application of ideas from Paper I to the problem of complexity. A CI problem is a compact description of potential interleavings between modules that decompose nicely. Instead of working directly with the modules, which can be messy, the relevant information is stored in a pair of matrices. This matrix description first appeared in Paper I, where it was applied both in the proof technique and in the mentioned counterexample to rectangle decomposable modules. In Paper II, we show that any CI problem can be encoded as a pair of modules in such a way that the CI problem is *solvable* if and only if the modules are 1-interleaved. This gives us a polytime reduction from the problem of deciding if CI problems are solvable to deciding if persistence modules are 1-interleaved. Thus, if one can show that CI problems are NP-hard, it follows that deciding existence of interleavings is NP-hard. We were, however, not able to show NP-hardness of CI problems.

A version of this paper without the appendices was published in the proceedings of The 34th International Symposium on Computational Geometry (SoCG 2018) [5].

2.3 Paper III

The main contribution of Paper III is a proof that deciding solvability of CI problems is NP-hard, which by the results in Paper II implies NP-hardness of deciding whether two multi-parameter persistence modules are δ-interleaved for some fixed δ.

In addition, we amend the encoding of CI problems as pairs of persistence modules so that any algorithm c-approximating the interleaving distance for a fixed $c < 3$ gives a polytime algorithm deciding solvability of CI problems, which is impossible unless $P = NP$. Thus, c-approximating the interleaving distance is NP-hard for $c < 3$. (1-approximation is the same as exact computation. The construction from Paper II would have given a constant of 2 instead of 3.)

We apply our proof strategy to show NP-hardness of some related problems in the latter parts of the paper. We state a theorem on one-sided stability where our technique is strong enough to settle all the cases not already known (up to the question of $P = NP$), except one. This case is deciding whether there is an injective morphism from a given module to another. We show that also this is NP-hard using a different proof, though still exploiting the idea of describing morphisms as matrices.

This paper was published in Foundations of Computational Mathematics (2019) [4].

2.4 Appendix

While the idea of representing interleavings as matrices play a prominent role in all the papers, there are some ideas that are lurking in the background but never made explicit. This is partly because they relate to problems we could not solve, and partly because the most elegant description of these ideas appear when we strip away the language of persistence modules and work directly on the level of graphs and matrices, which makes the usefulness to TDA less obvious. In the appendix, we formulate a conjecture that ties together stability and complexity even though the most natural way of phrasing it is not in terms of persistence modules.

Bibliography

[1] Gorô Azumaya. Corrections and supplementaries to my paper concerning Krull-Remak-Schmidt's theorem. *Nagoya Mathematical Journal*, 1:117–124, 1950.

[2] Ulrich Bauer, Xiaoyin Ge, and Yusu Wang. Measuring distance between reeb graphs. In *Proceedings of the thirtieth annual symposium on Computational geometry*, page 464. ACM, 2014.

[3] Ulrich Bauer, Elizabeth Munch, and Yusu Wang. Strong Equivalence of the Interleaving and Functional Distortion Metrics for Reeb Graphs. In *31st International Symposium on Computational Geometry (SoCG 2015)*, volume 34 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 461–475, 2015.

[4] Håvard Bakke Bjerkevik, Magnus Bakke Botnan, and Michael Kerber. Computing the interleaving distance is NP-hard. *Foundations of Computational Mathematics*, pages 1–35, 2019.

[5] Håvard Bakke Bjerkevik and Magnus Bakke Botnan. Computational Complexity of the Interleaving Distance. In *34th International Symposium on Computational Geometry (SoCG 2018)*, volume 99 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 13:1–13:15, 2018.

[6] Magnus Botnan and Michael Lesnick. Algebraic stability of zigzag persistence modules. *Algebraic & geometric topology*, 18(6):3133–3204, 2018.

[7] Peter A Brooksbank and Eugene M Luks. Testing isomorphism of modules. *Journal of Algebra*, 320(11):4020–4029, 2008.

[8] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J Guibas, and Steve Y Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 237–246. ACM, 2009.

[9] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.

[10] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and Its Applications*, 14(05):1550066, 2015.

[11] Donald R Sheehy. A multicover nerve for geometric inference. In *CCCG*, pages 309–314, 2012.

Paper I

Stability of Interval Decomposable Persistence Modules

Håvard Bakke Bjerkevik

Stability of Interval Decomposable Persistence Modules

Håvard Bakke Bjerkevik

Abstract

The algebraic stability theorem for $\mathbb{R}$-persistence modules is a fundamental result in topological data analysis. We present a stability theorem for n-dimensional rectangle decomposable persistence modules up to a constant $(2n-1)$ that is a generalization of the algebraic stability theorem, and also has connections to the complexity of calculating the interleaving distance. The proof given reduces to a new proof of the algebraic stability theorem with $n = 1$. We give an example to show that the bound cannot be improved for $n = 2$. We apply the same technique to prove stability results for zigzag modules and Reeb graphs, reducing the previously known bounds to a constant that cannot be improved, settling these questions.

1 Introduction

Persistent homology is a tool in topological data analysis used to determine the structure or shape of data sets. For example, given a point cloud $X \subset \mathbb{R}^n$ sampled from a subspace S of $\mathbb{R}^n$, we want to guess at the homology of S, which tells us something about how many "holes" S has in various dimensions. We can do this by defining $B(\epsilon)$ to be the union of the (open or closed) balls of radius ϵ centered at each point in X. Calculating homology, we get a group or vector space $H_n(B(\epsilon))$ for each $\epsilon \geq 0$, and the inclusions $B(\epsilon) \hookrightarrow B(\epsilon')$ induce morphisms $H_n(B(\epsilon)) \to H_n(B(\epsilon'))$ for $\epsilon \leq \epsilon'$. Such a collection of vector spaces and morphisms is called a *persistence module*, or $\mathbb{R}$-module, as the vector spaces are parametrized over $\mathbb{R}$. Under certain assumptions, we can decompose an $\mathbb{R}$-module into *interval modules* [15], which gives us a set of intervals uniquely determining the persistence module up to isomorphism. This set of intervals is the *barcode* of the persistence module. The intervals in the barcode are interpreted as corresponding to possible features of the space S, where one might interpret long intervals as more likely to describe actual features of S and short intervals as more likely to be the result of noise in the input data. In other words, we have an algorithm with a data set as input and a barcode as output. As data sets always carry a certain amount of noise, we would like this algorithm to be *stable* in the sense that a little change in the input data, or in the persistence modules, should not result in a big change in the barcode.

We measure the difference between persistence modules with the *interleaving distance* d_I, and the difference between barcodes with the *bottleneck distance* d_B. Proving stability then becomes a question of proving that the bottleneck distance is bounded by the interleaving distance, i.e. $d_B \leq C d_I$ for some constant C. Stability has been proved for persistence modules over $\mathbb{R}$ in [14, 12, 13, 5] in what is called the *algebraic stability theorem*, which implies the *isometry theorem* $d_I = d_B$.

Persistence modules can also be parametrized over other posets. A pair of filtrations $f, g : S \to \mathbb{R}$ of a topological space S gives rise to an $\mathbb{R}^2$-module which has a vector space V_p for each point in $p \in \mathbb{R}^2$ and linear maps $V_{(a,b)} \to V_{(c,d)}$ whenever $a \leq c$ and $b \leq d$, for instance by letting $V_{(a,b)}$ be $H_n(f^{-1}(-\infty, a) \cap g^{-1}(-\infty, b))$. Again, inclusions induce the linear maps on homology. With n filtrations instead of 2, we get an $\mathbb{R}^n$-module. Another example is *zigzag modules*, which are popular objects of study in topological data analysis [10, 20, 19]. These can arise from a sequence of subspaces $S_i \subset S$, where we also consider the intersections $S_i \cap S_{i+1}$ (or unions). In this case, we have

$$\cdots \subseteq S_{i-1} \supseteq S_{i-1} \cap S_i \subseteq S_i \supseteq S_i \cap S_{i+1} \subseteq \ldots,$$

which again gives rise to linear maps on homology. Defining interleavings and thus the interleaving distance is trickier than for $\mathbb{R}$-modules, but in fact one can do this by associating $\mathbb{R}^2$-modules called *block decomposable modules* to the zigzag modules. One can also associate block decomposable modules to *Reeb graphs*, which are of interest because of their ability to present geometrical information despite being relatively simple objects. See Section 3.

All these examples serve as motivation for why one would like to talk about stability for multi-parameter modules (that is, persistence modules parametrized over $\mathbb{R}^n$ for $n \geq 2$). Unfortunately, no isometry theorem is possible even for general $\mathbb{R}^2$-modules, because there is no nice decomposition theorem like in the one-parameter cases, meaning that d_B is not defined. The block decomposable modules, however, decompose nicely, and $d_B \leq \frac{5}{2} d_I$ has been shown for these [3]. This carries over to stability results for zigzag modules and Reeb graphs.

Our main contribution is a new method of proving stability for interval decomposable modules. We demonstrate several applications of this method. The first is Theorem 4.2:

Theorem. *Let $M = \bigoplus_{I \in B(M)} \mathbb{I}^I$ and $N = \bigoplus_{J \in B(N)} \mathbb{I}^J$ be rectangle decomposable $\mathbb{R}^n$-modules. If M and N are δ-interleaved, there exists a $(2n-1)\delta$-matching between $B(M)$ and $B(N)$.*

This is a generalization of the algebraic stability theorem for $\mathbb{R}$-modules, which is the case $n = 1$. For $n \geq 2$, the result is new. There already exist several proofs of the algebraic stability theorem, but our approach is different from the ones taken before, which allows this more general theorem, as well as the results below. Our method is combinatorial, which in our opinion reflects the true nature of the problem once some of the algebraic technicalities are stripped away. Also, our proof is fairly short in the case $n = 1$ compared to earlier proofs of the algebraic stability theorem. In Example 5.2, we construct rectangle decomposable modules M and N over $\mathbb{R}^2$ for which $d_I(M, N) = 1$ and $d_B(M, N) = 3$, disproving a conjecture made in an earlier version of [3] claiming that $d_B(M, N) = d_I(M, N)$ holds for all n-dimensional interval decomposable modules M and N whose barcodes only contain convex intervals. The example also shows that the bound in the theorem cannot be improved for $n = 2$. It is an open question if the bound can be improved for $n \geq 3$.

We do not know of any examples of rectangle decomposable modules arising naturally from real-world data sets. But as we discuss in Section 6, there is a strong link between the stability of these modules and the recent proof that calculating the interleaving distance between multi-parameter modules is NP-hard [9]. In particular, our way of viewing interleavings as pairs of matrices and our observation in Example 5.2 that the interleaving and bottleneck distances differ for rectangle decomposable modules served as inspiration for the approach used in [9]. The question of whether the hardness results can be strengthened is closely related to the question of whether Theorem 4.2 can be improved. Thus, even if rectangle decomposable modules never arise directly from data sets, the type of questions we consider can have an impact on practical applications.

Another reason why we give the proof in detail for rectangle decomposable modules instead of, say, block decomposable modules, is that this case demonstrates very well exactly when our method works and when it fails. The lesson to take home is that the method gives a bound $d_B \leq c d_I$ with a c that increases with the freedom we have in defining the intervals we consider. You need $2n$ coordinates to define an n-dimensional (hyper)rectangle, which gives a constant $2n - 1$ in the theorem.

Another application of our proof method gives Theorem 4.11:

Theorem. *Let M and N be δ-interleaved triangle decomposable modules. Then there is a δ-matching between $B(M)$ and $B(N)$.*

This is more immediately connected to practical applications. Theorem 4.11 implies $d_B \leq d_I$ for block decomposable modules, which is an improvement on the previous best known bound, $d_B \leq \frac{5}{2} d_I$. Since the opposite inequality $d_I \leq d_B$ holds trivially, our bound is the best possible. We discuss how stability results for zigzag modules and Reeb graphs follow in Section 3. The fact that our bound is optimal means that these stability problems are now settled.

We finish off Section 4 by showing stability for free modules.

We assume that all modules are pointwise finite dimensional (p.f.d.). In a previous version of this paper [8], we strengthened the theorems by removing this assumption.

2 Persistence modules, interleavings, and matchings

In this section we introduce some basic notation and definitions that we will use throughout the paper. Let k be a field that stays fixed throughout the text, and let **vec** be the category of finite dimensional vector spaces over k. We identify a poset with its poset category, which has the elements of the poset as objects, a single morphism $p \to q$ if $p \leq q$ and no morphism if $p \nleq q$.

Definition 2.1. *Let P be a poset category. A P-**persistence module** is a functor $P \to$ **vec**.*

If the choice of poset is obvious from the context, we usually write 'persistence module' or just 'module' instead of 'P-persistence module'.

For a persistence module M and $p \leq q \in P$, $M(p)$ is denoted by M_p and $M(p \to q)$ by $\phi_M(p,q)$. We refer to the morphisms $\phi_M(p,q)$ as the *internal morphisms* of M. M being a functor implies that $\phi_M(p,p) = id_{M_p}$, and that $\phi_M(q,r) \circ \phi_M(p,q) = \phi_M(p,r)$. Because the persistence modules are defined as functors, they automatically assemble into a category where the morphisms are natural transformations. This category is denoted by P-**mod**. Let $f : M \to N$ be a morphism between persistence modules. Such an f consists of a morphism associated to each $p \in P$, and these morphisms are denoted by f_p. Because f is a natural transformation, we have $\phi_N(p,q) \circ f_p = f_q \circ \phi_M(p,q)$ for all $p \leq q$.

Definition 2.2. *An **interval** is a subset $\varnothing \neq I \subseteq P$ that satisfies the following:*

- *If $p, q \in I$ and $p \leq r \leq q$, then $r \in I$.*
- *If $p, q \in I$, then there exist $p_1, p_2, \ldots, p_{2m} \in I$ for some $m \in \mathbb{N}$ such that $p \leq p_1 \geq p_2 \leq \cdots \geq p_{2m} \leq q$.*

We refer to the last point as the *connectivity axiom* for intervals.

Definition 2.3. *An **interval persistence module** or **interval module** is a persistence module M that satisfies the following: for some interval I, $M_p = k$ for $p \in I$ and $M_p = 0$ otherwise, and $\phi_M(p,q) = Id_k$ for points $p \leq q$ in I. We use the notation $\mathbb{I}^J$ for the interval module with J as its underlying interval.*

The definitions up to this point have been valid for all posets P, but we need some additional structure on P to get a notion of distance between persistence modules, which is essential to prove stability results. Since we will mostly be working with $\mathbb{R}^n$-persistence modules, we restrict ourselves to this case from now on. We define the poset structure on $\mathbb{R}^n$ by letting $(a_1, a_2, \ldots, a_n) \leq (b_1, b_2, \ldots, b_n)$ if and only if $a_i \leq b_i$ for $1 \leq i \leq n$. For $\epsilon \in \mathbb{R}$, we often abuse notation and write ϵ when we mean $(\epsilon, \epsilon, \ldots, \epsilon) \in \mathbb{R}^n$. We call an interval $I \subset \mathbb{R}^n$ *bounded* if it is bounded as a subset of $\mathbb{R}^n$ in the usual sense. That is, it is contained in a ball with finite radius.

Definition 2.4. *For $\epsilon \in [0, \infty)$, we define the **shift functor** $(\cdot)(\epsilon) : \mathbb{R}^n$**-mod** $\to \mathbb{R}^n$**-mod** by letting $M(\epsilon)$ be the persistence module with $M(\epsilon)_p = M_{p+\epsilon}$ and $\phi_{M(\epsilon)}(p,q) = \phi_M(p+\epsilon, q+\epsilon)$. For morphisms $f : M \to N$, we define $f(\epsilon) : M(\epsilon) \to N(\epsilon)$ by $f(\epsilon)_p = f_{p+\epsilon}$.*

We also define shift on intervals I by letting $I(\epsilon)$ be the interval for which $\mathbb{I}^{I(\epsilon)} = \mathbb{I}^I(\epsilon)$.

Define the morphism $\phi_{M,\epsilon} : M \to M(\epsilon)$ by $(\phi_{M,\epsilon})_p = \phi_M(p, p+\epsilon)$.

Definition 2.5. *An **ϵ-interleaving** between $\mathbb{R}^n$-modules M and N is a pair of morphisms $f : M \to N(\epsilon)$, $g : N \to M(\epsilon)$ such that $g(\epsilon) \circ f = \phi_{M,2\epsilon}$ and $f(\epsilon) \circ g = \phi_{N,2\epsilon}$.*

If there exists an ϵ-interleaving between M and N, then M and N are said to be ϵ-interleaved. An interleaving can be viewed as an 'approximate isomorphism', and a 0-interleaving is in fact an isomorphism. We call a module M *ϵ-significant* if $\phi_M(p, p+\epsilon) \neq 0$ for some p, and *ϵ-trivial* otherwise. M is 2ϵ-trivial if and only if it is ϵ-interleaved with the zero module. We call an interval I ϵ-significant if $\mathbb{I}^I$ is ϵ-significant, and ϵ-trivial otherwise.

Definition 2.6. *We define the **interleaving distance** d_I on persistence modules M and N by*

$$d_I(M,N) = \inf\{\epsilon \mid M \text{ and } N \text{ are } \epsilon\text{-interleaved}\}. \tag{1}$$

The interleaving distance intuitively measures how close the modules are to being isomorphic. The interleaving distance between two modules might be infinite, and the interleaving distance between two different, even non-isomorphic modules, might be zero. Apart from this, d_I satisfies the axioms for a metric, so d_I is an extended pseudometric.

Definition 2.7. *Suppose $M \cong \bigoplus_{I \in B} \mathbb{I}^I$ for a multiset*[1] *B of intervals. Then we call B the* ***barcode*** *of M, and write $B = B(M)$. We say that M is* ***interval decomposable****.*

Since the endomorphism ring of any interval module is isomorphic to k, it follows from Theorem 1 in [2] that if a persistence module M is interval decomposable, the decomposition is unique up to isomorphism. Thus the barcode is well-defined, even if we let M be a P-module for an arbitrary poset P. If M is a p.f.d. $\mathbb{R}$-module, it is interval decomposable [15], but this is not true for $\mathbb{R}$-modules or p.f.d. $\mathbb{R}^n$-modules in general. [21] gives an example showing the former, and the following is an example of a P-module for a poset P with four points that is not interval decomposable.

$$\begin{tikzcd} k \arrow[r, "\begin{pmatrix}1 & 0\end{pmatrix}"] & k^2 \arrow[r, "\begin{pmatrix}1 \\ 1\end{pmatrix}"] & k \\ & k \arrow[u, "\begin{pmatrix}0 & 1\end{pmatrix}"] & \end{tikzcd} \tag{2}$$

A corresponding $\mathbb{R}^2$-module that is not interval decomposable and is at most two-dimensional at each point can be constructed.

For multisets A, B, we define a *partial bijection* as a bijection $\sigma : A' \to B'$ for some subsets $A' \subset A$ and $B' \subset B$, and we write $\sigma : A \nrightarrow B$. We write coim $\sigma = A'$ and im $\sigma = B'$.

Definition 2.8. *Let A and B be multisets of intervals. An* ***ϵ-matching*** *between A and B is a partial bijection $\sigma : A \nrightarrow B$ such that*

- *all $I \in A \setminus$ coim σ are 2ϵ-trivial*
- *all $I \in B \setminus$ im σ are 2ϵ-trivial*
- *for all $I \in$ coim σ, $\mathbb{I}^I$ and $\mathbb{I}^{\sigma(I)}$ are ϵ-interleaved.*

If there is an ϵ-matching between $B(M)$ and $B(N)$ for persistence modules M and N, we say that M and N are *ϵ-matched*.

We have adopted this definition of ϵ-matching from [3], which differs from e.g. the one in [13], which allows two intervals I and J to be matched if $d_I(\mathbb{I}^I, \mathbb{I}^J) \leq \epsilon$ (or rather, this is equivalent to their definition). Conveniently, with the definition we have chosen, an ϵ-interleaving is easily constructed given an ϵ-matching. We feel that this is the more natural definition for this paper, as several of our results are phrased as statements about matchings and interleavings, and the interleaving distance might not come into the picture at all. The other definition is perhaps more natural in the context of 'persistence diagrams', where intervals are identified with points in a diagram, and the interleaving distance between the corresponding modules is simply the distance between the points. This is irrelevant to us, however, as we never consider persistence diagrams.

We can also define ϵ-matchings in the context of graph theory. A *matching in a graph* is a set of edges in the graph without common vertices, and a matching is said to *cover* a set S of vertices if all elements in S are adjacent to an edge in the matching. Let G_ϵ be the bipartite graph on $A \sqcup B$ with an edge between $I \in A$ and $J \in B$ if $\mathbb{I}^I$ and $\mathbb{I}^J$ are ϵ-interleaved. Then an ϵ-matching between A and B is a matching in G_ϵ such that the set of 2ϵ-significant intervals in $A \sqcup B$ is covered.

Definition 2.9. *The* ***bottleneck distance*** *d_B is defined by*

$$d_B(M, N) = \inf\{\epsilon \mid M \text{ and } N \text{ are } \epsilon\text{-matched}\} \tag{3}$$

for any interval decomposable M and N.

We might abuse notation and talk about $d_B(C, D)$, where C and D are barcodes.

[1] We will not be rigorous in our treatment of multisets. A multiset may contain multiple copies of one element, but we will assume that we have some way of separating the copies, so that we can treat the multiset as a set. If e.g. I and J are intervals in a multiset and we say that $I \neq J$, we mean that they are "different" elements of the multiset, not that they are different intervals.

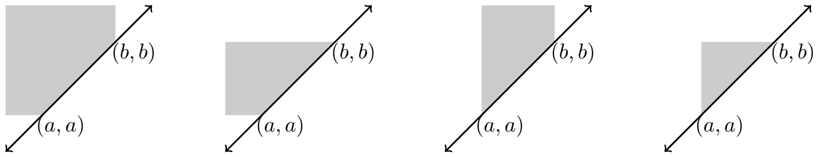


Figure 1: The intervals $[a,b]_{\mathrm{BL}}, [a,b)_{\mathrm{BL}}, (a,b]_{\mathrm{BL}}$ and $(a,b)_{\mathrm{BL}}$.

3 Zigzag modules and Reeb graphs

In this section we will give some intuition for how block decomposable modules relate to Reeb graphs and zigzag modules. We refer to [3] for a more detailed and rigorous treatment.

When explaining the connection to Reeb graphs and zigzag modules, it is more convenient to flip one of the axes in $\mathbb{R}^2$, so that we work with $\mathbb{R}^{\mathrm{op}} \times \mathbb{R}$ instead. This way, $(a,b) \leq (c,d)$ iff $c \leq a$ and $b \leq d$, or, equivalently, if $(a,b) \subseteq (c,d)$ as intervals, assuming $a < b$. Let $\mathbb{U} = \{(a,b) \in \mathbb{R}^{\mathrm{op}} \times \mathbb{R} \mid a \leq b\}$.

Definition 3.1. *An interval decomposable $\mathbb{R}^{op} \times \mathbb{R}$-module is called **block decomposable** if its barcode only contains intervals of the following types:*

- $[a,b]_{BL} = \{(c,d) \in \mathbb{U} \mid c \leq b, d \geq a\}$
- $[a,b)_{BL} = \{(c,d) \in \mathbb{U} \mid a \leq d < b\}$
- $(a,b]_{BL} = \{(c,d) \in \mathbb{U} \mid a < c \leq b\}$
- $(a,b)_{BL} = \{(c,d) \in \mathbb{U} \mid c > a, d < b\}$

We call these intervals *blocks*. Each interval intersects the diagonal in an $\mathbb{R}$-interval that is open, closed or half-open one way or the other depending on the type of the block.

3.1 Reeb graphs

There have been proposed several distances on Reeb graphs; see [4] for a summary, as well as references to various applications. The interleaving distance we consider was introduced in [16].

A Reeb graph is a topological graph G together with a continuous function $\gamma : G \to \mathbb{R}$ such that the level sets of γ are discrete. Let $S(\gamma) = \mathbb{R}^{\mathrm{op}} \times \mathbb{R} \to \mathbf{Set}$ be the functor sending (a,b) to the set of connected components of $\gamma^{-1}(a,b)$ and $S((a,b) \to (c,d))$ be induced by the inclusion $\gamma^{-1}(a,b) \subseteq \gamma^{-1}(c,d)$ for $c \leq a \leq b \leq d$. In Figure 2, γ is projection to a horizontal axis. Above the graph, the functor $S(\gamma)$ is shown, the shade of grey at (a,b) determined by the size of $S(\gamma)_{(a,b)}$.

Given two Reeb graphs (G_1,γ_1) and (G_2,γ_2), we get two functors $S(\gamma_1)$ and $S(\gamma_2)$, and we can talk about interleavings and interleaving distance by adjusting the definitions in the previous section. It turns out that this interleaving distance is at least as big as the one we get by replacing $S(\gamma_1)$ and $S(\gamma_2)$ by corresponding block decomposable modules M_{γ_1} and M_{γ_2}. In Figure 2, the blocks comprising this block decomposable module are exactly what you would guess by looking at the figure. In [3], $d_B(M,N) \leq \frac{5}{2} d_I(M,N)$ is proved for such modules; with Theorem 4.14, we have $d_B(M,N) = d_I(M,N)$.

There is also a barcode $L_0(\gamma)$ of $\mathbb{R}$-intervals (the level set persistence diagram [11]) associated to a Reeb graph (G,γ), which we can think of as arising from the intersection of $S(\gamma)$ with the diagonal $x = y$. This barcode is the same as $B(M_\gamma)$, except that $(a,b]_{BL}$ is replaced by $(a,b]$, and so on. It is not too hard to see that $d_B(L_0(\gamma_1), L_0(\gamma_2)) \leq 2 d_B(M_{\gamma_1}, M_{\gamma_2})$.[2] Altogether, this gives

$$\begin{aligned} d_B(L_0(\gamma_1), L_0(\gamma_2)) &\leq 2 d_B(M_{\gamma_1}, M_{\gamma_2}) \\ &= 2 d_I(M_{\gamma_1}, M_{\gamma_2}) \\ &\leq 2 d_I(S(\gamma_1), S(\gamma_2)). \end{aligned}$$

In other words:

[2] The reason for the constant 2 is that $(a,b)_{BL}$ is $(b-a)/4$-trivial, while (a,b) is not ϵ-trivial for $\epsilon < (b-a)/2$.

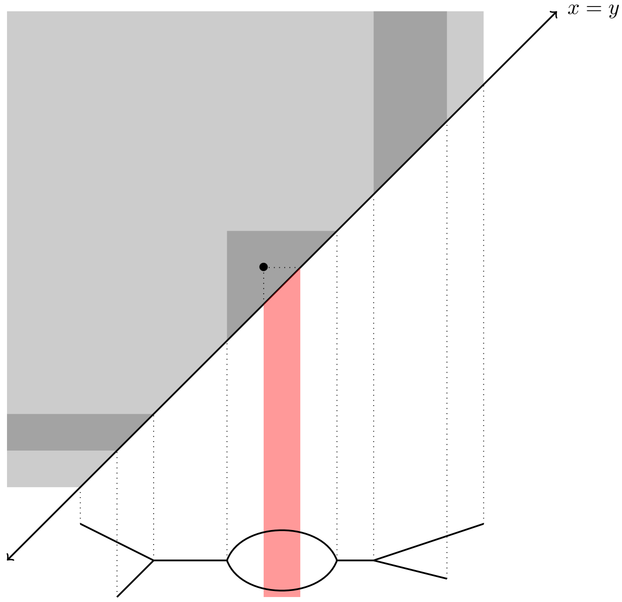


Figure 2: A Reeb graph (G, γ) with $S(\gamma)$ above. Evaluating $S(\gamma)$ at the point shown, we get the intersection of G with the red strip, which has two connected components.

Theorem 3.2. *For Reeb graphs γ_1, γ_2, the inequality $d_B(L_0(\gamma_1), L_0(\gamma_2)) \leq 2d_I(S(\gamma_1), S(\gamma_2))$ holds.*

Thus an easily computed bottleneck distance gives a lower bound for the interleaving distance between Reeb graphs. This improves the result in [3], which was again an improvement on [6], by lowering the constant in the inequality from 5 to 2, and this cannot be improved.

3.2 Zigzag modules

A zigzag module is a module over $\mathbb{ZZ} = \{(a, b) \in \mathbb{Z}^2 \mid a = b \vee a = b + 1\}$ taken as a sub-poset of $\mathbb{R}^{\mathrm{op}} \times \mathbb{R}$. Let $\mathbb{ZZ}|_{(a,b)}$ be the sub-poset of $\mathbb{ZZ}$ containing the elements $\{(c, d) \in \mathbb{ZZ} \mid a \leq c, d \leq b\}$. A zigzag module M gives rise to a block decomposable module M_{BL} defined by letting $M_{BL}(a, b)$ be the colimit of the restriction of M to $\mathbb{ZZ}|_{(a,b)}$. $M_{BL}((a, b) \to (c, d))$ is defined to be the induced morphism we get by the universal property of colimits for $(a, b) \leq (c, d)$. (This definition is given in [3], but something very similar is described in the discussions of pyramids in [11] and [7].) This way, we can define interleaving and bottleneck distance between zigzag modules by letting $d_I(M, N) = d_I(M_{BL}, N_{BL})$ and $d_B(M, N) = d_B(M_{BL}, N_{BL})$. Thus Theorem 4.14 holds if we replace 'block decomposable modules' by 'zigzag modules':

Theorem 3.3. *Let M and N be zigzag modules. If M and N are δ-interleaved, there exists a δ-matching between $B(M)$ and $B(N)$.*

This implies an isometry theorem for zigzag modules: $d_I(M, N) = d_B(M, N)$.

4 Higher-dimensional stability

The algebraic stability theorem for $\mathbb{R}$-modules states that an ϵ-interleaving between $\mathbb{R}$-modules M and N induces an ϵ-matching between $B(M)$ and $B(N)$, implying $d_I(M, N) = d_B(M, N)$, the isometry theorem.

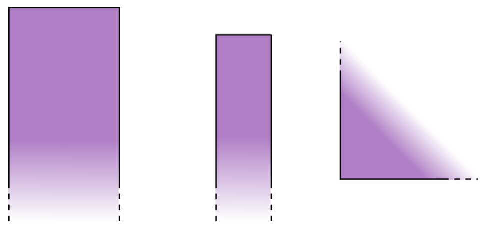


Figure 3: Three rectangles, where the left and middle rectangles are of the same type (unbounded downwards), while the last is of a different type (unbounded upwards and to the right). Assuming that it contains its boundary, the rightmost rectangle is also an example of a free interval, which we will define in a later section.

The main purpose of this paper is to find out when similar results for $\mathbb{R}^n$-modules hold. Our first result, Theorem 4.2, is a generalization of the algebraic stability theorem for $\mathbb{R}$-modules. Variations of the algebraic stability theorem have been proved several times already [14, 12, 13, 5], but this is a new proof with ideas that are applicable to more than just $\mathbb{R}$-modules.

4.1 Rectangle decomposable modules

For any interval $I \subset \mathbb{R}^n$, we let its projection on the i'th coordinate be denoted by I_i.

Definition 4.1. *A* ***rectangle*** *is an interval of the form* $R = R_1 \times R_2 \times \cdots \times R_n$.

Two rectangles R and S are of the *same type* if $R_i \setminus S_i$ and $S_i \setminus R_i$ are bounded for every i. For $n = 1$, we have four types of rectangles:

- intervals of finite length
- intervals of the form (a, ∞) or $[a, \infty)$
- intervals of the form $(-\infty, a)$ or $(-\infty, a]$
- $(-\infty, \infty)$,

for some $a \in \mathbb{R}$. We see that for $n \geq 1$, rectangles R and S are of the same type if R_i and S_i are of the same type for all $1 \leq i \leq n$. Examples of 2-dimensional rectangles are given in Figure 3.

In [13], *decorated numbers* were introduced. These are endpoints of intervals 'decorated' with a plus or minus sign depending on whether the endpoints are included in the interval or not. Let $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. A decorated number is of the form a^+ or a^-, where $a \in \overline{\mathbb{R}}$.[3] The notation is as follows for $a, b \in \overline{\mathbb{R}}$:

- $I = (a^+, b^+)$ if $I = (a, b]$
- $I = (a^+, b^-)$ if $I = (a, b)$
- $I = (a^-, b^+)$ if $I = [a, b]$
- $I = (a^-, b^-)$ if $I = [a, b)$.

We define decorated points in n dimensions for $n \geq 1$ as tuples $a = (a_1, a_2, \dots, a_n)$, where all the a_i's are decorated numbers. For an n-dimensional rectangle R and decorated points $(a_1, a_2, \dots, a_n)$ and $(b_1, b_2, \dots, b_n)$, we write $R = ((a_1, a_2, \dots, a_n), (b_1, b_2, \dots, b_n))$ if $R_i = (a_i, b_i)$ for all i. We define $\min_R$ and $\max_R$ as the decorated points for which $R = (\min_R, \max_R)$. We write a^* for decorated numbers with unknown 'decoration', so a^* is either a^+ or a^-.

[3]The decorated numbers $-\infty^-$ and ∞^+ are never used, as no interval contains points at infinity, but it does not matter whether we include these two points in the definition.

There is a total order on the decorated numbers given by $a^* < b^*$ for $a < b$, and $a^- < a^+$ for all $a, b \in \overline{\mathbb{R}}$. This induces a poset structure on decorated n-dimensional points given by $(a_1, a_2, \ldots, a_n) \leq (b_1, b_2, \ldots, b_n)$ if $a_i \leq b_i$ for all i. We can also add decorated numbers and real numbers by letting $a^+ + x = (a+x)^+$ and $a^- + x = (a+x)^-$ for $a \in \overline{\mathbb{R}}$, $x \in \mathbb{R}$. We add n-dimensional decorated points and n-tuples of real numbers coordinatewise.

If M is an interval decomposable $\mathbb{R}^n$-module and all $I \in B(M)$ are rectangles, M is *rectangle decomposable.*

Our goal is to prove the following theorem:

Theorem 4.2. *Let $M = \bigoplus_{I \in B(M)} \mathbb{I}^I$ and $N = \bigoplus_{J \in B(N)} \mathbb{I}^J$ be rectangle decomposable $\mathbb{R}^n$-modules. If M and N are δ-interleaved, there exists a $(2n-1)\delta$-matching between $B(M)$ and $B(N)$.*

The inequality $d_B(M, N) \leq (2n-1)d_I(M, N)$ for rectangle decomposable modules M and N immediately follows.

Fix $0 \leq \delta \in \mathbb{R}$. Assume that M and N are δ-interleaved, with interleaving morphisms $f : M \to N(\delta)$ and $g : N \to M(\delta)$. Recall that this means that $g(\delta) \circ f = \phi_{M,2\delta}$ and $f(\delta) \circ g = \phi_{N,2\delta}$. For any $I \in B(M)$, we have a canonical injection $\mathbb{I}^I \xrightarrow{\iota_I} M$ and projection $M \xrightarrow{\pi_I} \mathbb{I}^I$, and likewise, we have canonical morphisms $\mathbb{I}^J \xrightarrow{\iota_J} N$ and $N \xrightarrow{\pi_J} \mathbb{I}^J$ for $J \in B(N)$. We define

$$\begin{aligned} f_{I,J} &= \pi_J(\delta) \circ f \circ \iota_I : \mathbb{I}^I \to \mathbb{I}^J(\delta) \\ g_{J,I} &= \pi_I(\delta) \circ g \circ \iota_J : \mathbb{I}^J \to \mathbb{I}^I(\delta). \end{aligned} \tag{4}$$

We prove the theorem by a mix of combinatorial and geometric arguments. First we show that it is enough to prove the theorem under the assumption that all the rectangles in $B(M)$ and $B(N)$ are of the same type. Then we define a real-valued function α on the set of rectangles which in a sense measures, in the case $n = 2$, how far 'up and to the right' a rectangle is. There is a preorder $\leq_\alpha$ associated to α. The idea behind $\leq_\alpha$ is that if there is a nonzero morphism $\chi : \mathbb{I}^I \to \mathbb{I}^J(\epsilon)$ and $I \leq_\alpha J$, then I and J have to be close to each other. Finding pairs of intervals in $B(M)$ and $B(N)$ that are close is exactly what we need to construct a $(2n-1)\delta$-matching. Lemmas 4.6 and 4.7 say that such morphisms behave nicely in a precise sense that we will exploit when we prove Lemma 4.8. If we remove the conditions mentioning $\leq_\alpha$, Lemmas 4.6 and 4.7 are not even close to being true, so one of the main points in the proof of Lemma 4.8 is that we must exclude the cases that are not covered by Lemmas 4.6 and 4.7. We do this by proving that a certain matrix is upper triangular, where the 'bad cases' correspond to the elements above the diagonal and the 'good cases' correspond to elements on and below the diagonal.

Lemma 4.8 is what ties together the geometric and combinatorial parts of the proof of Theorem 4.2. While we prove Lemma 4.8 by geometric arguments, by Hall's marriage theorem the lemma is equivalent to a statement about matchings between $B(M)$ and $B(N)$. We have to do some combinatorics to get exactly the statement we need, namely that there is a $(2n-1)\delta$-matching between $B(M)$ and $B(N)$, and we do this after stating Lemma 4.8.

We begin by describing morphisms between rectangle modules.

Lemma 4.3. *Let $\chi : \mathbb{I}^I \to \mathbb{I}^J$ be a morphism between interval modules. Suppose $A = I \cap J$ is an interval. Then, for all $a, b \in A$, $\chi_a = \chi_b$ as k-endomorphisms.*

Proof. Suppose $a \leq b$ and $a, b \in A$. Then $\chi_b \circ \phi_{\mathbb{I}^I}(a, b) = \phi_{\mathbb{I}^J}(a, b) \circ \chi_a$. Since the ϕ-morphisms are identities, we get $\chi_a = \chi_b$ as k-endomorphisms. By the connectivity axiom for intervals, the equality extends to all elements in A. $\square$

Since the intersection of two rectangles is either empty or a rectangle, we can describe a morphism between two rectangle modules uniquely as a k-endomorphism if their underlying rectangles intersect. A k-endomorphism, in turn, is simply multiplication by a constant. Note that we could have relaxed the assumptions in the proof above and assumed that a is in I instead of in A, and still have gotten $\chi_a = \chi_b$. In particular, this means that if $0 \neq \chi : \mathbb{I}^I \to \mathbb{I}^J$, and I and J are rectangles, then $\min_{J_i} \leq \min_{I_i}$ for all i, which gives $\min_J \leq \min_I$. Similarly, $\max_J \leq \max_I$, and one can also see that $\min_I < \max_J$ must hold, or else $I \cap J = \varnothing$. We summarize these observations as a corollary of Lemma 4.3:

Corollary 4.4. *Let R and S be rectangles, and let $\chi : \mathbb{I}^R \to \mathbb{I}^S$ be a nonzero morphism. Then $min_S \leq min_R$ and $max_S \leq max_R$.*

This will come in handy when we prove Lemmas 4.5, 4.6, and 4.7.

We define a function $w : (B(M) \times B(N)) \sqcup (B(N) \times B(M)) \to k$ by letting $w(I, J) = x$ if $f_{I,J}$ is given by multiplication by x, and $w(I, J) = 0$ if $f_{I,J}$ is the zero morphism. $w(J, I)$ is given by $g_{J,I}$ in the same way.

With the definition of w, it is starting to become clear how combinatorics comes into the picture. We can now construct a bipartite weighted directed graph on $B(M) \sqcup B(N)$ by letting $w(I, J)$ be the weight of the edge from I to J. The reader is invited to keep this picture in mind, as a lot of what we do in the rest of the proof can be interpreted as statements about the structure of this graph.

The following lemma allows us to break up the problem and focus on the components of M and N with the same types separately.

Lemma 4.5. *Let R and T be rectangles of the same type, and S be a rectangle of a different type. Then $\psi\chi = 0$ for any pair $\chi : \mathbb{I}^R \to \mathbb{I}^S$, $\psi : \mathbb{I}^S \to \mathbb{I}^T$ of morphisms.*

Proof. Suppose $\psi, \chi \neq 0$. By Corollary 4.4, $\min_R \geq \min_S \geq \min_T$ and $\max_R \geq \max_S \geq \max_T$. We get $\min_{R_i} \geq \min_{S_i} \geq \min_{T_i}$ and $\max_{R_i} \geq \max_{S_i} \geq \max_{T_i}$ for all i, and it follows that if R and T are of the same type, then S is of the same type as R and T. □

Let $f' : M \to N(\delta)$ be defined by $f'_{I,J} = f_{I,J}$ for $I \in B(M)$ and $J \in B(N)$ if I and J are of the same type, and $f'_{I,J} = 0$ if they are not, and let $g' : N \to M(\delta)$ be defined analogously. Here f' and g' are assembled from $f'_{I,J}$ and $g'_{J,I}$ the same way f and g are from $f_{I,J}$ and $g_{J,I}$. Suppose $I, I' \in B(M)$. Then we have

$$\sum_{J \in B(N)} g_{J,I'}(\delta) f_{I,J} = \sum_{J \in B(N)} g'_{J,I'}(\delta) f'_{I,J}. \tag{5}$$

When I and I' are of different types, the left side is zero because f and g are δ-interleaving morphisms, and all the summands on the right side are zero by definition of f' and g'. When I and I' are of the same type, the equality follows from Lemma 4.5. This means that $g'(\delta)f' = g(\delta)f$. We also have $f'(\delta)g' = f(\delta)g$, so f' and g' are δ-interleaving morphisms. In particular, f' and g' are δ-interleaving morphisms when restricted to the components of M and N of a fixed type. If we can show that f' and g' induce a $(2n-1)\delta$-matching on each of the mentioned components, we will have proved Theorem 4.2. In other words, we have reduced the problem to the case where all the intervals in $B(M)$ and $B(N)$ are of the same type.

For a decorated number a^*, let $u(a^*) = a$ if $a \neq \pm\infty$ and $u(a^*) = 0$ otherwise. Let $a = (a_1, a_2, \ldots, a_n)$ be a decorated point. We define $P(a)$ to be the number of the decorated numbers a_i decorated with $+$, and we also define $\alpha(a) = \sum_{1 \leq i \leq n} u(a_i)$. What we really want to look at are rectangles and not decorated points by themselves, so we define $P(R) = P(\min_R) + P(\max_R)$ and $\alpha(R) = \alpha(\min_R) + \alpha(\max_R)$ for any rectangle R. Define an order $\leq_\alpha$ on decorated points given by $a \leq_\alpha b$ if either

- $\alpha(a) < \alpha(b)$, or
- $\alpha(a) = \alpha(b)$ and $P(a) \leq P(b)$

This defines a preorder. In other words, it is transitive ($R \leq_\alpha S \leq_\alpha T$ implies $R \leq_\alpha T$) and reflexive ($R \leq_\alpha R$ for all R). We write $R <_\alpha S$ if $R \leq_\alpha S$ and not $R \geq_\alpha S$.

The order $\leq_\alpha$ is one of the most important ingredients in the proof. The point is that if there is a nonzero morphism from $\mathbb{I}^R$ to $\mathbb{I}^S(\epsilon)$ and $R \leq_\alpha S$, then R and S have to be close to each other. If $\epsilon = 0$, R and S actually have to be equal. This 'closeness property' is expressed in Lemma 4.6, and is also exploited in Lemma 4.7. Finally, in the proof of Lemma 4.8, we make sure that we only have to deal with morphisms $g_{J,I'}(\delta) \circ f_{I,J}$ for $I \leq_\alpha I'$ and not $I >_\alpha I'$, so that our lemmas can be applied.

In Figure 4 we see two rectangles $R = (0, 4) \times (0, 4)$ and $S = (2, 5) \times (2, 5)$. There is no nonzero morphism from $\mathbb{I}^R$ to $\mathbb{I}^S$ or $\mathbb{I}^{S(1)}$, because $\min_R < \min_{S(\epsilon)}$ for all $\epsilon < 2$. This is connected to the fact that $\alpha(R) = 8 < 14 = \alpha(S)$, which can be interpreted to mean that R is 'further down and to the left' than S. The point of including $P(\alpha)$ in the definition of α is that e.g. $(a, b]$ is a tiny bit 'further to the right' than $[a, b)$, and this is a subtlety that P recognizes, and that matters in the proofs of Lemmas 4.6 and 4.7.

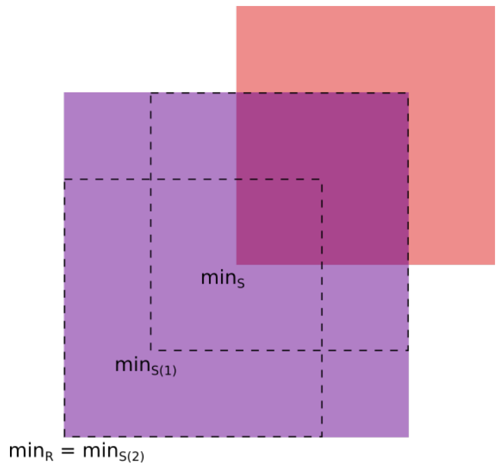


Figure 4: Rectangles $R = (0,4) \times (0,4)$ (purple), $S = (2,5) \times (2,5)$ (pink), $S(1) = (1,4) \times (1,4)$ (dotted border), and $S(2) = (0,3) \times (0,3)$ (dotted border).

Lemma 4.6. *Let R, S, and T be rectangles of the same type with $R \leq_\alpha T$. Suppose there are nonzero morphisms $\chi : \mathbb{I}^R \to \mathbb{I}^S(\epsilon)$ and $\psi : \mathbb{I}^S \to \mathbb{I}^T(\epsilon)$. Then $\mathbb{I}^S$ is $(2n-1)\epsilon$-interleaved with either $\mathbb{I}^R$ or $\mathbb{I}^T$.*

Proof. Since $\chi \neq 0$, we have

- $\min_S \leq \min_R + \epsilon$
- $\max_S \leq \max_R + \epsilon$.

This follows from Corollary 4.4.

Suppose $\mathbb{I}^R$ and $\mathbb{I}^S$ are not $(2n-1)\epsilon$-interleaved. Then either $\min_S + (2n-1)\epsilon \ngeq \min_R$ or $\max_S + (2n-1)\epsilon \ngeq \max_R$; let us assume the latter. (The former is similar.) In this case, there is an m such that $\max_{S_m} < \max_{R_m} - (2n-1)\epsilon$. For $i \neq m$, we have $\max_{S_i} \leq \max_{R_i} + \epsilon$ by the second bullet point. We get

$$\begin{aligned} \sum_{1 \leq i \leq n} u(\max_{S_i}) &\leq \left(\sum_{1 \leq i \leq n} u(\max_{R_i}) \right) - (2n-1)\epsilon + (n-1)\epsilon \\ &= \left(\sum_{1 \leq i \leq n} u(\max_{R_i}) \right) - n\epsilon. \end{aligned} \tag{6}$$

The first bullet point gives us

$$\sum_{1 \leq i \leq n} u(\min_{S_i}) \leq \left(\sum_{1 \leq i \leq n} u(\min_{R_i}) \right) + n\epsilon, \tag{7}$$

so we get $\alpha(S) \leq \alpha(R)$. If the inequality is strict, we have $S <_\alpha R$. If not, we have

- $u(\min_{S_i}) = u(\min_{R_i}) + \epsilon$ for all i
- $u(\max_{S_i}) = u(\max_{R_i}) + \epsilon$ for $i \neq m$
- $u(\max_{S_m}) = u(\max_{R_m}) - (2n-1)\epsilon$.

Because of the inequalities $\min_S \le \min_R + \epsilon$ and $\max_S \le \max_R + \epsilon$ (recall that these are inequalities of decorated points with the poset structure we defined earlier), we have $P(\min_{S_i}) \le P(\min_{R_i})$ for all i and $P(\max_{S_i}) \le P(\max_{R_i})$ for $i \ne m$. But since $\max_{S_m} < \max_{R_m} - (2n-1)\epsilon$, we have $P(\max_{S_m}) < P(\max_{R_m})$, so $S <_\alpha R$. Similarly, we can prove $T <_\alpha S$ if $\mathbb{I}^S$ and $\mathbb{I}^T$ are not $(2n-1)\epsilon$-interleaved, so we have $T <_\alpha R$, which is a contradiction. □

Lemma 4.7. *Let R, S, and T be rectangles of the same type with R and T $(4n-2)\epsilon$-significant and $\alpha(R) \le \alpha(T)$. Suppose there are nonzero morphisms $\chi : \mathbb{I}^R \to \mathbb{I}^S(\epsilon)$ and $\psi : \mathbb{I}^S \to \mathbb{I}^T(\epsilon)$. Then $\psi(\epsilon) \circ \chi \ne 0$.*

The constant $(4n-2)$ can be improved on for $n > 1$, but since the constant $(2n-1)$ in Lemma 4.6 is optimal, strengthening Lemma 4.7 will not help us get a better constant in Theorem 4.2.

Proof. Suppose that χ and ψ are nonzero, but $\psi(\epsilon) \circ \chi = 0$. We have

- $\min_R + 2\epsilon \ge \min_T$
- $\min_{R_m} + 2\epsilon \ge \max_{T_m}$ for some m
- $\max_R + 2\epsilon \ge \max_T$
- $\max_{R_m} \ge \max_{T_m} + (4n-4)\epsilon$.

The first and third statements hold because $\chi, \psi \ne 0$. (See Corollary 4.4.) The second is equivalent to $\min_R \not< \max_{T(2\epsilon)}$. If this did not hold, R and $T(2\epsilon)$ would intersect, and $\psi(\epsilon) \circ \chi$ would be nonzero in this intersection, which is a contradiction. The fourth statement follows from the second and the fact that R is $(4n-2)\epsilon$-significant.

Since T is $(4n-2)\epsilon$-significant, $\min_T + (4n-2)\epsilon < \max_T$. Thus the second bullet point implies that $\min_{R_m} + 2\epsilon > \min_{T_m} + (4n-2)\epsilon$. The first point gives $\min_{R_i} \ge \min_{T_i} - 2\epsilon$ for $i \ne m$. In a similar fashion, we get from the last two points that $\max_{R_m} \ge \max_T + (4n-4)\epsilon$ and $\max_{R_i} \ge \max_{T_i} - 2\epsilon$ for $i \ne m$. From all this, we get

$$\begin{aligned}
\alpha(R) &= \sum_{1 \le i \le n} u(\min_{R_i}) + u(\max_{R_i}) \\
&\ge u(\min_{T_m}) + u(\max_{T_m}) + 2(4n-4)\epsilon + \sum_{i \ne m} (u(\min_{T_i}) + u(\max_{T_i}) - 4\epsilon) \\
&= \alpha(T) + (4n-4)\epsilon \\
&\ge \alpha(T).
\end{aligned} \tag{8}$$

Equality only holds if $u(\min_{T_m}) + (4n-2)\epsilon = u(\max_{T_m})$, $u(\min_{R_m}) + (4n-2)\epsilon = u(\max_{R_m})$, and $n = 1$. This means that $R = R_1 = T = T_1 = [u(\min_R), u(\min_R) + 2\epsilon]$. As we see, $R \cap T(2\epsilon) = [u(\min_R), u(\min_R)] \ne \varnothing$, so $\psi(\epsilon) \circ \chi \ne 0$. □

We define a function μ by

$$\mu(I) = \{J \in B(N) \mid I \text{ and } J \text{ are } (2n-1)\delta\text{-interleaved}\} \tag{9}$$

for I in $B(M)$. In other words, $\mu(I)$ contains all the intervals that can be matched with I in a $(2n-1)\delta$-matching. Let $I \in B(M)$ be $(4n-2)\delta$-significant, and pick $p \in \mathbb{R}^n$ such that $p, p + (4n-2)\delta \in I$. Then, $p + (2n-1)\delta \in J$ for every $J \in \mu(I)$. Since M and N are p.f.d., this means that $\mu(I)$ is a finite set. For $A \subset B(M)$, we write $\mu(A) = \bigcup_{I \in A} \mu(I)$.

Lemma 4.8. *Let A be a finite subset of $B(M)$ containing no $(4n-2)\delta$-trivial elements. Then $|A| \le |\mu(A)|$.*

Before we prove Lemma 4.8, we show that it implies that there is a $(2n-1)\delta$-matching between $B(M)$ and $B(N)$ and thus completes the proof of Theorem 4.2.

Let G_μ be the undirected bipartite graph on $B(M) \sqcup B(N)$ with an edge between I and J if $J \in \mu(I)$. Observe that G_μ is the same as the graph $G_{(2n-1)\delta}$ we defined when we gave the graph theoretical definition of an ϵ-matching (in this case, $(2n-1)\delta$-matching) in section 2. Following that definition, a $(2n-1)\delta$-matching is a matching in G_μ that covers the set of all $(4n-2)\delta$-significant elements in $B(M)$ and $B(N)$.

For a subset S of a graph G, let $A_G(S)$ be the neighbourhood of S in G, that is, the set of vertices in G that are adjacent to at least one vertex in S. We now apply Hall's marriage theorem [18] to bridge the gap between Lemma 4.8 and the statement we want to prove about matchings.

Theorem 4.9 (Hall's theorem)**.** *Let G be a bipartite graph on bipartite sets X and Y such that $A_G(\{x\})$ is finite for all $x \in X$. Then the following are equivalent:*

- *for all $X' \subset X$, $|X'| \leq |A_G(X')|$*
- *there exists a matching in G covering X.*

One of the two implications is easy, since if $|X'| > |A_G(X')|$ for some $X' \subset X$, then there is no matching in G covering X'. It is the other implication we will use, namely that the first statement is sufficient for a matching in G covering X to exist.

Letting X be the set of $(4n-2)\delta$-significant intervals in $B(M)$ and Y be $B(N)$, Hall's theorem and Lemma 4.8 give us a matching σ in the graph G_μ covering all the $(4n-2)\delta$-significant elements in $B(M)$.[4] By symmetry, we also have a matching τ in G_μ covering all the $(4n-2)\delta$-significant elements in $B(N)$. Neither of these is necessarily a $(2n-1)\delta$-matching, however, as each of them only guarantees that all the $(4n-2)$-significant intervals in one of the barcodes are matched. We will use σ and τ to construct a $(2n-1)\delta$-matching. This construction is similar to one used to prove the Cantor-Bernstein theorem [1, pp. 110-111].

Let H be the undirected bipartite graph on $B(M) \sqcup B(N)$ for which the set of edges is the union of the edges in the matchings σ and τ. Let C be a connected component of H. Suppose the submatching of σ in C does not cover all the $(4n-2)\delta$-significant elements of C. Then there is a $(4n-2)\delta$-significant $J \in C \cap B(N)$ that is not matched by σ. If we view σ and τ as partial bijections $\sigma : B(M) \nrightarrow B(N)$ and $\tau : B(N) \nrightarrow B(M)$, we can write the connected component of J, which is C, as $\{J, \tau(J), \sigma(\tau(J)), \tau(\sigma(\tau(J))), \dots\}$. Either this sequence is infinite, or it is finite, in which case the last element is $(4n-2)\delta$-trivial. In either case, we get that the submatching of τ in C covers all $(4n-2)\delta$-significant elements in C.

By this argument, there is a $(2n-1)\delta$-matching in each connected component of H. We can piece these together to get a $(2n-1)\delta$-matching in $B(M) \sqcup B(N)$, so Lemma 4.8 completes the proof of Theorem 4.2.

Proof of Lemma 4.8. Because $\leq_\alpha$ is a preorder, we can order $A = \{I_1, I_2, \dots, I_r\}$ so that $I_i \leq_\alpha I_{i'}$ for all $i \leq i'$. Write $\mu(A) = \{J_1, J_2, \dots, J_s\}$. For $I \in B(M)$, we have

$$\begin{aligned}
\phi_{\mathbb{I}^I, 2\delta} &= \pi_I(2\delta) g(\delta) f|_I \\
&= \pi_I(2\delta) \left(\sum_{J \in B(N)} g|_J \pi_J \right) (\delta) f|_I \\
&= \sum_{J \in B(N)} \pi_I(2\delta) g|_J(\delta) \pi_J(\delta) f|_I \\
&= \sum_{J \in B(N)} g_{J,I}(\delta) f_{I,J}.
\end{aligned} \tag{10}$$

Also, $\sum_{J \in B(N)} g_{J,I'}(\delta) f_{I,J} = 0$ for $I \neq I' \in B(M)$, since $\phi_{M,2\delta}$ is zero between different components of M. Lemma 4.6 says that if $g_{J,I'}(\delta) f_{I,J} \neq 0$ and $I \leq_\alpha I'$, then J is $(2n-1)\delta$-interleaved with either I or J'. This means that if $i < i'$, then

$$\begin{aligned}
0 &= \sum_{J \in B(N)} g_{J,I_{i'}}(\delta) f_{I_i,J} \\
&= \sum_{J \in \mu(A)} g_{J,I_{i'}}(\delta) f_{I_i,J},
\end{aligned} \tag{11}$$

[4] Strictly speaking, Lemma 4.8 says nothing about infinite A, but the case with A countably infinite follows from the finite cases. Each interval in A contains a rational point, so since M is p.f.d., the cardinality of A is at most finite times countably infinite, which is countable. Thus we have covered all the possible cases.

as $g_{J,I_{i'}}(\delta)f_{I_i,J} = 0$ for all J that are not $(2n-1)\delta$-interleaved with either I_i or $I_{i'}$. Similarly,

$$\begin{aligned}\phi_{\mathbb{I}^{I_i},2\delta} &= \sum_{J\in B(N)} g_{J,I_i}(\delta)f_{I_i,J} \\ &= \sum_{J\in\mu(A)} g_{J,I_i}(\delta)f_{I_i,J}.\end{aligned} \tag{12}$$

Writing this in matrix form, we get

$$\begin{bmatrix} g_{J_1,I_1}(\delta) & \dots & g_{J_s,I_1}(\delta) \\ \vdots & \ddots & \vdots \\ g_{J_1,I_r}(\delta) & \dots & g_{J_s,I_r}(\delta)\end{bmatrix}\begin{bmatrix} f_{I_1,J_1} & \dots & f_{I_r,J_1} \\ \vdots & \ddots & \vdots \\ f_{I_1,J_s} & \dots & f_{I_r,J_s}\end{bmatrix} = \begin{bmatrix} \phi_{M_{\mathbb{I}^{I_1}},2\delta} & ? & \dots & ? \\ 0 & \phi_{M_{\mathbb{I}^{I_2}},2\delta} & \dots & ? \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_{M_{\mathbb{I}^{I_r}},2\delta}\end{bmatrix}.$$

That is, on the right-hand side we have the internal morphisms of the I_i on the diagonal, and 0 below the diagonal.

Recall that a morphism between rectangle modules can be identified with a k-endomorphism, and that in our notation, $f_{I,J}$ and $g_{J,I}$ are given by multiplication by $w(I,J)$ and $w(J,I)$, respectively. For an arbitrary morphism ψ between rectangle modules, we introduce the notation $w(\psi) = c$ if ψ is given by multiplication by c, and 0 otherwise. A consequence of Lemma 4.7 is that $w(g_{J,I_{i'}}(\delta)f_{I_i,J}) = w(g_{J,I_{i'}})w(f_{I_i,J}) = w(J,I_i)w(I_{i'},J)$ whenever $I_i \leq_\alpha I_{i'}$, in particular if $i \leq i'$. We get

$$\begin{aligned}1 &= w\left(\phi_{\mathbb{I}^I,2\delta}\right) \\ &= w\left(\sum_{J\in\mu(A)} g_{J,I_i}(\delta)f_{I_i,J}\right) \\ &= \sum_{J\in\mu(A)} w(g_{J,I_i}(\delta)f_{I_i,J}) \\ &= \sum_{J\in\mu(A)} w(J,I_i)w(I_i,J),\end{aligned} \tag{13}$$

and similarly $0 = \sum_{J\in\mu(A)} w(J,I_{i'})w(I_i,J)$ for $i \leq i'$. Again we can interpret this as a matrix equation:

$$\begin{bmatrix} w(J_1,I_1) & \dots & w(J_s,I_1) \\ \vdots & \ddots & \vdots \\ w(J_1,I_r) & \dots & w(J_s,I_r)\end{bmatrix}\begin{bmatrix} w(I_1,J_1) & \dots & w(I_r,J_1) \\ \vdots & \ddots & \vdots \\ w(I_1,J_s) & \dots & w(I_r,J_s)\end{bmatrix} = \begin{bmatrix} 1 & ? & \dots & ? \\ 0 & 1 & \dots & ? \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1\end{bmatrix}.$$

That is, the right-hand side is an $r \times r$ upper triangular matrix with 1's on the diagonal. The right-hand side has rank $|A|$ and the left-hand side has rank at most $|\mu(A)|$, so the lemma follows immediately from this equation. □

4.2 Block decomposable modules

Next, we prove stability for block decomposable modules, which, as explained in Section 3, implies stability for zigzag modules and Reeb graphs. Let $\mathbb{R}^2_+ = \{(x,y) \in \mathbb{R}^2 \mid x + y \geq 0\}$.

Definition 4.10. *A* ***triangle*** *is a nonempty set T of the form $\{(x,y) \in \mathbb{R}^2 \mid x < a, y < b\} \cap \mathbb{R}^2_+$ for some $(a,b) \in (\mathbb{R} \cup \{\infty\})^2$ with $a + b > 0$.*

It follows that triangles are intervals. For a triangle $T = \{(x,y) \in \mathbb{R}^2 \mid x < a, y < b\} \cap \mathbb{R}^2_+$, we write $\max_T = (a,b) \in (\mathbb{R}\cup\{\infty\})^2$. If T is bounded, $\max_T$ is the maximal element in the closure of T, as illustrated in Figure 5. A *triangle decomposable module* is an interval decomposable $\mathbb{R}^2$-module whose barcode only contains triangles. Observe that triangles correspond exactly to blocks of the form $(a,b)_{BL}$ under the poset isomorphism between $\mathbb{R}^{\mathrm{op}} \times \mathbb{R}$ and $\mathbb{R}^2$ flipping the x-axis.

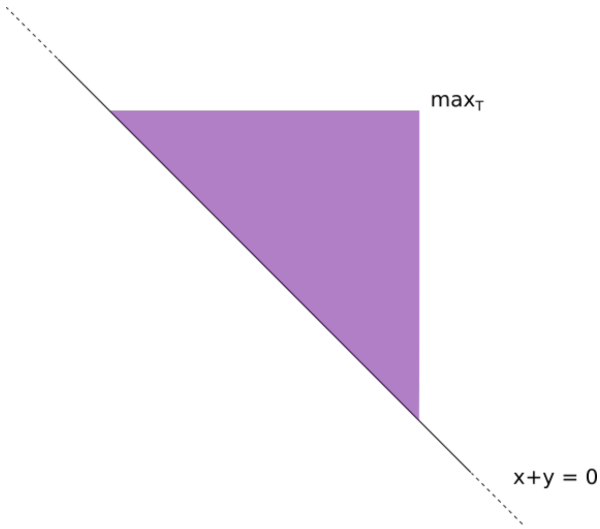


Figure 5: A bounded triangle T.

Theorem 4.11. *Let M and N be δ-interleaved triangle decomposable modules. Then there is a δ-matching between $B(M)$ and $B(N)$.*

To prove this, we can split the triangles into sets of different 'types', as we did with the rectangles. We get four different types of triangles T, depending on whether $\max_T$ is of the form (a, b), (∞, b), (a, ∞), or (∞, ∞) for $a, b \in \mathbb{R}$. Now a result analogous to Lemma 4.5 holds, implying that it is enough to show Theorem 4.11 under the assumption that the barcodes only contain intervals of a single type. The case in which the triangles are bounded is the hardest one, and the only one we will prove. So from now on, we assume all triangles to be bounded.

Again, we reuse parts of the proof of Theorem 4.2. For $I \in B(M)$, we define $\nu(I) = \{J \in B(N) \mid I \text{ and } J \text{ are } \delta\text{-interleaved}\}$. The discussion about Hall's theorem is still valid, so we only need to prove the analogue of Lemma 4.8 for ν. Define $\alpha(T) = m_1 + m_2$, where $\max_T = (m_1, m_2)$. The only things we need to complete the proof of the analogue of Lemma 4.8 for triangle decomposable modules are the following analogues of Lemmas 4.6 and 4.7:

Lemma 4.12. *Let R, S, and T be triangles with $\alpha(R) \leq \alpha(T)$. Suppose there are morphisms $f : \mathbb{I}^R \to \mathbb{I}^S(\epsilon)$ and $g : \mathbb{I}^S \to \mathbb{I}^T(\epsilon)$ such that $g(\epsilon) \circ f \neq 0$. Then $\mathbb{I}^S$ is ϵ-interleaved with either $\mathbb{I}^R$ or $\mathbb{I}^T$.*

Lemma 4.13. *Let R, S, and T be triangles with T 2ϵ-significant and $\alpha(R) \leq \alpha(T)$. Suppose there are nonzero morphisms $f : \mathbb{I}^R \to \mathbb{I}^S(\epsilon)$ and $g : \mathbb{I}^S \to \mathbb{I}^T(\epsilon)$. Then $g(\epsilon) \circ f \neq 0$.*

Proof of Lemma 4.12. Suppose $\mathbb{I}^R$ and $\mathbb{I}^S$ are not ϵ-interleaved. Then $\max_S \ngeq \max_R - \epsilon$. But at the same time, $\max_R \geq \max_S - \epsilon$, which gives $\alpha(R) > \alpha(S)$. Assuming that $\mathbb{I}^S$ and $\mathbb{I}^T$ are not ϵ-interleaved, either, we also get $\alpha(S) > \alpha(T)$. Thus $\alpha(R) > \alpha(T)$, a contradiction. $\square$

Proof of Lemma 4.13. For all triangles I, we treat $\min_I$ and $\max_I$ as undecorated points. We have $\max_T - \epsilon \leq \max_S$ and $\max_S - \epsilon \leq \max_R$, so $\max_T - 2\epsilon \leq \max_R$. Because T is 2ϵ-significant, $\max_T - 2\epsilon - \epsilon' \in \mathbb{R}^2_+$ for some $\epsilon' > 0$. Combining these facts, we get $\max_T - 2\epsilon - \epsilon' \in R$, so $(g(\epsilon) \circ f)_{\max_T - 2\epsilon - \epsilon'} \neq 0$. $\square$

Theorem 4.11 implies $d_B(M, N) = d_I(M, N)$ for block decomposable M and N such that $B(M)$ and $B(N)$ only have blocks of the form $(a, b)_{BL}$ (so no closed or half-closed blocks). Our proof technique extends easily to prove the same equality for all block decomposable M and N. In fact, $d_B(M, N) \leq d_I(M, N)$ in the case where all the intervals in the barcodes are of the form $[a, b]_{BL}$ follows from Theorem 4.16 below with $n = 2$ by the correspondence $[a, b]_{BL} \leftrightarrow \langle(-a, b)\rangle$, while the two cases with half-open blocks are both essentially the algebraic stability theorem. In the end we could stitch the cases together by something similar to Lemma 4.5 and the discussion following it. We omit the details, and anyway the closed and half-open cases are taken care of in [3]. Thus, either by appealing to previous work for the other cases or using our own methods, we get

Theorem 4.14. *Let M and N be block decomposable modules. If M and N are δ-interleaved, there exists a δ-matching between $B(M)$ and $B(N)$.*

4.3 Free modules

Definition 4.15. *We define a **free interval** as an interval of the form $\langle p\rangle := \{q \mid q \geq p\} \subset \mathbb{R}^n$.*

For a free interval R, we define $\min_R$ by $R = \langle \min_R\rangle$.[5] We define a *free $\mathbb{R}^n$-module* as an interval decomposable module whose barcode only contains free intervals. It is easy to see that free intervals are rectangles, so it follows from Theorem 4.2 that $d_B(M, N) \leq (2n-1)d_I(M, N)$ for free modules M, N. But because of the geometry of free modules, this result can be strengthened.

Theorem 4.16. *Let M and N be free δ-interleaved $\mathbb{R}^n$-modules with $n \geq 2$. Then there is a $(n-1)\delta$-matching between $B(M)$ and $B(N)$.*

We already did most of the work while proving Theorem 4.2, and there are some obvious simplifications. Firstly, free intervals are ϵ-significant for all $\epsilon \geq 0$. Secondly, for all nonzero $f : \mathbb{I}^R \to \mathbb{I}^S$ and $g : \mathbb{I}^S \to \mathbb{I}^T$ with R, S, T free, gf is nonzero. For $I \in B(M)$, define $\nu(I) = \{J \in B(N) \mid I \text{ and } J \text{ are } (n-1)\delta\text{-interleaved}\}$. By the arguments in the proof of Theorem 4.2, we only need to prove Lemma 4.8 with μ replaced by ν. Lemmas 4.6 and 4.7 still hold for free modules, but we need to sharpen Lemma 4.6.

Lemma 4.17. *Let R, S, and T be free intervals with $R \leq_\alpha T$. Suppose there are morphisms $0 \neq f : \mathbb{I}^R \to \mathbb{I}^S(\epsilon)$ and $0 \neq g : \mathbb{I}^S \to \mathbb{I}^T(\epsilon)$. Then $\mathbb{I}^S$ is $(n-1)\epsilon$-interleaved with either $\mathbb{I}^R$ or $\mathbb{I}^T$.*

Proof. In this proof, we treat $\min_I$ and $\max_I$ as undecorated points for all free intervals I, so that we can add them. We have $\min_S \leq \min_R + \epsilon$. Suppose $\mathbb{I}^R$ and $\mathbb{I}^S$ are not $(n-1)\epsilon$-interleaved. Then $\min_S + (n-1)\epsilon \not\geq \min_R$, so for some m, we must have $\min_{S_m} < \min_{R_m} - (n-1)\epsilon$. We get

$$\begin{aligned}\alpha(S) &= \sum_{1\leq i\leq n} \min_{S_i} \\ &< \min_{R_m} - (n-1)\epsilon + \sum_{i\neq m} (\min_{R_i} + \epsilon) \\ &= \sum_{1\leq i\leq n} \min_{R_i} \\ &= \alpha(R).\end{aligned} \tag{14}$$

We can also prove that $\alpha(T) < \alpha(S)$ if $\mathbb{I}^S$ and $\mathbb{I}^T$ are not $(n-1)\epsilon$-interleaved, so we have $\alpha(T) < \alpha(R)$, a contradiction. □

5 Counterexamples to a general algebraic stability theorem

Theorem 4.2 gives an upper bound of $(2n-1)$ on d_B/d_I for rectangle decomposable modules that increases with the dimension. An obvious question is whether it is possible to improve this constant, or if for each $C < 2(n-1)$ there exist pairs M, N of modules for which $d_B(M, N) > Cd_I(M, N)$, in which case the bound is optimal. We know that $d_B(M, N) \geq d_I(M, N)$ for any M and N whenever the bottleneck

[5] This makes $\min_R$ an undecorated point, while we have previously defined $\min_{-}$ as decorated points, but this does not matter, as we will not need decorated points in this subsection.

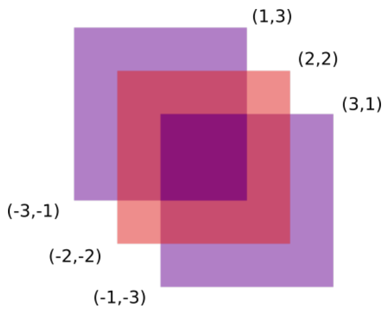


Figure 6: M and N. I_1 and I_2 are the light purple squares, I_3 is deep purple, and J is pink.

distance is defined, so for $n = 1$, the constant is optimal. For $n > 1$, however, it turns out that the equality $d_B(M, N) = d_I(M, N)$ does not always hold, and the geometry becomes more confusing when n increases. In dimension 2, we give an example of rectangle decomposable modules M and N with $d_B(M, N) = 3d_I(M, N)$ in Example 5.2, which means that the bound is optimal for $n = 2$, as well. This is a counterexample to a conjecture made in a previous version of [3] which claims that interval decomposable $\mathbb{R}^n$-modules M and N such that $B(M)$ and $B(N)$ only contain convex intervals are ϵ-matched if they are ϵ-interleaved.

Example 5.1. *Let* $B(M) = \{I_1, I_2, I_3\}$[6] *and* $B(N) = \{J\}$*, where*

- $I_1 = (-3, 1) \times (-1, 3)$
- $I_2 = (-1, 3) \times (-3, 1)$
- $I_3 = (-1, 1) \times (-1, 1)$
- $J = (-2, 2) \times (-2, 2)$.

See Figure 6. We can define 1*-interleaving morphisms* $f : M \to N(1)$ *and* $g : N \to M(1)$ *by letting* $w(I_1, J) = w(I_2, J) = w(I_3, J) = w(J, I_1) = w(J, I_2) = 1$ *and* $w(J, I_3) = -1$*, where* w *is defined as in the proof of Theorem 4.2. On the other hand, in any matching between* $B(M)$ *and* $B(N)$ *we have to leave either* I_1 *or* I_2 *unmatched, and they are* ϵ*-significant for all* $\epsilon < 4$*. In fact, any possible matching between* $B(M)$ *and* $B(N)$ *is a* 2*-matching. Thus* $d_I(M, N) = 1$ *and* $d_B(M, N) = 2$*.*

A crucial point is that even though $w(I_1, J)$, $w(J, I_2)$, $w(I_2, J)$, and $w(J, I_1)$ are all nonzero, both $g_{J,I_2} \circ f_{I_1,J}$ and $g_{J,I_1} \circ f_{I_2,J}$ are zero. To do the same with one-dimensional intervals, we would have to shrink I_1 and I_2 so much that they no longer would be 2-significant (see Lemma 4.7), and then they would not need to be matched in a 1-matching. This shows how the geometry of higher dimensions can allow us to construct examples that would not work in lower dimensions.

Next, we give an example of rectangle decomposable $\mathbb{R}^2$-modules M and N such that $d_B(M, N) = 3d_I(M, N)$, proving that our upper bound of $2(n - 1)$ is the best possible for $n = 2$.

Example 5.2. *Let* $B(M) = \{I_1, I_2, I_3\}$ *and* $B(N) = \{J_1, J_2, J_3\}$*, where*

- $I_1 = (0, 10) \times (1, 11)$
- $I_2 = (0, 12) \times (-1, 11)$
- $I_3 = (2, 10) \times (1, 9)$
- $J_1 = (1, 11) \times (0, 10)$
- $J_2 = (1, 9) \times (0, 12)$
- $J_3 = (-1, 11) \times (2, 10)$.

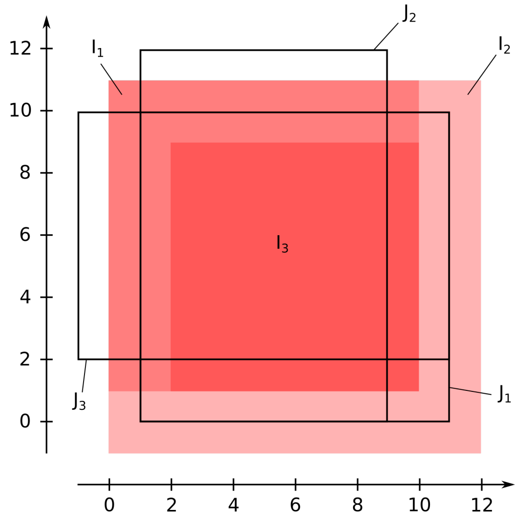


Figure 7: I_1, I_2, and I_3 are the filled pink rectangles, and J_1, J_2, and J_3 are the black rectangles without fill.

The rectangles in $B(M)$ and $B(N)$ are shown in Figure 7.

We give an example of 1-interleaving morphisms f and g that we write on matrix form. In the first matrix, $w(I_i, J_j)$ is in row i, column j. In the second, $w(J_j, I_i)$ is in row j, column i.

$$f : \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \qquad g : \begin{bmatrix} -1 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix}. \tag{15}$$

This means that M and N are 1-interleaved, but they are not ϵ-interleaved for any $\epsilon < 1$, so $d_I(M, N) = 1$.

Let $\epsilon < 3$. We see that the difference between $max_{I_2} = (12, 11)$ and $max_{J_2} = (9, 12)$ is 3 in the first coordinate, so I_2 and J_2 are not ϵ-interleaved, and thus they cannot be matched in an ϵ-matching. In fact, I_i and J_j cannot be matched in an ϵ-matching for any $i, j \in \{2, 3\}$ by similar arguments. Since I_2 and I_3 cannot both be matched with J_1, one of them has to be left unmatched, but since both I_2 and I_3 are 6-significant, this means that there is no ϵ-matching between $B(M)$ and $B(N)$. On the other hand, any bijection between $B(M)$ and $B(N)$ is a 3-matching, so $d_B(M, N) = 3$.

There is a strong connection between n-dimensional rectangle decomposable modules and $2n$-dimensional free modules. This is related to the fact that we need $2n$ coordinates to determine an n-dimensional rectangle, and also $2n$ coordinates to determine a $2n$-dimensional free interval. The following example illustrates this connection, as we simply rearrange the coordinates of $\min_R$, $\max_R$ for all rectangles R involved in Example 5.2 to get 4-dimensional free modules with similar properties as in Example 5.2.

Example 5.3. *Let $B(M) = \{I_1, I_2, I_3\}$ and $B(N) = \{J_1, J_2, J_3\}$, where*

- $I_1 = \langle(0, 1, 10, 11)\rangle$
- $I_2 = \langle(0, -1, 12, 11)\rangle$
- $I_3 = \langle(2, 1, 10, 9)\rangle$

[6] Here we use subscripts to index different intervals, not to indicate projections, as we did earlier.

- $J_1 = \langle(1, 0, 11, 10)\rangle$
- $J_2 = \langle(1, 0, 9, 12)\rangle$
- $J_3 = \langle(-1, 2, 11, 10)\rangle$.

(Compare with the intervals I_i and J_j in Example 5.2.) We have 1-interleaving morphisms defined the same way as in Example 5.2. Just as in that example, we can deduce that there is nothing better than a 3-matching between $B(M)$ and $B(N)$, so $d_B(M, N) = 3$ and $d_I(M, N) = 1$.

As a consequence of this example, we get that our upper bound of $d_B/d_I \leq n - 1$ for free n-dimensional modules cannot be improved on for $n = 4$.

6 Relation to the complexity of calculating interleaving distance

The interleaving distance between arbitrary persistence modules is on the surface not easy to find, as naively trying to construct interleaving morphisms can quickly lead you to consider a complicated set of equations for which it is not clear that one can decide if there is a solution in polynomial time. For $\mathbb{R}$-modules, however, the interval decomposition theorem plus the algebraic stability theorem gives us a polynomial time algorithm to compute d_I: decompose the modules into intervals and find the bottleneck distance. Since $d_I = d_B$, this gives us the interleaving distance. When it exists, one can compute the bottleneck distance in polynomial time also in two dimensions [17], but the approach fails for general $\mathbb{R}^n$-modules already at the first step, as we do not have a nice decomposition theorem. But in the recent proof that calculating interleaving distance is NP-hard [9], it is the failure of the second step that is exploited. Specifically, a set of modules that decompose nicely into interval modules (*staircase modules*, to be precise) is constructed, but for these, d_I and d_B are different. It turns out that calculating d_I for these corresponds to deciding whether *CI problems* are solvable, which is shown to be NP-hard.

Though rectangle modules are not considered in the NP-hardness proof, they have similar properties to staircase modules,[7] and Example 5.2 is essentially a CI problem with a corresponding pair of modules. Importantly, it shows that $d_I = d_B$ does not hold in general for modules corresponding to CI problems. This crucial observation, which appeared first in a preprint of this paper [8], opened the door to proving NP-hardness of calculating d_I by the approach used in [9].

In [9], it is also shown that *c-approximating d_I* is NP-hard for $c < 3$, where an algorithm is said to c-approximate d_I if it returns a number in the interval $[d_I(M, N), cd_I(M, N)]$ for any input pair M, N of modules. Whether the approach by CI problems can be used to prove hardness of c-approximation for $c \geq 3$ is closely related to Theorem 4.2. It can be shown that if $d_B(M, N) \leq cd_I(M, N)$ for any pair M, N of rectangle decomposable modules, the same holds for staircase modules, and therefore there is a polynomial time algorithm c-approximating d_I for these, meaning that the strategy of going through CI problems will not give a proof that c-approximation of d_I is NP-hard. On the other hand, if one can find an example of rectangle decomposable modules M and N such that $d_B(M, N) = cd_I(M, N)$ for $c > 3$, one might be able to use that to increase the constant 3 in the approximation hardness result. Thus there is a strong link between stability of rectangle decomposable modules and the only successful method so far known to the author of determining the complexity of computing or approximating multiparameter interleaving distance.

7 Acknowledgements

I would like to thank my supervisors Gereon Quick and Nils Baas for invaluable support and help. I would also like to thank Peter Landweber for detailed comments on several drafts of this text, Steve Oudot for feedback on the first arXiv version and Magnus Bakke Botnan for interesting discussions.

[7]The only significant difference in this setting is that in a fixed dimension, rectangle modules are defined by a limited number of coordinates, or "degrees of freedom", while there is no such restriction on staircase modules even in dimension 2.

References

[1] Martin Aigner and Günter M. Ziegler. *Proofs from THE BOOK*. Springer, Berlin, 4th edition, 2010.

[2] Gorô Azumaya. Corrections and supplementaries to my paper concerning Krull-Remak-Schmidt's theorem. *Nagoya Mathematical Journal*, 1:117–124, 1950.

[3] Magnus Bakke Botnan and Michael Lesnick. Algebraic stability of zigzag persistence modules. *Algebraic & Geometric Topology*, 18, 04 2016.

[4] Ulrich Bauer, Xiaoyin Ge, and Yusu Wang. Measuring distance between reeb graphs. In *Proceedings of the thirtieth annual symposium on Computational geometry*, page 464. ACM, 2014.

[5] Ulrich Bauer and Michael Lesnick. Induced matchings and the algebraic stability of persistence barcodes. *Journal of Computational Geometry*, 6(2):162–191, 2015.

[6] Ulrich Bauer, Elizabeth Munch, and Yusu Wang. Strong Equivalence of the Interleaving and Functional Distortion Metrics for Reeb Graphs. In *31st International Symposium on Computational Geometry (SoCG 2015)*, volume 34 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 461–475. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015.

[7] Paul Bendich, Herbert Edelsbrunner, Dmitriy Morozov, Amit Patel, et al. Homology and robustness of level and interlevel sets. *Homology, Homotopy and Applications*, 15(1):51–72, 2013.

[8] Håvard Bakke Bjerkevik. Stability of higher-dimensional interval decomposable persistence modules. *arXiv preprint arXiv:1609.02086v2*, 2016.

[9] Håvard Bakke Bjerkevik, Magnus Bakke Botnan, and Michael Kerber. Computing the interleaving distance is np-hard. *arXiv preprint arXiv:1811.09165*, 2018.

[10] Gunnar Carlsson and Vin De Silva. Zigzag persistence. *Foundations of computational mathematics*, 10(4):367–405, 2010.

[11] Gunnar Carlsson, Vin De Silva, and Dmitriy Morozov. Zigzag persistent homology and real-valued functions. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 247–256. ACM, 2009.

[12] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J Guibas, and Steve Y Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 237–246. ACM, 2009.

[13] Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve Oudot. *The structure and stability of persistence modules*. Springer, 2016.

[14] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.

[15] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and Its Applications*, 14(05):1550066, 2015.

[16] Vin De Silva, Elizabeth Munch, and Amit Patel. Categorified reeb graphs. *Discrete & Computational Geometry*, 55(4):854–906, 2016.

[17] Tamal K. Dey and Cheng Xin. Computing Bottleneck Distance for 2-D Interval Decomposable Modules. In *34th International Symposium on Computational Geometry (SoCG 2018)*, volume 99 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 32:1–32:15. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018.

[18] Philip Hall. On representatives of subsets. *J. London Math. Soc*, 10(1):26–30, 1935.

[19] Woojin Kim and Facundo Memoli. Stable signatures for dynamic metric spaces via zigzag persistent homology. *arXiv preprint arXiv:1712.04064*, 2017.

[20] Steve Y Oudot and Donald R Sheehy. Zigzag zoology: Rips zigzags for homology inference. *Foundations of Computational Mathematics*, 15(5):1151–1186, 2015.

[21] Cary Webb. Decomposition of graded modules. *Proceedings of the American Mathematical Society*, 94(4):565–571, 1985.

Paper II

Computational Complexity of the Interleaving Distance

Håvard Bakke Bjerkevik   Magnus Bakke Botnan

Version without appendices published in the proceedings of SoCG 2018

Computational Complexity of the Interleaving Distance*

Håvard Bakke Bjerkevik† Magnus Bakke Botnan‡

Abstract

The interleaving distance is arguably the most prominent distance measure in topological data analysis. In this paper, we provide bounds on the computational complexity of determining the interleaving distance in several settings. We show that the interleaving distance is NP-hard to compute for persistence modules valued in the category of vector spaces. In the specific setting of multidimensional persistent homology we show that the problem is at least as hard as a matrix invertibility problem. Furthermore, this allows us to conclude that the interleaving distance of interval decomposable modules depends on the characteristic of the field. Persistence modules valued in the category of sets are also studied. As a corollary, we obtain that the isomorphism problem for Reeb graphs is graph isomorphism complete.

1 Introduction

For a category $\mathcal{C}$ and a poset $\mathbf{P}$ we define a $\mathbf{P}$-*indexed (persistence) module valued in* $\mathcal{C}$ to a be a functor $M : \mathbf{P} \to \mathcal{C}$. We will denote the associated functor category by $\mathcal{C}^{\mathbf{P}}$. If $M, N \in \mathcal{C}^{\mathbf{P}}$ then M and N are of the same *type*. Such functors appear naturally in applications, and most commonly when $\mathbf{P} = \mathbf{R}^n$, n-tuples of real numbers under the normal product order, and $\mathcal{C} = \mathbf{Vec}_{\mathbb{K}}$, the category of vector spaces over the field $\mathbb{K}$, or $\mathcal{C} = \mathbf{Set}$, the category of sets. The field $\mathbb{K}$ is assumed to be finite. We suppress notation and simply write $\mathbf{Vec}$ when $\mathbb{K}$ is an arbitrary finite field. The notation $p \in \mathbf{P}$ denotes that p is an object of $\mathbf{P}$.

Remark 1.1. Throughout the paper we make use of basic concepts from category theory. The reader unfamiliar to such ideas will find the necessary background material in the first few pages of [11].

Assume that $h : X \to \mathbb{R}$ is a continuous function of "Morse type", a generalization of a Morse function on a compact manifold. Roughly, a real-valued function is of Morse type if the homotopy type of the fibers changes at finite set of values; see [16] for a precise definition. We shall now briefly review four different scenarios in which functors of the aforementioned form can be associated to h.

Let $\mathrm{H}_p : \mathbf{Top} \to \mathbf{Vec}_{\mathbb{K}}$ denote the p-th singular homology functor with coefficients in $\mathbb{K}$, and let $\pi_0 : \mathbf{Top} \to \mathbf{Set}$ denote the functor giving the set of path-components. We also associate the

*M. B. Botnan has been supported by the DFG Collaborative Research Center SFB/TR 109 "Discretization in Geometry and Dynamics". This work was partially carried out while the authors were visitors to the Hausdorff Center for Mathematics, Bonn, during the special Hausdorff program on applied and computational topology.

†Norwegian University of Science and Technology, Trondheim, Norway; havard.bjerkevik@ntnu.no

‡TU München, Munich, Germany; botnan@ma.tum.de

two following functors to h whose actions on morphisms are given by inclusions:

$$\mathcal{S}^{\uparrow}(h) : \mathbf{R} \to \mathbf{Top} \qquad\qquad \mathcal{S}(h) : \mathbf{R}^2 \to \mathbf{Top}$$
$$\mathcal{S}^{\uparrow}(h)(t) = \{x \in X \mid h(x) \le t\} \qquad\qquad \mathcal{S}(h)(-s,t) = \{x \in X \mid s \le h(x) \le t\}$$

- *Persistent Homology* studies the evolution of the homology of the sublevel sets of h and is perhaps the most prominent tool in topological data analysis [16]. Specifically, the *p-th sublevel set persistence module associated to h* is the functor $\mathrm{H}_p\mathcal{S}^{\uparrow}(h) : \mathbf{R} \to \mathbf{Vec}$. Importantly, such a module is completely determined by a collection of intervals $\mathcal{B}(\mathrm{H}_p\mathcal{S}^{\uparrow}(h))$ called the *barcode* of $\mathrm{H}_p\mathcal{S}^{\uparrow}(h)$. This collection of intervals is then in turn used to extract topological information from the data at hand. In Fig. 1 we show the associated barcode for $p = 0$ and $p = 1$ for a function of Morse type.

- Upon replacing H_p by π_0 in the above construction we get a *merge tree*. That is, the *merge tree associated to h* is the functor $\tau^h : \pi_0\mathcal{S}^{\uparrow}(h) : \mathbf{R} \to \mathbf{Set}$. A merge tree captures the evolution of the path components of the sublevel sets of h and can be, as the name indicates, be visualized as (a disjoint union of) rooted trees. See Fig. 1 for an example.

- The two aforementioned examples used sublevel sets. A richer invariant is obtained by considering interlevel sets: define the *p-th interlevel set persistence of h* to be the functor $\mathrm{H}_p\mathcal{S}(h) : \mathbf{R}^2 \to \mathbf{Vec}$. Analogously to above, such a module is completely determined by a collection $\mathcal{B}(\mathrm{H}_p\mathcal{S}(h))$ of simple regions in $\mathbb{R}^2$. However, it is often the collection of intervals $\mathcal{L}_p(h)$ obtained by the intersection of these regions with the anti-diagonal $y = -x$ which are used in data analysis. We refer the reader to [9] for an in-depth treatment. In Fig. 1 we show an example of the 0-th interlevel set barcode. Observe how the endpoints of the intervals correspond to different types of features of the Reeb graph.

- Just as interlevel set persistence is a richer invariant than sublevel set persistence, the *Reeb graph* is richer in structure than the merge tree. Specifically, we define the functor $\mathrm{Reeb}^h := \pi_0\mathcal{S}(h) : \mathbf{R}^2 \to \mathbf{Set}$. Just as for Merge trees, Reeb^h admits a visualization of a graph; see Fig. 1. In particular, this appealing representation has made Reeb graphs a popular objects of study in computational geometry and topology, and they have found many applications in data visualization and exploratory data analysis.

These are all examples of topological invariants arising from a single real-valued function. There are many settings for which it is more fruitful to combine a collection of real-valued functions into a single function $g : X \to \mathbb{R}^n$ [15]. By combining them into a single function we not only learn how the data looks from the point of view of each function (i.e. a type of measurement) but how the different functions (measurements) interact. One obvious way to assign a (algebraic) topological invariant to g is to filter it by sublevel sets. That is, define $\mathcal{S}^{\uparrow}(g) : \mathbf{R}^n \to \mathbf{Top}$ by $\mathcal{S}^{\uparrow}(g)(t) = \{x \in X \mid g(x) \le t\}$. The associated functor $\mathrm{H}_p\mathcal{S}^{\uparrow}(g) : \mathbf{R}^n \to \mathbf{Vec}$ is an example of an *n-dimensional persistence module*. We saw above that for $n = 1$ this functor is completely described by a collection of intervals. This is far from true for $n \ge 2$: there exists no way to describe such functors by interval-like regions in higher-dimensional Euclidean space. Even the task of parameterizing such (indecomposable) modules is known to be a *hopeless* problem (so-called *wild representation type*) [4].

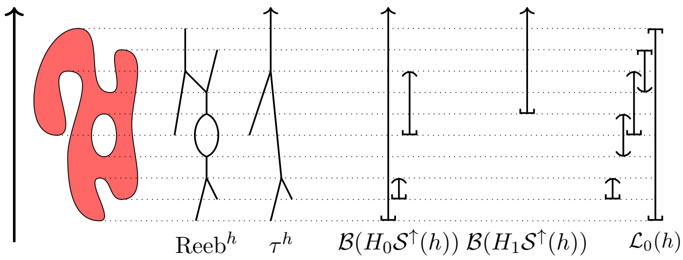


Figure 1: The height function of the solid shape is of Morse type. The associated Reeb graph, merge tree, sublevel set barcodes, and interlevel set barcode are shown to the right.

1.1 The Interleaving Distance

Different types of distances have been proposed on various types of persistence modules with values in **Vec** [6, 12, 17, 25, 27]. Of all these, the *interleaving distance* is arguably the most prominent for the following reasons: the theory of interleavings lies at the core of the theoretical foundations of 1-dimensional persistence, notably through the Isometry Theorem (Theorem 2.6). Furthermore, it was shown by Lesnick that when $\mathbb{K}$ is a prime field, the interleaving distance is the *most discriminative* of all *stable* metrics on such modules. We refer to [25] for the precise statement. As we shall see, it is also an immediate consequence of Theorem 2.6 that the interleaving distance for 1-dimensional persistence modules can be computed in polynomial time.

Lesnick's result generalizes to n-dimensional persistence modules, but the computational complexity of computing the interleaving distance of such modules remains unknown. An efficient algorithm to compute the interleaving distance could carry a profound impact on topological data analysis: the standard pipeline for 1-dimensional persistent homology is to first compute the barcode and then perform analysis on the collection of intervals. However, for multi-dimensional persistence there is no way of defining the barcode. With an efficient algorithm for computing the interleaving distance at hand it would still not be clear how to analyze the persistence modules individually, but we would have a theoretical optimal way of comparing them. This in turn could be used in clustering, kernel methods, and other kinds of data analysis widely applied in the 1-dimensional setting.

Complexity

The purpose of this paper is to determine the computational complexity of computing the interleaving distance. To make this precise, we need to associate a notion of size to the persistence modules.

Definition 1.2. Let $\mathbf{P}$ denote a poset category and $M : \mathbf{P} \to \mathcal{C}$.

- For $\mathcal{C} = \mathbf{Vec}$, define the *total dimension of* M to be $\dim M = \sum_{p\in\mathbf{P}} \dim M_p$.
- For $M : \mathbf{Z} \to \mathbf{Set}$, define the *total cardinality of* M to be $|M| = \sum_{p\in\mathbf{P}} |M_p|$.

The input size will be the total dimension or the total cardinality and for the the remaining of the paper **we shall always assume** that those quantities are finite. The following shows that there exists an algorithm, polynomial in the input size, which determines whether or not two $\mathbf{P}$-indexed modules valued in $\mathbf{Vec}$ are isomorphic.

Theorem 1.3 ([10]). *Let* $\mathbf{P}$ *be a finite poset and* $M, M' : \mathbf{P} \to \mathbf{Vec}$. *There exists a deterministic algorithm which decides if* $M \cong M'$ *in* $\mathcal{O}\left((\dim M + \dim M')^6\right)$.

This result will be important to us in what ensues because the strongest of interleavings, the 0-interleaving, is by definition a pair of inverse isomorphisms. Furthermore, by choosing an appropriate basis for each vector space, an isomorphism between M and M' is nothing more than a collection of matrices with entries in a finite field. Likewise a δ-interleaving will be nothing more than a collection of matrices over a finite field satisfying certain constraints. When $\mathcal{C} = \mathbf{Set}$ the morphisms are specified by collections of functions between finite sets. Hence, the decision problems considered in this paper **are trivially in NP**.

Furthermore, it is an immediate property of the Morse type of h, that the modules considered above are *discrete*. Intuitively, we say that an $\mathbf{R}^n$-indexed persistence module M is discrete if there exists a $\mathbf{Z}^n$-indexed persistence module containing all the information of M; see Appendix B. In practice, persistence modules arising from data will be discrete. Hence, when it comes to algorithmic questions we shall restrict ourselves to the setting in which $\mathbf{P} = \mathbf{Z}^n$ or a slight generalization thereof. Importantly, the modules considered in this paper can be δ-interleaved only for $\delta \in \{0, 1, 2, \ldots\}$.

Contributions

The contributions of this paper are summarized in Table 1. Concretely, a cell in Table 1 gives a complexity bound on the decision problem of deciding if two modules of the given type are δ-interleaved. It is an easy consequence of the definition of the interleaving distance that this is at least as hard as determining the distance itself. The cells with a shaded background indicate that novel contributions to that complexity bound is provided in this paper. Recall that we have defined the input size to be $n = \dim M + \dim M'$ when the modules are valued in $\mathbf{Vec}$, and $n = |M| + |M'|$ when the modules are valued in $\mathbf{Set}$. Observe that any non-trivial functor $M : \mathbf{Z}^m \to \mathbf{Set}$ must have $|M| = \infty$. Hence, when we talk about interleavings of such functors, we shall assume that they are completely determined by a restriction to a finite sub-grid. The input size is then the total cardinalities of the restrictions. We will now give a brief summary of the cells of Table 1.

- $\mathbf{Z} \to \mathbf{Vec}$. [$\delta \geq 0$] This bound is achieved by first determining the barcodes of the persistence modules and then using Theorem 2.6 to obtain the interleaving distance. The complexity of this is $\mathcal{O}(\mathrm{FindBarcode} + \mathrm{Match}) = \mathcal{O}(n^\omega + n^{1.5}\log n) = \mathcal{O}(n^\omega)$ where ω is the matrix multiplication exponent[23]. The details can be found in Appendix C. In [26], the complexity is shown to be $\mathcal{O}(n^\omega + n^2 \log^2 n)$ for essentially the same problem, but with a slightly different input size n.

- $\mathbf{Z} \to \mathbf{Set}$. [$\delta = 0$] Essentially isomorphism of rooted trees; see Appendix E. [$\delta \geq 1$] This follows from arguments in [1].

- $\mathbf{Z}^2 \to \mathbf{Vec}$. [$\delta = 0$] This is Theorem 1.3 for $\mathbf{P} = \mathbf{Z}^2$. [$\delta \geq 1$] A *constrained invertibility* (CI) problem is a triple (P, Q, n) where P and Q are subsets of $\{1, 2, \ldots, n\}^2$. We say that a CI-problem (P, Q, n) is *solvable* if there exists an invertible $n \times n$ matrix M such that $M_{i,j} = 0$ for all $(i, j) \in P$ and $M^{-1}_{i',j'} = 0$ for all $(i', j') \in Q$. We call (M, M^{-1}) a *solution* of (P, Q, n).

type/δ	$\mathbf{Z} \to \mathbf{Vec}$	$\mathbf{Z} \to \mathbf{Set}$	$\mathbf{Z}^2 \to \mathbf{Vec}$	$\mathbf{Z}^2 \to \mathbf{Set}$	$\mathbf{Z}^{L,C} \to \mathbf{Vec}_{\mathbb{Z}/2\mathbb{Z}}$
$\delta = 0$	$\mathcal{O}(n^\omega)$	$\mathcal{O}(n)$	$\mathcal{O}(n^6)$	GI-complete	$\mathcal{O}(n^6)$
$\delta \geq 1$	$\mathcal{O}(n^\omega)$	NP-complete	CI-hard	NP-complete	NP-complete

Table 1: The complexity of checking for δ-interleavings between modules M and M'. If the target category is $\mathbf{Vec}$ then $n = \dim M + \dim M'$, and if the target category is $\mathbf{Set}$ then $n = |M| + |M'|$. Here ω is the matrix multiplication exponent.

In Section 4 we show that a CI-problem is solvable if and only if an associated pair of $\mathbf{Z}^2$-indexed modules is 1-interleaved. Thus, the interleaving problem is *constrained invertibility-hard* (CI-hard).

- $\mathbf{Z}^2 \to \mathbf{Set}$. [$\delta = 0$] Reeb graphs are a particular type of functors $\mathbf{Z}^2 \to \mathbf{Set}$ and deciding if two Reeb graphs are isomorphic is graph isomorphism-hard (GI-hard) [20]. In Appendix D we strengthen this result by showing that the isomorphism problem for $\mathbf{Z}^2 \to \mathbf{Set}$ is in fact GI-*complete*. This also implies that Reeb graph isomorphism is GI-complete. [$\delta \geq 1$] This follows from $\mathbf{Z} \to \mathbf{Set}$.

- $\mathbf{Z}^{L,C} \to \mathbf{Vec}_{\mathbb{Z}/2\mathbb{Z}}$. For two sets L and C, define $\mathbf{Z}^{L \to C}$ to be the poset generated by the following disjoint union of posets $\mathbf{Z}^{L \to C} := \bigsqcup_{l \in L, c \in C} \mathbf{Z}$ with the added relation $(l, t) < (c, t)$ for every $l \in L$, $c \in C$ and $t \geq 3$. This poset is a mild generalization of a disjoint union of $\mathbf{Z}$'s. [$\delta = 0$] Immediate from Theorem 1.3. [$\delta \geq 1$] Follows from a reduction from 3-SAT; see Section 3. This shows that computing the generalized interleaving distance of [12] for $\mathbf{Vec}$-valued persistence modules is NP-complete in general.

2 Preliminaries

For $\mathbf{P}$ a poset and $\mathcal{C}$ an arbitrary category, $M : \mathbf{P} \to \mathcal{C}$ a functor, and $a, b \in \mathbf{P}$, let $M_a = M(a)$, and let $\varphi_M(a, b) : M_a \to M_b$ denote the morphism $M(a \leq b)$.

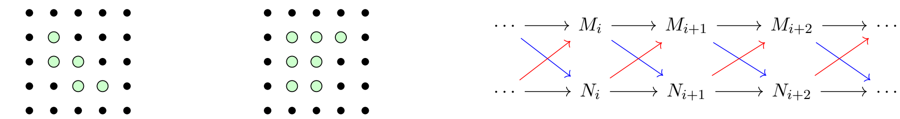
2.1 Interleavings

In this section we review the theory of interleavings for $\mathbf{Z}^n$-indexed modules. For a treatment of the $\mathbf{R}^n$-indexed setting see [25]. For a discussion on interleavings over arbitrary posets see [12].

For $u \in \mathbf{Z}^n$, define the *u-shift functor* $(-)(u) : \mathcal{C}^{\mathbf{Z}^n} \to \mathcal{C}^{\mathbf{Z}^n}$ on objects by $M(u)_a = M_{u+a}$, together with the obvious internal morphisms, and on morphisms $f : M \to N$ by $f(u)_a = f(u + a) : M(u)_a \to N(u)_a$. For $u \in \{0, 1, \ldots\}^n$, let $\varphi_M^u : M \to M(u)$ be the morphism whose restriction to each M_a is the linear map $\varphi_M(a, a + u)$. For $\delta \in \{0, 1, 2 \ldots\}$ we will abuse notation slightly by letting $(-)(\delta)$ denote the $\delta(1, \ldots, 1)$-shift functor, and letting φ_M^δ denote $\varphi_M^{\delta(1,\ldots,1)}$.

Definition 2.1. Given $\delta \in \{0, 1, \ldots\}$, a δ-interleaving between $M, N : \mathbf{Z}^n \to \mathcal{C}$ is a pair of morphisms $f : M \to N(\delta)$ and $g : N \to M(\delta)$ such that $g(\delta) \circ f = \varphi_M^{2\delta}$ and $f(\delta) \circ g = \varphi_N^{2\delta}$.

We call f and g *δ-interleaving morphisms*. If there exists a δ-interleaving between M and N, we say M and N are δ-interleaved. The *interleaving distance* $d_I : \mathrm{Ob}(\mathcal{C}^{\mathbf{Z}^n}) \times \mathrm{Ob}(\mathcal{C}^{\mathbf{Z}^n}) \to [0, \infty]$ is given by $d_I(M, N) = \min\{\delta \in \{0, 1, \ldots\} \mid M \text{ and } N \text{ are } \delta\text{-interleaved}\}$. Here we set $d_I(M, N) = \infty$ if there does not exist a δ-interleaving for any δ.



(a) An interval. (b) Not an interval. (c) A 1-interleaving.

Figure 2: (a) is an interval in $\mathbf{Z}^2$ whereas (b) is not. (c) The persistence modules M and N are 1-interleaved if and only if there exist diagonal morphisms such that the diagram in (c) commutes.

2.2 Interval Modules and the Isometry Theorem

Let $\mathcal{C} = \mathbf{Vec}$. An *interval* of a poset $\mathbf{P}$ is a subset $\mathcal{J} \subset \mathbf{P}$ such that

1. $\mathcal{J}$ is non-empty.
2. If $a, c \in \mathcal{J}$ and $a \leq b \leq c$, then $b \in \mathcal{J}$.
3. [connectivity] For any $a, c \in \mathcal{J}$, there is a sequence $a = b_0, b_1, \ldots, b_l = c$ of elements of $\mathcal{J}$ with b_i and b_{i+1} comparable for $0 \leq i \leq l-1$.

We refer to a collection of intervals in $\mathbf{P}$ as a *barcode (over* $\mathbf{P}$*)*.

Definition 2.2. For $\mathcal{J}$ an interval in $\mathbf{P}$, the interval module $I^{\mathcal{J}}$ is the $\mathbf{P}$-indexed module such that

$$I^{\mathcal{J}}_a = \begin{cases} \mathbb{K} & \text{if } a \in \mathcal{J}, \\ 0 & \text{otherwise.} \end{cases} \qquad \varphi_{I^{\mathcal{J}}}(a,b) = \begin{cases} \mathrm{id}_{\mathbb{K}} & \text{if } a \leq b \in I, \\ 0 & \text{otherwise.} \end{cases}$$

We say a persistence module M is *decomposable* if it can be written as $M \cong V \oplus W$ for non-trivial persistence modules V and W; otherwise, we say that M is *indecomposable*.

A $\mathbf{P}$-indexed module M is *interval decomposable* if there exists a collection $\mathcal{B}(M)$ of intervals in $\mathbf{P}$ such that $M \cong \bigoplus_{\mathcal{J} \in \mathcal{B}(M)} I^{\mathcal{J}}$. We call $\mathcal{B}(M)$ the *barcode* of M. This is well-defined by the Azumaya–Krull–Remak–Schmidt theorem [3].

Theorem 2.3 (Structure of 1-D Modules [19, 28])**.** *Suppose* $M : \mathbf{P} \to \mathbf{Vec}$ *for* $\mathbf{P} \in \{\mathbf{R}, \mathbf{Z}\}$ *and* $\dim M_p < \infty$ *for all* $p \in \mathbf{P}$*. Then* M *is interval decomposable.*

Remark 2.4. Such a decomposition theorem exists only for *very* special choices of $\mathbf{P}$. Two other scenarios appearing in applications are *zigzags* [8, 14] and exact bimodules [18]. The latter is a specific type of $\mathbf{R}^2$-indexed persistence modules.

Corollary 2.5. *Let* $\mathbf{P} = \bigsqcup_{i \in \Lambda} \mathbf{Z}$ *be the poset given as a disjoint union of* $\mathbf{Z}$*'s (i.e. elements in different components are incomparable). If* $M : \mathbf{P} \to \mathbf{Vec}$ *satisfies* $\dim M_p < \infty$ *for all* $p \in \mathbf{P}$*, then* M *is interval decomposable.*

Proof. Apply Theorem 2.3 to each of the components of $\mathbf{P}$ independently. This gives $\mathcal{B}(M) = \bigsqcup_{i \in \Lambda} \mathcal{B}(M|_{(i,\mathbf{Z})})$. □

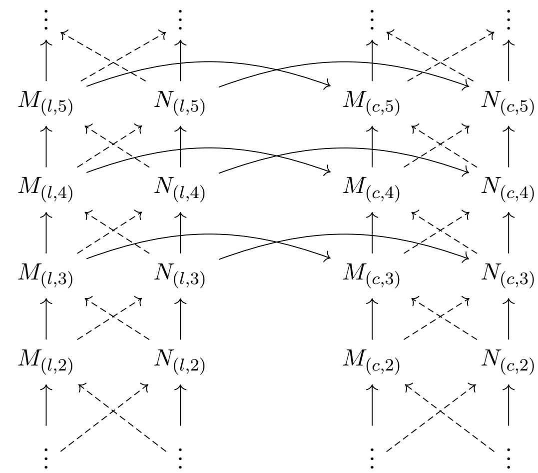
At the very core of topological data analysis are the *isometry theorems*. They say that for certain choices of interval decomposable modules, the interleaving distance coincides with a completely combinatorial distance on their associated barcodes. This combinatorial distance d_B is called the *bottleneck distance* and is defined in Appendix A. Importantly, for any two barcodes, if the interleaving distance between each pair of interval modules in the barcodes is known, the associated bottleneck distance can be computed by solving a bipartite matching problem. This, in turn, implies that the interleaving distance can be efficiently computed whenever an isometry theorem holds. See Appendix C for an example.

Theorem 2.6 (Isometry Theorem [5, 7, 17, 25])**.** *Suppose $M, N : \mathbf{Z} \to \mathbf{Vec}$ satisfy $\dim M_i < \infty$ and $\dim N_i < \infty$ for all $i \in \mathbf{Z}$. Then $d_I(M, N) = d_B(\mathcal{B}(M), \mathcal{B}(N))$.*

Remark 2.7. Continuing on the remark to Theorem 2.3. An isometry theorem also holds for zigzags and exact bimodules [7, 9]. Although there might be other classes of interval decomposable modules for which an isometry theorem holds, the result is not true in general. See Appendix F for an example of interval decomposable modules in $\mathbf{Z}^2$ for which $2d_I(M, N) = d_B(\mathcal{B}(M), \mathcal{B}(N))$, and see [9] for a general conjecture. This shows that a matching of the barcodes will not determine the interleaving distance even in the case of very well-behaved modules.

3 NP-completeness

In this section we shall prove that it is NP-hard to decide if two modules $M, N \in \mathbf{Vec}^{\mathbf{Z}^{L\to C}}$ are 1-interleaved. Recall that for two sets L and C, we define $\mathbf{Z}^{L\to C}$ to be the disjoint union $\bigsqcup_{l\in L, c\in C} \mathbf{Z}$ with the added relations $(l, t) < (c, t)$ for all $l \in L, c \in C$, and $t \geq 3$. Define the *u-shift functor* $(-)(u) : \mathcal{C}^{\mathbf{Z}^{L\to C}} \to \mathcal{C}^{\mathbf{Z}^{L\to C}}$ on objects by $M(u)_{(p,t)} = M_{(p,t+u)}$, together with the obvious internal morphisms, and on morphisms $f : M \to N$ by $f(u)_{(p,t)} = f_{(p,t+u)} : M(u)_{(p,t)} \to N(u)_{(p,t)}$. That is, the shift functor simply acts on each of the components independently. With the shift-functor defined, we define a δ-interleaving of $\mathbf{Z}^{L\to C}$-indexed modules precisely as in Section 2.1. Thus, we see that a δ-interleaving is simply a collection of δ-interleavings over each disjoint component of $\mathbf{Z}$ which satisfy the added relations. Indeed, a 1-interleaving is equivalent to the existence of dashed morphisms in the following diagram for all $l \in L$ and $c \in C$:



We saw in Corollary 2.5 that $M : \mathbf{Z}^{L \to \emptyset} \to \mathbf{Vec}$ is interval decomposable. By applying Theorem 2.6 to each disjoint component independently, the following is easy to show. Here the bottleneck distance is generalized in the obvious way, i.e. matching each component independently.

Corollary 3.1 (Isometry Theorem for Disjoint Unions)**.** *Let L be any set, and $M, N : \mathbf{Z}^{L \to \emptyset} \to \mathbf{Vec}$ such that $\dim M_p < \infty$ and $\dim N_p < \infty$ for all $p \in \mathbf{Z}^{L \to \emptyset}$. Then $d_I(M, N) = d_B(\mathcal{B}(M), \mathcal{B}(N))$.*

In particular, the interleaving distance between M and N can be effectively computed through a bipartite matching. As we shall see, this is not true for $C \neq \emptyset$. The remainder of this section is devoted to proving the following theorem:

Theorem 3.2. *Unless P=NP, there exists no algorithm, polynomial in $n = \dim M + \dim N$, which decides if $M, N : \mathbf{Z}^{L \to C} \to \mathbf{Vec}_{\mathbb{Z}/2\mathbb{Z}}$ are 1-interleaved.*

3.1 The Proof

We shall prove Theorem 3.2 by a reduction from 3-SAT. Let ψ be a boolean formula in 3-CNF defined on literals $L = \{x_1, x_2, \ldots, x_{n_l}\}$ and clauses $C = \{c_1, c_2, \ldots, c_{n_c}\}$. We shall assume that the literals of each clause are distinct and *ordered*. That is, the clause c_i is specified by the three distinct literals $\{x_{i_1}, x_{i_2}, x_{i_3}\}$ wherein $i_1 < i_2 < i_3$. Determining if ψ is satisfiable is well-known to be NP-complete. For the entirety of the proof $\mathbb{K} = \mathbb{Z}/2\mathbb{Z}$.

Step 1: Defining the representations. Associate to ψ two functors $M, N : \mathbf{Z}^{L \to C} \to \mathbf{Vec}_{\mathbb{Z}/2\mathbb{Z}}$ in the following way: For all literals $x_j \in L$ define

$$M_{(x_j,1)} \longrightarrow M_{(x_j,2)} \longrightarrow M_{(x_j,3)} \longrightarrow M_{(x_j,4)} = \mathbb{K} \xrightarrow{1} \mathbb{K} \xrightarrow{1} \mathbb{K} \xrightarrow{1} \mathbb{K}$$

$$N_{(x_j,1)} \longrightarrow N_{(x_j,2)} \longrightarrow N_{(x_j,3)} \longrightarrow N_{(x_j,4)} = \mathbb{K} \xrightarrow{1} \mathbb{K} \xrightarrow{(1;1)} \mathbb{K}^2 \longrightarrow 0$$

and for every clause c_i in ψ define

$$M_{(c_i,1)} \longrightarrow M_{(c_i,2)} \longrightarrow M_{(c_i,3)} \longrightarrow M_{(c_i,4)} = 0 \longrightarrow \mathbb{K} \xrightarrow{1} \mathbb{K} \xrightarrow{1} \mathbb{K}$$

$$N_{(c_i,1)} \longrightarrow N_{(c_i,2)} \longrightarrow N_{(c_i,3)} \longrightarrow N_{(c_i,4)} = 0 \longrightarrow \mathbb{K}^3 \xrightarrow{1} \mathbb{K}^3 \longrightarrow 0$$

For any other $p \in \mathbf{Z}^{L \to C}$, $M_p = N_p = 0$. Next we specify the remaining non-trivial morphisms: let $c_i = z_{i_1} \vee z_{i_2} \vee z_{i_3}$ be a clause in ψ, where $z_{j_l} = x_{j_l}$ or $z_{j_l} = \neg x_{j_l}$, and for which $i_1 < i_2 < i_3$. For $s = 1, 2, 3$ define $H_s : \mathbb{K}^2 \to \mathbb{K}^3$ by $e_1 \to u \cdot e_s$ and $e_2 \to (1-u) \cdot e_s$, where $u = 1$ if $z_{i_s} = x_{i_s}$ and $u = 0$ if $z_{i_s} = \neg x_{i_s}$. Here e_d is the d-th standard basis vector of $\mathbb{K}^3$. Given this we define the following for $s = 1, 2, 3$:

$$\begin{tikzcd}
M_{(x_{i_s},4)} \arrow[r] & M_{(c_i,4)} & = & \mathbb{K} \arrow[r, "1"] & \mathbb{K} \\
M_{(x_{i_s},3)} \arrow[r] \arrow[u] & M_{(c_i,3)} \arrow[u] & = & \mathbb{K} \arrow[r, "1"] \arrow[u, "1"] & \mathbb{K} \arrow[u, "1"]
\end{tikzcd}$$

and

$$N_{(x_{i_s},3)} \longrightarrow N_{(c_i,3)} = \mathbb{K}^2 \xrightarrow{H_s} \mathbb{K}^3.$$

Clearly $\dim M + \dim N = \mathcal{O}(n_c + n_l)$. Thus, the total dimension is polynomial in the input size of 3-SAT.

Step 2: Showing the reduction. Observe that M and N are 1-interleaved if and only if there exist dashed morphisms such that the below diagram is commutative for every literal x_{i_s} and for every clause c_i containing x_{i_s}:

$$\begin{tikzcd}
M_{(x_{i_s},5)} = 0 & 0 = N_{(x_{i_s},5)} & M_{(c_i,5)} = 0 & 0 = N_{(c_i,5)} \\
M_{(x_{i_s},4)} = \mathbb{K} \arrow[u] \arrow[ur, dashed] \arrow[rr, bend left] & 0 = N_{(x_{i_s},4)} \arrow[u] \arrow[ul, dashed] \arrow[rr, bend left] & M_{(c_i,4)} = \mathbb{K} \arrow[u] \arrow[ur, dashed] & 0 = N_{(c_i,4)} \arrow[u] \arrow[ul, dashed] \\
M_{(x_{i_s},3)} = \mathbb{K} \arrow[u] \arrow[ur, dashed] \arrow[rr, bend left] & \mathbb{K}^2 = N_{(x_{i_s},3)} \arrow[u] \arrow[ul, dashed] \arrow[rr, bend left, "H_s"] & M_{(c_i,3)} = \mathbb{K} \arrow[u] \arrow[ur, dashed] & \mathbb{K}^3 = N_{(c_i,3)} \arrow[u] \arrow[ul, dashed] \\
M_{(x_{i_s},2)} = \mathbb{K} \arrow[u] \arrow[ur, dashed] & \mathbb{K} = N_{(x_{i_s},2)} \arrow[u, "(1;1)"'] \arrow[ul, dashed] & M_{(c_i,2)} = \mathbb{K} \arrow[u] \arrow[ur, dashed] & \mathbb{K}^3 = N_{(c_i,2)} \arrow[u, "I_3"'] \arrow[ul, dashed] \\
M_{(x_{i_s},1)} = \mathbb{K} \arrow[u] \arrow[ur, dashed] & \mathbb{K} = N_{(x_{i_s},1)} \arrow[u] \arrow[ul, dashed] & M_{(c_i,1)} = 0 \arrow[u] & 0 = N_{(c_i,1)} \arrow[u]
\end{tikzcd} \tag{1}$$

We shall see there are few degrees of freedom in the choice of interleaving morphisms. Indeed, consider the left part of the above diagram:

$$
\begin{tikzcd}
M & 0 \arrow[r] & \mathbb{K} \arrow[r, "1"] \arrow[dr] & \mathbb{K} \arrow[r, "1"] \arrow[dr] & \mathbb{K} \arrow[r, "1"] \arrow[dr] & \mathbb{K} \arrow[r, "1"] \arrow[dr] & 0 \\
N & 0 \arrow[r] & \mathbb{K} \arrow[r, "1"'] \arrow[ur] & \mathbb{K} \arrow[r, "(1;1)"'] \arrow[ur] & \mathbb{K}^2 \arrow[r] \arrow[ur, dashed] & 0 \arrow[r] \arrow[ur] & 0
\end{tikzcd}
$$

We leave it to the reader to verify that if M and N are 1-interleaved, then all the solid diagonal morphisms in the above diagram are completely determined by commutativity. For the dashed morphism $(\varphi_{x_{i_s}}, \varphi_{\neg x_{i_s}}) : \mathbb{K}^2 \to \mathbb{K}$ there are two choices: by commutativity it must satisfy $(\varphi_{x_{i_s}}, \varphi_{\neg x_{i_s}}) \cdot (1;1) = 1$ and thus $\varphi_{x_{i_s}} + \varphi_{\neg x_{i_s}} = 1$. As $\mathbb{K} = \mathbb{Z}/2\mathbb{Z}$, this implies that precisely one of $\varphi_{x_{i_s}}$ and $\varphi_{\neg x_{i_s}}$ is multiplication by 1. This corresponds to a choice of truth value for x_{i_s}: $\varphi_{x_{i_s}} = 1 \iff x_{i_s} = \text{True}$ and $\varphi_{\neg x_{i_s}} = 1 \iff x_{i_s} = \text{False}$. Next, consider the right part of 1:

$$
\begin{tikzcd}
M & 0 \arrow[r] & \mathbb{K} \arrow[r, "1"] \arrow[dr, dashed] & \mathbb{K} \arrow[r, "1"] \arrow[dr] & \mathbb{K} \arrow[r] \arrow[dr] & 0 \\
N & 0 \arrow[r] & \mathbb{K}^3 \arrow[r, "1"'] \arrow[ur, dashed] & \mathbb{K}^3 \arrow[r] \arrow[ur, dashed] & 0 \arrow[r] \arrow[ur] & 0
\end{tikzcd}
$$

There are three non-trivial morphisms, out of which two are equal by commutativity. Let $Z_1^i : \mathbb{K} \to \mathbb{K}^3$ and $Z_2^i : \mathbb{K}^3 \to \mathbb{K}$ denote the two unspecified morphisms. Returning to (1), we see that Z_2^i must satisfy the following for $s \in \{1,2,3\}$:

$$
\begin{tikzcd}
\mathbb{K} \arrow[rr, "1"] & & \mathbb{K} \\
& \mathbb{K}^2 \arrow[ul, "(\varphi_{x_{i_s}}, \varphi_{\neg x_{i_s}})"] \arrow[rr, "H_s"] & & \mathbb{K}^3 \arrow[ul, dashed, "Z_2^i"']
\end{tikzcd}
$$

Thus, Z_2^i restricted to its s-th component equals either $\varphi_{x_{i_s}}$ or $\varphi_{\neg x_{i_s}}$, depending on whether x_{i_s} or its negation $\neg x_{i_s}$ appears in the clause c_i. This implies that Z_2^i is given by

$$
Z_2^i = \begin{bmatrix} \varphi_{z_{i_1}} & \varphi_{z_{i_2}} & \varphi_{z_{i_3}} \end{bmatrix}
$$

Hence, if M and N are to be 1-interleaved, then there are no degrees of freedom in choosing Z_2^i after the $\varphi_{x_{i_s}}$ are specified. However, Z_1^i only needs to satisfy $Z_2^i \circ Z_1^i = 1$. As this is the sole restriction imposed on Z_1^i, we see that this can be satisfied if and only if $Z_2^i \neq 0$, which is true if and only if $z_{i_s} = \text{True}$ for at least one $s \in \{1,2,3\}$.

Theorem 3.3. *Let ψ be a boolean formula as above. Then ψ is satisfiable if and only if the associated persistence modules $M, N : \mathbf{Z}^{L \to C} \to \mathbf{vec}$ are 1-interleaved.*

Proof. Summarizing the above: we have that M and N are 1-interleaved if and only if we can choose morphisms $(\varphi_{x_{i_s}}, \varphi_{\neg x_{i_s}})$ such that $Z_2^i \neq 0$ for all clauses c_i. This means precisely that we can choose truth values for each x_{i_s} such that every clause $c_i = z_{i_1} \vee z_{i_2} \vee z_{i_3}$ evaluates to true. This shows that a 1-interleaving implies that ψ is satisfiable. Conversely, if ψ is satisfiable, then we see that the morphisms defined by $\varphi_{x_{i_s}} = 1 \iff x_{i_s} = \text{True}$ and $\varphi_{\neg x_{i_s}} = 1 \iff x_i = \text{False}$ satisfy $Z_2^i \neq 0$ for every clause c_i. Thus, M and N are 1-interleaved. $\square$

Remark 3.4. Let $i : \mathbf{P} \hookrightarrow \mathbf{Q}$ be an inclusion of posets and $M : \mathbf{P} \to \mathbf{Vec}$. There are multiple functorial ways of extending M to a representation $E(M) : \mathbf{Q} \to \mathbf{Vec}$, e.g. by means of left or right Kan extensions. This is a key ingredient in one of the more recent proofs of Theorem 2.6; see [13] for details. However, if we impose the condition that $E(M) \circ i \cong M$ then such an extension need not exist. Indeed, Theorem 3.3 implies that the associated decision problem is NP-complete.

4 Interleavings of Multidimensional Persistence Modules

Recall that a constrained invertibility (CI) problem is a triple (P, Q, n) where P and Q are subsets of $\{1, 2, \dots, n\}^2$, and that a CI-problem is solvable if there exists an invertible $n \times n$ matrix M such that $M_{(i,j)} = 0$ for all $(i, j) \in P$ and $M^{-1}_{(i',j')} = 0$ for all $(i', j') \in Q$. We shall show that a CI-problem is solvable if and only if a pair of associated persistence modules $\mathbf{Z}^2 \to \mathbf{Vec}$ is 1-interleaved. Hence, if deciding solvability is NP-hard, then so is computing the interleaving distance for multidimensional persistence modules.

Example 4.1. Let $P = \{(2, 2), (3, 3)\}, Q = \{(2, 3), (3, 2)\} \subset \{1, 2, 3\}^2$. Then $(P, Q, 3)$ is solvable by

$$M = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \quad M^{-1} = \begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}.$$

Example 4.2. Let $P = \{(1, 1), (1, 3)\}, Q = \{(2, 1)\} \subset \{1, 2, 3\}^2$. Then $(P, Q, 3)$ is not solvable, as $(MN)_{(1,1)} = 0$ for all 3×3-matrices M, N with $M_{(1,1)} = M_{(1,3)} = N_{(2,1)} = 0$. Note that it matters that we view P and Q as subsets of $\{1, 2, 3\}^2$ and not of $\{1, \dots, n\}^2$ for some $n > 3$, in which case (P, Q) would be solvable.

Example 4.3. Observe that a CI-problem $(P, \emptyset, n)$ reduces to a bipartite matching problem. Build a graph G on $2n$ vertices $\{v_1, \dots, v_n, u_1, \dots, u_n\}$ with an edge from v_i to u_j if $(i, j) \notin P$. Then the CI-problem is solvable if and only if there exists a perfect matching of G.

A CI-problem can be seen as a problem of choosing weights for the edges in a directed simple graph: Given (P, Q, n), let G be the bipartite directed simple graph with vertices $\{u_1, \dots, u_n, v_1, \dots, v_n\}$, an edge from u_i to v_j if $(i, j) \notin P$, and an edge from v_j to u_i if $(j, i) \notin Q$. Solving (P, Q, n) is then equivalent to weighting the edges in G with elements from $\mathbb{K}$ so that

$$\sum_{j=1}^{n} w(u_i, v_j) w(v_j, u_i) = 1$$

for all i, and

$$\sum_{j=1}^{n} w(u_i, v_j) w(v_j, u_{i'}) = 0$$

for all $i \neq i'$, where $w(u, v)$ is the weight of the edge from u to v if there is one, and 0 if not. If the weights are elements of $\mathbb{Z}/2\mathbb{Z}$, this is equivalent to picking a subset of the edges such that there is an odd number of paths of length two from any vertex to itself and an even number of paths of length two from any vertex to any other vertex.

Fix a CI-problem (P, Q, n) and let $m = |P| + |Q|$. We will construct $\mathbf{Z}^2$-indexed modules M and N that are 1-interleaved if and only if (P, Q, n) is solvable, and that are zero outside a grid of size $(2m+3) \times (2m+3)$ in $\mathbf{Z}^2$. The dimension of each vector space $M_{(a,b)}$ or $N_{(a,b)}$ is bounded by n, so the total dimensions of M and N are polynomial in n.

For $p \in \mathbf{Z}^2$, let $\langle p \rangle = \{q \in \mathbf{Z}^2 \mid p \leq q \leq (2m+2, 2m+2)\}$. Let $\mathcal{W}$ be the interval $\bigcup_{k=0}^{m} \langle (2m-2k, 2k) \rangle$, and for $i \in \{1, 2, \ldots, m\}$, let $x_i = (2m-2i+1, 2i-1)$; see Fig. 3.

Write $P = \{(p_1, q_1), \ldots, (p_r, q_r)\}$ and $Q = \{(p_{r+1}, q_{r+1}), \ldots, (p_m, q_m)\}$. We define $M = \bigoplus_{i=1}^{n} I^{\mathcal{I}_i}$ and $N = \bigoplus_{i=1}^{n} I^{\mathcal{J}_i}$, where $\mathcal{I}_i$ and $\mathcal{J}_i$ are constructed as follows: let $\mathcal{I}_i^0 = \mathcal{J}_i^0 = \mathcal{W}$ for all i. For $k = 1, 2, \ldots, r$, let

$$\mathcal{I}_i^k = \begin{cases} \mathcal{I}_i^{k-1} \cup \langle x_k - (1,1) \rangle, & \text{if } i = p_k \\ \mathcal{I}_i^{k-1} \cup \langle x_k \rangle, & \text{if } i \neq p_k \end{cases}, \qquad \mathcal{J}_i^k = \begin{cases} \mathcal{J}_i^{k-1}, & \text{if } i = q_k \\ \mathcal{J}_i^{k-1} \cup \langle x_k \rangle, & \text{if } i \neq q_k \end{cases}$$

and for $k = r+1, \ldots, m$, let

$$\mathcal{I}_i^k = \begin{cases} \mathcal{I}_i^{k-1}, & \text{if } i = q_k \\ \mathcal{I}_i^{k-1} \cup \langle x_k \rangle, & \text{if } i \neq q_k \end{cases}, \qquad \mathcal{J}_i^k = \begin{cases} \mathcal{J}_i^{k-1} \cup \langle x_k - (1,1) \rangle, & \text{if } i = p_k \\ \mathcal{J}_i^{k-1} \cup \langle x_k \rangle, & \text{if } i \neq p_k \end{cases}$$

and let $\mathcal{I}_i = \mathcal{I}_i^m$ and $\mathcal{J}_i = \mathcal{J}_i^m$. This way, we ensure that there is no nonzero morphism from $I^{\mathcal{I}_i}$ to $I^{\mathcal{J}_j}(1)$ when $(i,j) \in P$, and no nonzero morphism from $I^{\mathcal{J}_j}$ to $I^{\mathcal{I}_i}(1)$ when $(j,i) \in Q$. In all other cases, there exist nonzero morphisms.

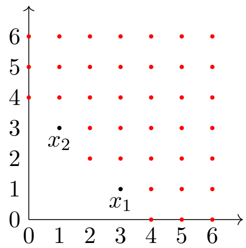


Figure 3: The interval $\mathcal{W}$ for $m = 2$ along with $x_1 = (3,1)$ and $x_2 = (1,3)$.

Lemma 4.4. *Suppose $(i,j) \notin P$. Then there is an isomorphism $\mathrm{Hom}(I^{I_i}, I^{\mathcal{J}_j}(1)) \cong \mathbb{K}$. In particular, any morphism $f \in \mathrm{Hom}(I^{I_i}, I^{\mathcal{J}_j}(1))$ is completely determined by $f_{(2m+1,2m+1)}$: if f_p is nonzero, then $f_p = f_{(2m+1,2m+1)}$.*

The same holds if $(j,i) \notin Q$ instead of $(i,j) \notin P$, and $\mathcal{I}_i$ and $\mathcal{J}_j$ are interchanged. As f_p is a $\mathbb{K}$-endomorphism, this implies that any f can be identified with an element of $\mathbb{K}$.

Proof. Let $f : I^{\mathcal{I}_i} \to I^{\mathcal{J}_j}(1)$ be nonzero. If $p \notin \mathcal{I}_i$ or $p \not\leq (2m+1, 2m+1)$, $f_p = 0$. For $(2m+1, 2m+1) \geq p \in \mathcal{I}_i$, we have $p + (1,1) \in \mathcal{J}_i$ by construction and the fact that $(i,j) \notin P$, so $\varphi_{I^{\mathcal{J}_j}}(p+(1,1), (2m+2, 2m+2))$ is nonzero and hence the identity. We get

$$\begin{aligned} f_p &= \varphi_{I^{\mathcal{J}_j}(1)}(p, (2m+1, 2m+1)) \circ f_p \\ &= f_{(2m+1,2m+1)} \circ \varphi_{I^{\mathcal{I}_i}}(p, (2m+1, 2m+1)) = f_{(2m+1,2m+1)}. \end{aligned}$$

□

Describing a morphism from $M = \bigoplus_{i=1}^{n} I^{\mathcal{I}_i}$ to $N(1) = \bigoplus_{j=1}^{n} I^{\mathcal{J}_j}(1)$ is the same as describing morphisms from $I^{\mathcal{I}_i}$ to $I^{\mathcal{J}_j}(1)$ for all i and j[1]. We have just proved that these can be identified with elements of $\mathbb{K}$, so we conclude that any $f : M \to N(1)$ is uniquely defined by an $n \times n$-matrix A_f where the entry (i,j) is the element in $\mathbb{K}$ corresponding to the morphism $I^{\mathcal{I}_i} \to I^{\mathcal{J}_j}(1)$ given by f. Note that we get the same result by writing $f_{(2m,2m)} = f_{(2m+1,2m+1)} : \mathbb{K}^n \to \mathbb{K}^n$ as a matrix,

[1] $\mathrm{Hom}(\oplus_i M_i, \oplus_j N_j) \cong \oplus_i \oplus_j \mathrm{Hom}(M_i, N_j)$.

where each copy of $\mathbb{K}$ in the domain and codomain comes from one of the interval modules $I^{\mathcal{I}_i}$ or $I^{\mathcal{J}_j}(1)$, respectively.

If we also have a morphism $g : N \to M(1)$, we can define a matrix A_g symmetrically, and similarly A_g is $g_{(2m,2m)} = g_{(2m+1,2m+1)} : \mathbb{K}^n \to \mathbb{K}^n$ in matrix form.

Theorem 4.5. *With f and g as above, (f, g) is a 1-interleaving if and only if A_f and A_g are inverse matrices.*

Proof. Suppose (f, g) is a 1-interleaving. The internal morphism $\varphi_M((2m, 2m), (2m + 2, 2m + 2))$ is the identity on $\mathbb{K}^n$, and is by definition of interleaving the same as

$$g(1)_{(2m,2m)} \circ f_{(2m,2m)} = g_{(2m+1,2m+1)} \circ f_{(2m,2m)} = A_g A_f.$$

Thus $A_g A_f$ is the identity matrix, and so A_f and A_g are inverses of each other, as both are $n \times n$-matrices.

Suppose A_f and A_g are inverse matrices. We must check that at every point $p \in \mathbb{Z}^2$, $\varphi_M(p, p + (2, 2)) = g(1)_p \circ f_p$. If $p \nleq (2m, 2m)$ or $M(p) = 0$, both sides are zero. If $p \leq (2m, 2m)$ and $M(p) \neq 0$,

$$g(1)_p \circ f_p = \varphi_M(p + (2, 2), (2m + 2, 2m + 2)) \circ g(1)_p \circ f_p,$$

since $\varphi_M(p + (2, 2), (2m + 2, 2m + 2))$ must be the identity by construction of M and the fact that $M(p) \neq 0$. This is equal to

$$\begin{aligned}
&\varphi_M(p + (2, 2), (2m + 2, 2m + 2)) \circ g_{p+(1,1)} \circ f_p \\
&= g_{(2m+1,2m+1)} \circ \varphi_N(p + (1, 1), (2m + 1, 2m + 1)) \circ f_p \\
&= g_{(2m+1,2m+1)} \circ f_{(2m,2m)} \circ \varphi_M(p, (2m, 2m)) \\
&= A_g A_f \circ \varphi_M(p, (2m, 2m)) = \varphi_M(p, (2m, 2m)) = \varphi_M(p, p + (2, 2)).
\end{aligned}$$

□

We have proved that defining morphisms $f : M \to N(1)$ and $g : N \to M(1)$ is the same as choosing $n \times n$-matrices A_f and A_g such that the entries corresponding to the elements of P and Q are zero, and that (f, g) is a 1-interleaving if and only if A_f and A_g are inverse matrices. Thus M and N are 1-interleaved if and only if the CI-problem (P, Q, n) is solvable.

We constructed M and N by setting all the interval modules comprising M and N equal to $I^{\mathcal{W}}$, then modifying them in m steps each, where the complexity of each step is clearly polynomial in n. Thus the complexity of constructing M and N is polynomial in n, and so are the total dimensions of M and N. Taking n^2 as the input complexity of solving a CI-problem (P, Q, n), we have proved a reduction implying the following theorem:

Theorem 4.6. *Determining the interleaving distance for modules $\mathbf{Z}^2 \to \mathbf{Vec}$ is CI-hard.*

Remark 4.7. We give an example in Appendix F of a CI-problem whose associated matrices M and N satisfy $d_I(M, N) = 1$ and $d_B(\mathcal{B}(M), \mathcal{B}(N)) = 2$. This shows that it is not enough to find the bottleneck distance of the barcodes of M and N to decide whether M and N are 1-interleaved and thus whether the CI-problem is solvable. In fact, recent work shows that d_B can be efficiently computed [21].

We end this paper with the somewhat surprising observation that the interleaving distance of the above interval decomposable modules depends the characteristic char($\mathbb{K}$) of the underlying field $\mathbb{K}$. That is, let $M, N : \mathbf{Z}^2 \to \mathbf{Vec}_{\mathbb{K}}$, $M', N' : \mathbf{Z}^2 \to \mathbf{Vec}_{\mathbb{K}'}$, $\mathbb{K} \neq \mathbb{K}'$, and for which $\mathcal{B}(M) = \mathcal{B}(M')$ and $\mathcal{B}(N) = \mathcal{B}(N')$. Clearly, any matching distance d would satisfy $d(\mathcal{B}(M), \mathcal{B}(N)) = d(\mathcal{B}(M'), \mathcal{B}(N'))$, but it is not always true that $d_I(M, N) = d_I(M', N')$.

For a fixed $n \geq 2$, let $Q = \{(2,2), \dots, (n+2, n+2)\}$ and $P = \{(1,1)\} \cup \{2, \dots, n+2\}^2 \setminus Q$. Then the CI-problem $(P, Q, n+2)$ is solvable if and only if the characteristic of $\mathbb{K}$ divides n. We will only prove this for $n = 2$ for clarity, but the argument easily generalizes to all n.

Assume that (M, M^{-1}) is a solution to $(P, Q, 4)$:

$$M = \begin{bmatrix} 0 & ? & ? & ? \\ a & ? & 0 & 0 \\ b & 0 & ? & 0 \\ c & 0 & 0 & ? \end{bmatrix}, \quad M^{-1} = \begin{bmatrix} ? & d & e & f \\ ? & 0 & ? & ? \\ ? & ? & 0 & ? \\ ? & ? & ? & 0 \end{bmatrix}$$

Here we have put the entries corresponding to the elements of P and Q equal to 0, and left the rest as unknown. The entries we will use in the calculations that follow are labeled a, b, c, d, e, f. We see that $(MM^{-1})_{(2,2)} = ad$, $(MM^{-1})_{(3,3)} = be$, $(MM^{-1})_{(4,4)} = cf$, that is, $ad = be = cf = 1$. At the same time, $(M^{-1}M)_{(1,1)} = ad + be + cf$, so we get $1 = 1 + 1 + 1$, or $2 = 0$. Thus char($\mathbb{K}$) $= 2$, and in this case we can put all the unknowns in M and M^{-1} above equal to 1 to obtain a solution. (For $n > 2$, we put the nonzero elements on the diagonal of M equal to -1.)

Our motivation for introducing CI-problems was working towards determining the computational complexity of calculating the interleaving distance. While the last examples say little about complexity, they illustrate the underlying philosophy of our approach: By considering CI-problems, we can avoid the confusing geometric aspects of persistence modules and interleavings. E.g., in the case above, working with persistence modules over a 23×23 size grid is reduced to looking at a pair of 4×4-matrices.

5 Discussion

The problem of determining the computational complexity of computing the interleaving distance for multidimensional persistence modules (valued in **Vec**) was first brought up in Lesnick's thesis [24]. Although it has been an important open question for several years, a non-trivial lower bound on the complexity class has not yet been given. In light of Theorem 1.3, one might hope that tools from computational algebra can be efficiently extended to the setting of interleavings. Theorem 3.2 is an argument against this, as it shows that the problem of computing the interleaving distance is NP-hard in general. This leads us to conjecture that the problem of computing the interleaving distance for multidimensional persistence modules is also NP-hard. Unfortunately, writing down the conditions for an interleaving becomes intractable already for small grids. To make the decision problem more accessible to researchers in other fields of mathematics and computer science, we have shown that the problem is at least as hard as an easy to state matrix invertibility problem. We speculate that this problem is also NP-hard. If that is not the case, then an algorithm would provide valuable insight into the interleaving problem for interval decomposable modules.

A Bottleneck Distance

A *matching* σ between multisets S and T (written as $\sigma : S \nrightarrow T$) is a bijection $\sigma : S \supseteq S' \to T' \subset T$. Formally, we regard σ as a relation $\sigma \subset S \times T$ where $(s, t) \in \sigma$ if and only if $s \in S'$ and $\sigma(s) = t$.

We call S' and T' the *coimage* and *image* of σ, respectively, and denote them by $\operatorname{coim}\sigma$ and $\operatorname{im}\sigma$. If $w \in \operatorname{coim}\sigma \cup \operatorname{im}\sigma$, we say that σ *matches* w.

We say intervals $\mathcal{J}, \mathcal{K} \subset \mathbf{Z}^n$ are δ-interleaved if $I^{\mathcal{J}}$ and $I^{\mathcal{K}}$ are δ-interleaved. Similarly, we say $\mathcal{J}$ is 2δ-trivial if $I^{\mathcal{J}}$ is δ-interleaved with the 0-module, i.e. the module $I^{\emptyset}$. For $\mathcal{C}$ a barcode over $\mathbf{Z}^n$ and $\delta \geq 0$, define $\mathcal{C}_\delta \subset \mathcal{C}$ to be the multiset of intervals in $\mathcal{C}$ that are not δ-trivial.

Define a *δ-matching* between barcodes $\mathcal{C}$ and $\mathcal{D}$ to be a matching $\sigma : \mathcal{C} \nrightarrow \mathcal{D}$ satisfying the following properties:

1. $\mathcal{C}_{2\delta} \subset \operatorname{coim}\sigma$ and $\mathcal{D}_{2\delta} \subset \operatorname{im}\sigma$.
2. If $\sigma(\mathcal{J}) = \mathcal{K}$, then $\mathcal{J}$ and $\mathcal{K}$ are δ-interleaved.

For barcodes $\mathcal{C}$ and $\mathcal{D}$, we define the bottleneck distance d_B by

$$d_B(\mathcal{C}, \mathcal{D}) = \min\{\delta \in \{0, 1, 2, \ldots\} \mid \exists \text{ a } \delta\text{-matching between } \mathcal{C} \text{ and } \mathcal{D}\}.$$

It is not hard to check that d_B is an extended pseudometric. In particular, it satisfies the triangle inequality.

B Discrete Modules

We define an *(injective) n-D grid* to be a function $\mathcal{G} : \mathbb{Z}^n \to \mathbb{R}^n$ given by

$$\mathcal{G}(z_1, \ldots, z_n) = (\mathcal{G}_1(z_1), \ldots, \mathcal{G}_n(z_n))$$

for strictly increasing functions $\mathcal{G}_i : \mathbb{Z} \to \mathbb{R}$ with $\lim_{i\to-\infty} = -\infty$ and $\lim_{i\to\infty} = \infty$.

Define $\mathrm{fl}_{\mathcal{G}} : \mathbb{R}^n \to \mathrm{im}(\mathcal{G})$ by $\mathrm{fl}_{\mathcal{G}}(t) = \max\{s \in \mathrm{im}(\mathcal{G}) \mid s \leq t\}$.

For $\mathcal{G}$ an n-D grid, we let $E_{\mathcal{G}} : \mathcal{C}^{\mathbf{Z}^n} \to \mathcal{C}^{\mathbf{R}^n}$:

1. For M a $\mathbf{Z}^n$-indexed persistence module and $a, b \in \mathbf{R}^n$,

$$E_{\mathcal{G}}(M)_a = M_y, \qquad \varphi_{E_{\mathcal{G}}(M)}(a, b) = \varphi_M(y, z),$$

   where $y, z \in \mathbf{Z}^2$ are given by $\mathcal{G}(y) = \mathrm{fl}_{\mathcal{G}}(a)$ and $\mathcal{G}(z) = \mathrm{fl}_{\mathcal{G}}(b)$.

2. The action of $E_{\mathcal{G}}$ on morphisms is the obvious one.

Let

$$(-)|_{\mathcal{G}} : \mathcal{C}^{\mathbf{R}^n} \to \mathcal{C}^{\mathbf{Z}^n}$$

denote the restriction along $\mathcal{G}$.

We say that $M : \mathbf{R}^n \to \mathcal{C}$ is *discrete* if there exists an n-grid $\mathcal{G}$ such that $M \cong E_{\mathcal{G}}(M|_{\mathcal{G}})$. Clearly, if M and N are discrete then we may choose a grid $\mathcal{G}$ such that $M \cong E_{\mathcal{G}}(M|_{\mathcal{G}})$ and $N \cong E_{\mathcal{G}}(N|_{\mathcal{G}})$.

C Interleavings of Functors $\mathbf{Z} \to \mathbf{Vec}$

It is well-known [15] and easy to see that a persistence module $M : \mathbf{Z} \to \mathbf{Vec}$ is completely determined by its associated *rank invariant* rk_M,

$$\mathrm{rk}_M(a, b) = \mathrm{rank}(\varphi_M(a, b)), \qquad a \leq b \in \mathbf{Z}.$$

The rank of an $m_1 \times m_2$-matrix can be calculated in $\mathcal{O}(m_1 m_2^{\omega-1})$ [22], where ω is the matrix multiplication exponent. Let $d_i = \dim M_i$ and $d = \dim M = \sum_i d_i$, and assume that we are given a list of all i such that M_i is nonzero. The cost of calculating $\mathrm{rk}_M(i,j)$ for all pairs $i<j$ in the list is at most

$$\sum_{i<j} C d_i d_j^{\omega-1} \le C\left(\sum_i d_i\right)\left(\sum_i d_i^{\omega-1}\right) \le C\left(\sum_i d_i\right)\left(\sum_i d_i\right)^{\omega-1} \le C\left(\sum_i d_i\right)^{\omega} = Cd^{\omega}$$

for a sufficiently large constant C. This shows that the complexity of computing rk_M is $\mathcal{O}(d^\omega)$. Note that the number of intervals $[a,b]$ in the barcode $\mathcal{B}(M)$ is $\mathrm{rk}_M(a,b)-\mathrm{rk}_M(a-1,b)-\mathrm{rk}_M(a,b+1)+\mathrm{rk}_M(a-1,b+1)$. Thus, once we got the rank invariant, we can extract $\mathcal{B}(M)$ in $\mathcal{O}(d^2)$ operations. In conclusion, we have provided an algorithm which computes $\mathcal{B}(M)$ from M in $\mathcal{O}(d^\omega+d^2)=\mathcal{O}(d^\omega)$ operations.

Observe that $|\mathcal{B}(M)|\le \dim M$. Now, assume that we are given barcodes $\mathcal{B}(M)$ and $\mathcal{B}(N)$, with $n=\dim M+\dim N$, and we want to decide if M and N are δ-interleaved. By Theorem 2.6, this is equivalent to deciding if there is a δ-matching between $\mathcal{B}(M)$ and $\mathcal{B}(N)$, which can be done in $\mathcal{O}(b^{1.5}\log b)$, where $b=|\mathcal{B}(M)|+|\mathcal{B}(N)|\le n$ [23]. Thus, we can decide if $M,N:\mathbf{Z}\to\mathbf{Vec}$ are δ-interleaved in $\mathcal{O}(n^\omega)$.

D The Isomorphism Problem for $\mathbf{Z}^2\to\mathbf{Set}$

The isomorphism problem for Reeb graphs can be rephrased as an isomorphism problem of $\mathbf{Z}^2$-indexed persistence modules. Indeed, following [9], one sees that Reeb graphs can be viewed as functors $\mathbf{R}^2\to\mathbf{Set}$, and by [20] it follows that these functors are discrete in the sense of Appendix B. As the isomorphism problem for Reeb graphs is graph isomorphism hard [20], it follows immediately that the same is true for modules $\mathbf{Z}^2\to\mathbf{Set}$. We shall show that these problems are in fact graph isomorphism *complete*. Since we have chosen the total cardinality as the input size, and every functor $\mathbf{Z}^2\to\mathbf{Set}$, except the one sending everything to the empty set, has infinite total cardinality, we consider functors $[n]^2\to\mathbf{Set}$ instead. Here $[n]^2$ is $\{1,2,\dots,n\}^2$ considered as a full subcategory of $\mathbf{Z}^2$.

Let $M,N:[n]^2\to\mathbf{Set}$. We shall associate a pair of multigraphs to M and N in a way that ensures that M and N are isomorphic if and only if the associated multigraphs are isomorphic. The isomorphism problem for multigraphs is GI-complete [29].

An isomorphism between $M,N:[n]^2\to\mathbf{Set}$ is a natural isomorphism, i.e. a natural transformation with a two-sided inverse. Concretely, such an isomorphism f consists of bijections $f_p:M_p\to N_p$ for all $p\in[n]^2$ that commute with the internal morphisms of M and N, meaning that $f_{p+(0,1)}\circ\varphi_M(p,p+(0,1))=\varphi_N(p,p+(0,1))\circ f_p$ and $f_{p+(1,0)}\circ\varphi_M(p,p+(1,0))=\varphi_N(p,p+(1,0))\circ f_p$ hold whenever everything is defined. It is not hard to check that f^{-1} defined by $\left(f^{-1}\right)_p=(f_p)^{-1}$ is an inverse of f.

Given modules $M,N:[n]^2\to\mathbf{Set}$, we may assume that their pointwise cardinalities are the same, since if not, we can immediately conclude that they are not isomorphic. Let $c=|M|=|N|$. We also assume that M_p and N_p are nonempty on $p=(1,1)$, and for at least one $p\in\{1\}\times[n]\cup[n]\times\{1\}$. This implies $c\ge n$. We define the graph $G(M)=(V,E)$ as follows.

- $V=\bigcup_{p\in[n]^2}M_p\cup\{T\}$.
- There is a single edge between $x\in M_p$ and $y\in M_q$ if $\varphi_M(p,q)(x)=y$ and either $q=p+(0,1)$ or $q=p+(1,0)$.

- For $x \in M_{(a,b)}$, there are $n(a-1)+b$ edges between x and T.

Except from the ones described, there are no edges in $G(M)$. We can visualize $G(M)$ as the graph we get by putting $|M_p|$ vertices at each point in $[n]^2$ and short horizontal and vertical edges given by the internal morphisms of M, and in addition one vertex T which is incident to a certain number of edges from each other vertex. We have $|V| = c+1$ and $|E| \leq 2c^2 + cn^2 \leq (2+c)c^2$, since at most $2c^2$ edges come from the internal morphisms of M and cn^2 is an upper bound on the number of edges incident to T. In other words $|V| + |E|$ is polynomial in c. Defining $G(N) = (V', E')$ analogously with T' in place for T, we get the same for $|V'| + |E'|$.

Now we consider what an isomorphism $f : V \to V'$ from $G(M)$ to $G(N)$ must look like. Except for cases with $c \leq 2$, both graphs have exactly one vertex that is adjacent to all other vertices, so T must be sent to $T' \in V'$. Since there are $n(a-1)+b$ edges between $x \in M_{(a,b)}$ and T, there must be $n(a-1)+b$ edges between $f(x)$ and $f(T) = T'$, implying $f(x) \in N_{(a,b)}$. Thus the restriction of f to M_p is a bijection $M_p \to N_p$ for each $p \in [n]^2$.

It is easy to see that f is functorial. That is, there is an edge between $x \in M_p$ and $y \in M_q$ if and only if there is an edge between $f(x) \in N_p$ and $f(y) \in N_q$. Hence, we conclude that f defines an isomorphism between M and N in the obvious way.

Remark D.1. With small adjustments, the reduction from isomorphism of functors $[n]^2 \to \mathbf{Set}$ to isomorphism of multigraphs would work just as well for any poset category $\mathbf{P}$ in place of $[n]^2$. This shows that determining isomorphism between $\mathbf{Set}$-valued functors is at most as hard as GI regardless of the poset category.

E The Isomorphism Problem for $\mathbf{Z} \to \mathbf{Set}$

We consider functors $[n] \to \mathbf{Set}$, where $[n] = \{1, 2, \ldots, n\}$ is a subcategory of $\mathbf{Z}$, as in Appendix D. A rooted tree is a tree with one vertex chosen as the root, and an isomorphism between two rooted trees is a graph isomorphism that sends the root of one tree to the root of the other. We will show that deciding whether functors $[n] \to \mathbf{Set}$ are isomorphic is linear in the total cardinality by reducing it to checking isomorphism between rooted trees, which is known to be linear in the number of vertices [2, p. 85].

Given $M : [n] \to \mathbf{Set}$, let $T(M)$ be the rooted tree with vertex set $\bigcup_{k=1}^{n} M_k \sqcup \{r\}$, where we choose r as the root and there is an edge between $x \in M_k$ and $y \in M_{k+1}$ if $\varphi_M(k, k+1)(x) = y$. For persistence modules M and N, an isomorphism between $T(M)$ and $T(N)$ is a function that sends the root of $T(M)$ to the root of $T(N)$ and restricts to a bijection from M_k to N_k for each k. Moreover, f preserves parent-child relations, which means that for $x \in M_k$, $k < n$, $\varphi_N(k, k+1)(f(x)) = f(\varphi_M(k, k+1)(x))$. This is exactly what it takes for f restricted to $\bigcup_{k=1}^{n} M_k$ to define a natural transformation from M to N. Thus, $T(M)$ and $T(N)$ are isomorphic as rooted trees if and only if M and N are isomorphic as functors.

The number of vertices of $T(M)$ is one more than the total cardinality of M. Assuming that for each k we are given a list of tuples $(x, \varphi_M(k, k+1)(x))$, where x runs through the elements of M_k, we have exactly the information needed to run the algorithm in [2, p. 84] for checking isomorphism of $T(M)$ and $T(N)$ in linear time. Thus, deciding whether merge trees are isomorphic can be done in time linear in $|M| + |N|$.

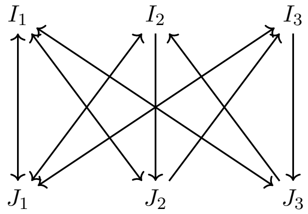


Figure 4: Graph illustrating possible nonzero morphisms between interval modules; see Appendix F.

F Example $d_I \neq d_B$

Consider the CI-problem $(P, Q, 3)$, where $P = \{(2,3), (3,2)\}$ and $Q = \{(2,2), (3,3)\}$. Applying the algorithm in Section 4, we get modules $M = I^{\mathcal{I}_1} \oplus I^{\mathcal{I}_2} \oplus I^{\mathcal{I}_3}$ and $N = I^{\mathcal{J}_1} \oplus I^{\mathcal{J}_2} \oplus I^{\mathcal{J}_3}$, where

$$
\begin{aligned}
I_1 &= \mathcal{W} \cup \langle x_1 \rangle \cup \langle x_2 \rangle \cup \langle x_3 \rangle \cup \langle x_4 \rangle, \\
I_2 &= \mathcal{W} \cup \langle x_1 - (1,1) \rangle \cup \langle x_2 \rangle \cup \langle x_4 \rangle, \\
I_3 &= \mathcal{W} \cup \langle x_1 \rangle \cup \langle x_2 - (1,1) \rangle \cup \langle x_3 \rangle, \\
J_1 &= \mathcal{W} \cup \langle x_1 \rangle \cup \langle x_2 \rangle \cup \langle x_3 \rangle \cup \langle x_4 \rangle, \\
J_2 &= \mathcal{W} \cup \langle x_1 \rangle \cup \langle x_3 - (1,1) \rangle \cup \langle x_4 \rangle, \\
J_3 &= \mathcal{W} \cup \langle x_2 \rangle \cup \langle x_3 \rangle \cup \langle x_4 - (1,1) \rangle.
\end{aligned}
$$

By Example 4.1, $(P, Q, 3)$ is solvable, which means that M and N are 1-interleaved. Since they are not isomorphic, $d_I(M, N) = 1$.

The graph in Figure 4 has an edge from A to B if A and B are in different barcodes and there is a nonzero morphism from $\mathcal{I}^A$ to $\mathcal{I}^B(1)$. (A double-headed arrow means an edge in each direction.) In a 1-matching between $\mathcal{B}(M)$ and $\mathcal{B}(N)$, if there is one, we need to match each I_i with a J_j, and each corresponding pair of interval modules needs to be 1-interleaved. Specifically, there needs to be a nonzero morphism both from $\mathcal{I}^{I_i}$ to $\mathcal{I}^{J_j}(1)$ and from $\mathcal{I}^{J_j}$ to $\mathcal{I}^{I_i}(1)$, that is, there must be edges in both directions between I_i and J_j in the graph. We see that both I_2 and I_3 can only be matched with J_1, and J_1 can only be matched with one of them. Thus there is no 1-matching between $\mathcal{B}(M)$ and $\mathcal{B}(N)$. On the other hand, all the intervals are 2-interleaved, so any bijection between $\mathcal{B}(M)$ and $\mathcal{B}(N)$ gives a 2-matching. In other words, $d_B(\mathcal{B}(M), \mathcal{B}(N)) = 2$.

References

[1] P. K. Agarwal, K. Fox, A. Nath, A. Sidiropoulos, and Y. Wang. Computing the gromov-hausdorff distance for metric trees. In *International Symposium on Algorithms and Computation*, pages 529–540. Springer, 2015.

[2] A. V. Aho and J. E. Hopcroft. *The design and analysis of computer algorithms*. Pearson Education India, 1974.

[3] G. Azumaya. Corrections and supplementaries to my paper concerning Krull–Remak–Schmidt's theorem. *Nagoya Mathematical Journal*, 1:117–124, 1950.

[4] M. Barot. *Introduction to the representation theory of algebras*. Springer, 2014.

[5] U. Bauer and M. Lesnick. Induced matchings and the algebraic stability of persistence barcodes. *Journal of Computational Geometry*, 6(2):162–191, 2015.

[6] S. Biasotti, A. Cerri, P. Frosini, D. Giorgi, and C. Landi. Multidimensional size functions for shape comparison. *Journal of Mathematical Imaging and Vision*, 32(2):161–179, 2008.

[7] H. B. Bjerkevik. Stability of higher-dimensional interval decomposable persistence modules. *arXiv preprint arXiv:1609.02086*, 2016.

[8] M. B. Botnan. Interval decomposition of infinite zigzag persistence modules. *Proceedings of the American Mathematical Society*, 145(8):3571—-3577, 2017.

[9] M. B. Botnan and M. Lesnick. Algebraic stability of zigzag persistence modules. *arXiv preprint arXiv:1604.00655*, 2016.

[10] P. A. Brooksbank and E. M. Luks. Testing isomorphism of modules. *Journal of Algebra*, 320 (11):4020–4029, 2008.

[11] P. Bubenik and J. A. Scott. Categorification of persistent homology. *Discrete & Computational Geometry*, 51(3):600–627, 2014.

[12] P. Bubenik, V. de Silva, and J. Scott. Metrics for generalized persistence modules. *Foundations of Computational Mathematics*, 15(6):1501–1531, 2015.

[13] P. Bubenik, V. de Silva, and V. Nanda. Higher interpolation and extension for persistence modules. 2017.

[14] G. Carlsson and V. de Silva. Zigzag persistence. *Foundations of computational mathematics*, 10 (4):367–405, 2010.

[15] G. Carlsson and A. Zomorodian. The theory of multidimensional persistence. *Discrete & Computational Geometry*, 42(1):71–93, 2009. ISSN 0179-5376.

[16] G. Carlsson, V. de Silva, and D. Morozov. Zigzag persistent homology and real-valued functions. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 247–256. ACM, 2009.

[17] F. Chazal, D. Cohen-Steiner, M. Glisse, L. J. Guibas, and S. Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 237–246. ACM, 2009.

[18] J. Cochoy and S. Oudot. Decomposition of exact pfd persistence bimodules. *arXiv preprint arXiv:1605.09726*, 2016.

[19] W. Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and Its Applications*, 14(05):1550066, 2015.

[20] V. de Silva, E. Munch, and A. Patel. Categorified Reeb graphs. *arXiv preprint arXiv:1501.04147*, 2015.

[21] T. K. Dey and C. Xin. Computing bottleneck distance for 2-d interval decomposable modules. *arXiv preprint arXiv:1803.02869*, 2018.

[22] O. H. Ibarra, S. Moran, and R. Hui. A generalization of the fast lup matrix decomposition algorithm and applications. *Journal of Algorithms*, 3(1):45–56, 1982.

[23] M. Kerber, D. Morozov, and A. Nigmetov. Geometry helps to compare persistence diagrams. *Journal of Experimental Algorithmics (JEA)*, 22(1):1–4, 2017.

[24] M. Lesnick. Multidimensional interleavings and applications to topological inference. *arXiv preprint arXiv:1206.1365*, 2012.

[25] M. Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Foundations of Computational Mathematics*, 15(3):613–650, 2015.

[26] N. Milosavljević, D. Morozov, and P. Skraba. Zigzag persistent homology in matrix multiplication time. In *Proceedings of the twenty-seventh annual symposium on Computational geometry*, pages 216–225. ACM, 2011.

[27] M. Scolamiero, W. Chachólski, A. Lundman, R. Ramanujam, and S. Öberg. Multidimensional persistence and noise. *Foundations of Computational Mathematics*, 17(6):1367–1406, 2017.

[28] C. Webb. Decomposition of graded modules. *Proceedings of the American Mathematical Society*, 94(4):565–571, 1985.

[29] V. N. Zemlyachenko, N. M. Korneenko, and R. I. Tyshkevich. Graph isomorphism problem. *Journal of Soviet Mathematics*, 29(4):1426–1481, 1985.

Paper III

Computing the Interleaving Distance Is NP-Hard

Håvard Bakke Bjerkevik   Magnus Bakke Botnan   Michael Kerber

Published in Foundations of Computational Mathematics

Foundations of Computational Mathematics
https://doi.org/10.1007/s10208-019-09442-y

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

Computing the Interleaving Distance is NP-Hard

Håvard Bakke Bjerkevik[1] · Magnus Bakke Botnan[2] · Michael Kerber[3]

Received: 22 November 2018 / Revised: 8 August 2019 / Accepted: 10 September 2019

Abstract
We show that computing the interleaving distance between two multi-graded persistence modules is NP-hard. More precisely, we show that deciding whether two modules are 1-interleaved is NP-complete, already for bigraded, interval decomposable modules. Our proof is based on previous work showing that a constrained matrix invertibility problem can be reduced to the interleaving distance computation of a special type of persistence modules. We show that this matrix invertibility problem is NP-complete. We also give a slight improvement in the above reduction, showing that also the approximation of the interleaving distance is NP-hard for any approximation factor smaller than 3. Additionally, we obtain corresponding hardness results for the case that the modules are indecomposable, and in the setting of one-sided stability. Furthermore, we show that checking for injections (resp. surjections) between persistence modules is NP-hard. In conjunction with earlier results from computational algebra this gives a complete characterization of the computational complexity of one-sided stability. Lastly, we show that it is in general NP-hard to approximate distances induced by noise systems within a factor of 2.

Keywords NP-hardness · Persistent homology · Interleavings · Matrix completion problems

Mathematics Subject Classification 15A83 · 55U99

Communicated by Herbert Edelsbrunner.

Magnus Bakke Botnan
m.b.botnan@vu.nl

Håvard Bakke Bjerkevik
havard.bjerkevik@ntnu.no

Michael Kerber
kerber@tugraz.at

[1] Institutt for matematiske fag, NTNU, 7491 Trondheim, Norway

[2] Department of Mathematics, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands

[3] Institut für Geometrie, TU Graz, 8010 Graz, Austria

Published online: 11 November 2019

1 Introduction

1.1 Motivation and Problem Statement

A *persistence module* M over $\mathbb{R}^d$ is a collection of vector spaces $\{M_p\}_{p\in\mathbb{R}^d}$ and linear maps $M_{p\to q}: M_p \to M_q$ whenever $p \leq q$, with the property that $M_{p\to p}$ is the identity map and the linear maps are composable in the obvious way. For $d = 1$, we will talk about *single-parameter persistence*, and for $d \geq 2$, we will use the term *multi-parameter persistence*.

Persistence, particularly in its single-parameter version, has recently gained a lot of attention in applied fields, because one of its instantiations is *persistent homology*, which studies the evolution of homology groups when varying a real scale parameter. The observation that topological features in real data sets carry important information to analyze and reason about the contained data has given rise to the term *topological data analysis* (*TDA*) for this research field, with various connections to application areas, e.g., [2,9,16–18].

A recurring task in TDA is the comparison of two persistence modules. The natural notion in terms of algebra is by interleavings of two persistence modules: given two persistence modules M and N as above and some $\epsilon > 0$, an ϵ*-interleaving* is the assignment of maps $\phi_p: M_p \to N_{p+\epsilon}$ and $\psi_p: N_p \to M_{p+\epsilon}$ which commute with each other and the internal maps of M and N. The *interleaving* distance is then just the infimum over all ϵ for which an interleaving exists.

A desirable property for any distance on persistence modules is *stability*, meaning informally that a small change in the input data set should only lead to a small distortion of the distance. At the same time, we aim for a *sensitive* measure, meaning that the distance between modules should be generally as large as possible without violating stability. As an extreme example, the distance measure that assigns 0 to all pairs of modules is maximally stable, but also maximally insensitive. Lesnick [14] proved that among all stable distances for single- or multi-parameter persistence, the interleaving distance is the most sensitive one over prime fields. This makes the interleaving distance an interesting measure to be used in applications and raises the question of how costly it is to compute the distance [14, Sec. 1.3 and 7]. Of course, for the sake of computation, a suitable finiteness condition must be imposed on the modules to ensure that they can be represented in finite form; we postpone the discussion to Sect. 3, and simply call such modules *of finite type*.

The complexity of computing the interleaving distance is well understood for the single-parameter case. The *isometry theorem* [8,14] states the equivalence of the interleaving distance and the *bottleneck distance*, which is defined in terms of the *persistence diagrams* of the persistence modules and can be reduced to the computation of a min cost bottleneck matching in a complete bipartite graph [11]. That matching, in turn, can be computed in $O(n^{1.5} \log n)$ time, and efficient implementations have been developed recently [13].

The described strategy, however, fails in the multi-parameter case, simply because the two distances do not match for more than one parameter: even if the multi-parameter persistence module admits a decomposition into *intervals* (which are "nice" indecomposable elements, see Sect. 3), it has been proved that the interleaving distance and the multi-parameter extension of the bottleneck distance are arbitrarily far from each other [5, Example 9.1]. Another example where the interleaving and bottleneck distances differ is given in [3, Example 4.2]; moreover, in this example the pair of persistence modules has the property that potential interleavings can be written on a particular matrix form, later formalized by the introduction of *CI problems* in [4]. A consequence is that the strategy of computing interleaving distance by computing the bottleneck distance fails also in this special case.

1.2 Our Contributions

We show that, for $d = 2$, the computation of the interleaving distance of two persistence modules of finite type is NP-hard, even if the modules are assumed to be decomposable into intervals. In [4], it is proved that the problem is CI-hard, where CI is a combinatorial problem related to the invertibility of a matrix with a prescribed set of zero elements. This is done by associating a pair of modules to each CI problem such that the modules are 1-interleaved if and only if the CI problem has a solution. We "finish" this proof by showing that CI is NP-complete, hence proving the main result. The hardness result on CI is independent of all topological concepts required for the rest of the paper and potentially of independent interest in other algorithmic areas.

Moreover, we slightly improve the reduction from [4] that asserts the CI-hardness of the interleaving distance, showing that also obtaining a $(3 - \epsilon)$-approximation of the interleaving distance is NP-hard to obtain for every $\epsilon > 0$. This result follows from the fact that our improved construction takes an instance of a CI problem and returns a pair of persistence modules which are 1-interleaved if the instance has a solution and are 3-interleaved if no solution exists. We mention that for *rectangle decomposable* modules in $d = 2$, a subclass of interval decomposable modules, it is known that the bottleneck distance 3-approximates the interleaving distance [3, Theorem 3.2], and can be computed in polynomial time. While this result does not directly extend to all interval decomposable modules, it gives reason to hope that a 3-approximation of the interleaving distance exists for a larger class of modules.

We also extend our hardness result to related problems: we show that it is NP-complete to compute the interleaving distance of two *indecomposable* persistence modules (for $d = 2$). We obtain this result by "stitching" together the interval decomposables from our main result into two indecomposable modules without affecting their interleaving distance. We remark that the restriction of computing the interleaving distance of indecomposable *interval* modules has recently been shown to be in P [10].

Bauer and Lesnick [1] showed that the existence of an interleaving pair, for modules indexed over $\mathbb{R}$, is equivalent to the existence of a single morphism with kernel and cokernel of a corresponding "size". While the equivalence does not hold in general, the two concepts are still closely related for $d > 1$. Using this, we obtain as a corollary to the aforementioned results that it is in general NP-complete to decide if there exists a morphism whose kernel and cokernel have size bounded by a given parameter. We also show that it is NP-complete to decide if there exists a surjection (dually, an injection) from one persistence module to another. Together with the result of [6], this gives a complete characterization of the computational complexity of "one-sided stability". Furthermore, we remark that this gives an alternative proof of the fact that checking for injections (resp. surjections) between modules over a finite-dimensional algebra (over a finite field) is NP-hard. This was first shown in [12, Theorem 1.2] (for arbitrary fields). The paper concludes with a result showing that it is in general NP-hard to approximate distances induced by noise systems (as introduced by Scolamiero et al. [19]) within a factor of 2.

1.3 Outline

We begin with the hardness proof for CI in Sect. 2. In Sect. 3, we discuss the representation-theoretic concepts needed in the paper. In Sect. 4, we describe our improved reduction scheme from interleaving distance to CI. In Sect. 5, we prove the hardness for indecomposable modules. In Sect. 6, we prove our hardness result for one-sided stability. A result closely related to one-sided stability can be found in Sect. 7 where we discuss a particular distance induced by a noise system. We conclude in Sect. 8.

2 The CI Problem

Throughout the paper, we set $\mathbb{F}$ to be any finite field with a constant number of elements. We write $\mathbb{F}^{n\times n}$ for the set of $n \times n$-matrices over $\mathbb{F}$, and $P_{ij} \in \mathbb{F}$ for the entry of P at the position at row i and column j. We write I_n for the $n \times n$-unit matrix. The *constrained invertibility problem* asks for a solution of the equation $AB = I_n$, when certain entries of A and of B are constrained to be zero. Formally, using the notation $[n] := \{1, \dots, n\}$, we define the language

$$\begin{aligned}\mathrm{CI} := \{(n, P, Q) \mid\ & P \subseteq [n]\times[n] \wedge Q \subseteq [n]\times[n] \wedge \exists A, B \in \mathbb{F}^{n\times n} :\\ & \big(\forall (i, j) \in P : A_{i,j} = 0 \wedge \forall (i, j) \in Q : B_{i,j} = 0 \wedge AB = I_n\big)\}.\end{aligned}$$

We can write CI-instances in a more visual form, for instance writing

$$\begin{pmatrix} * & * & * \\ * & 0 & * \\ * & * & 0 \end{pmatrix}\begin{pmatrix} * & * & * \\ * & * & 0 \\ * & 0 & * \end{pmatrix} = I_3$$

instead of $(3, \{(2, 2), (3, 3)\}, \{(2, 3), (3, 2)\})$. Indeed, the CI problem asks whether in the above matrices, we can fill the $*$-entries with field elements to satisfy the equation.

In the above example, this is indeed possible, for instance by choosing

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \mathbf{0} & 1 \\ 1 & 1 & \mathbf{0} \end{pmatrix} \quad B = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & \mathbf{0} \\ 1 & \mathbf{0} & -1 \end{pmatrix}.$$

We sometimes also call A and B a *satisfying assignment*. In contrast, the instance

$$\begin{pmatrix} 0 & * & 0 \\ * & * & * \\ * & * & * \end{pmatrix} \begin{pmatrix} * & * & * \\ 0 & * & * \\ * & * & * \end{pmatrix} = I_3$$

has no solution, because the $(1, 1)$ entry of the product on the left is always 0, no matter what values are chosen. Note that the existence of a solution also depends on the characteristic of the base field. For an example, see Chapter 4, page 13 in [4].

The CI problem is of interest to us, because we will see in Sect. 4 that CI *reduces* to the problem of computing the interleaving distance, that is, a polynomial time algorithm for computing the interleaving distance will allow us to decide whether a triple (n, P, Q) is in CI, also in polynomial time. Although the definition of CI is rather elementary and appears to be useful in different contexts, we are not aware of any previous work studying this problem (apart from [4]).

It is clear that CI is in NP because a valid choice of the matrices A and B can be checked in polynomial time. We want to show that CI is NP-hard as well. It will be convenient to do so in two steps. First, we define a slightly more general problem, called *generalized constrained invertibility (GCI)*, and show that GCI reduces to CI. Then, we proceed by showing that 3SAT reduces to GCI, proving the NP-hardness of CI.

2.1 Generalized Constrained Invertibility

We generalize from the above problem in two ways: first, instead of square matrices, we allow that $A \in \mathbb{F}^{n\times m}$ and $B \in \mathbb{F}^{m\times n}$ (where m is an additional input). Second, instead of forcing $AB = I_n$, we only require that AB coincides with I_n in a fixed subset of entries over $[n] \times [n]$. Formally, we define

$$\begin{aligned} GCI := \{(n, m, P, Q, R) \mid\ & P \subseteq [n] \times [m] \wedge Q \subseteq [m] \times [n] \wedge R \subseteq [n] \\ & \times [n] \wedge \exists A \in \mathbb{F}^{n\times m}, B \in \mathbb{F}^{m\times n} : \\ & \big(\forall (i, j) \in P : A_{i,j} = 0 \wedge \forall (i, j) \in Q : B_{i,j} \\ = 0 \wedge \forall (i, j) \in R : (AB)_{i,j} = (I_n)_{i,j}\big)\}. & \end{aligned}$$

Again, we use the following notation

$$\begin{pmatrix} * & * & * \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} * & 0 \\ 0 & * \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ * & * \end{pmatrix}$$

for the GCI-instance $(2, 3, \{(2, 1), (2, 2), (2, 3)\}, \{(1, 2), (2, 1), (3, 1), (3, 2)\}, \{(1, 1), (1, 2)\})$. This instance is indeed in GCI, as for instance,

$$\begin{pmatrix} 1 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & 1 \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ 0 & 0 \end{pmatrix}.$$

GCI is indeed generalizing CI, as we can encode any CI-instance by setting $m = n$ and $R = [n] \times [n]$. Hence, CI trivially reduces to GCI. We show, however, that also the converse is true, meaning that the problems are computationally equivalent. We will need the following lemma which follows from linear algebra:

Lemma 1 *Let $M \in \mathbb{F}^{n\times m}$, $N \in \mathbb{F}^{m\times n}$ with $m > n$ such that $MN = I_n$. Then there exist matrices $M' \in \mathbb{F}^{(m-n)\times m}$, $N' \in \mathbb{F}^{m\times(m-n)}$ such that*

$$\begin{bmatrix} M \\ M' \end{bmatrix} \begin{bmatrix} N & N' \end{bmatrix} = I_m.$$

Proof Pick $M'' \in \mathbb{F}^{(m-n)\times m}$ so that $\begin{bmatrix} M \\ M'' \end{bmatrix}$ has full rank. This is possible, as the row vectors of M are linearly independent, so we can pick the rows in M'' iteratively such that they are linearly independent of each other and those in M. Let $M' = M'' - M''NM$, which gives

$$M'N = (M'' - M''NM)N = M''N - M''NI_n = 0_{(m-n)\times n}.$$

Since

$$\begin{bmatrix} M \\ M'' \end{bmatrix} = \begin{bmatrix} I_n & 0_{n\times(m-n)} \\ M''N & I_{m-n} \end{bmatrix} \begin{bmatrix} M \\ M' \end{bmatrix},$$

$\begin{bmatrix} M \\ M' \end{bmatrix}$ also has full rank, which means that it has an inverse. Let N' be the last $m - n$ columns of this inverse matrix. We get

$$\begin{bmatrix} M \\ M' \end{bmatrix} \begin{bmatrix} N & N' \end{bmatrix} = \begin{bmatrix} I_n & 0_{n\times(m-n)} \\ 0_{(m-n)\times n} & I_{m-n} \end{bmatrix} = I_m.$$

□

Lemma 2 *GCI is polynomial time reducible to CI.*

Proof Fix a GCI-instance (n, m, P, Q, R). We have to define a polynomial time algorithm to compute a CI-instance (n', P', Q') such that

$$(n, m, P, Q, R) \in GCI \Leftrightarrow (n', P', Q') \in CI.$$

Write the GCI-instance as $AB = C$, where A and B are matrices with 0 and $*$ entries (of dimensions $n \times m$ and $m \times n$, respectively), and C is an $n \times n$-matrix with 1 or $*$ entries on the diagonal, and 0 or $*$ entries away from the diagonal (as in the example above).

Define the matrix I_n^* as the matrix with 0 away from the diagonal and $*$ on the diagonal. Moreover, let $\bar{C}$ denote the matrix C with all 1-entries replaced by 0-entries. Now, consider the GCI-instance

$$\begin{bmatrix} A & I_n^* \end{bmatrix} \begin{bmatrix} B \\ \bar{C} \end{bmatrix} = I_n \tag{1}$$

which can be formally written as $(n, n+m, P', Q', [n]\times[n])$ for some choices $P' \supseteq P$, $Q' \supseteq Q$.

We claim that the original instance is in GCI if and only if the extended instance is in GCI. First, assume that $AB = C$ has a solution (that is, an assignment of field elements to $*$ entries that satisfies the equation). Then, we pick all diagonal entries in I_n^* as 1, so that the matrix becomes I_n. Also, we pick $\bar{C}$ to be $I_n - AB$; this is indeed possible, as an entry in $\bar{C}$ is fixed only if the corresponding positions of I_n and AB coincide. With these choices, we have that

$$\begin{bmatrix} A & I_n^* \end{bmatrix} \begin{bmatrix} B \\ \bar{C} \end{bmatrix} = AB + I_n(I_n - AB) = I_n,$$

as required.

Conversely, if there is a solution for the extended instance, write X for the assignment of I_n^* and Y for the assignment of $\bar{C}$. Then $AB + XY = I_n$. Now fix any index $(i, j) \in R$ and consider the equation in that entry. By construction $Y_{i,j} = 0$, and multiplication by the diagonal matrix X does not change this property. It follows that $(AB)_{(i,j)} = (I_n)_{i,j}$, which means that $AB = C$ has a solution. Hence, the two instances are indeed equivalent.

To finish the proof, we observe that (1) is in GCI if and only if

$$\begin{bmatrix} A & I_n^* \\ *_{m\times m} & *_{m\times n} \end{bmatrix} \begin{bmatrix} B & *_{m\times m} \\ \bar{C} & *_{n\times m} \end{bmatrix} = I_{n+m} \tag{2}$$

is in GCI, where $*_{a\times b}$ is simply the $a \times b$ matrix only containing $*$ entries. Formally written, this instance corresponds to $(n+m, n+m, P', Q', [n+m]\times[n+m])$. To see the equivalence, if (1) is in GCI, Lemma 1 asserts that there are indeed choices for the $*$-matrices to solve (2) as well. In the opposite direction, a satisfying assignment of the involved matrices in (2) also yields a valid solution for (1) when restricted to the upper n rows and left n columns, respectively.

Combining everything, we see that (n, m, P, Q, R) is in GCI if and only if $(n + m, n+m, P', Q', [n+m]\times[n+m])$ is in GCI. The latter, however, is equivalent to the CI-instance $(n+m, P', Q')$. The conversion can clearly be performed in polynomial time, and the statement follows. □

2.2 Hardness of GCI

We describe now how an algorithm for how deciding GCI can be used to decide satisfiability of 3SAT formulas. Let ϕ be a 3CNF formula with n variables and m clauses. We construct a GCI-instance that is satisfiable if and only if ϕ is satisfiable.

In what follows, we will often label some $*$ entries in matrices with variables when we want to talk about the possible assignments of the corresponding entries.

The first step is to build a "gadget" that allows us to encode the truth value of a variable in the matrix. Consider the instance

$$\begin{pmatrix} * & 0 & * \\ 0 & * & * \end{pmatrix} \begin{pmatrix} x & 0 \\ 0 & y \\ * & * \end{pmatrix} = I_2.$$

In any solution to this equation, not both x and y can be zero because otherwise, the right matrix would have rank at most 1. Furthermore, when extending the instance by one row/column

$$\left(\begin{array}{ccc} a & b & 0 \\ \hline * & 0 & * \\ 0 & * & * \end{array}\right) \left(\begin{array}{c|cc} 0 & x & 0 \\ 0 & 0 & y \\ 0 & * & * \end{array}\right) = \begin{pmatrix} * & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

we see that both $ax = 0$ and $by = 0$ must hold, which is then only possible if at least one entry a or b is equal to 0. In fact, there is a solution with $a \neq 0$, and a solution with $b \neq 0$, for instance

$$\begin{pmatrix} 1 & 0 & \mathbf{0} \\ 0 & \mathbf{0} & 1 \\ \mathbf{0} & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{0} & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \\ \mathbf{0} & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix},$$
$$\begin{pmatrix} 0 & 1 & \mathbf{0} \\ 1 & \mathbf{0} & 0 \\ \mathbf{0} & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{0} & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 \\ \mathbf{0} & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix}.$$

The intuition is that for a variable x_i appearing in ϕ, we interpret x_i to be true if $a \neq 0$, and to be false if $b \neq 0$. We build such a gadget for each variable. A crucial observation is that we can do so with all variable entries placed in the same row. This works essentially by concatenating the variable gadgets, in a block-like fashion. We show the construction for three variables as an example.

$$\begin{pmatrix} a_0 & b_0 & 0 & a_1 & b_1 & 0 & a_2 & b_2 & 0 & * \\ * & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & * & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & * & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & 0 \end{pmatrix} \begin{pmatrix} 0 & * & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 & 0 & 0 \\ 0 & * & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & 0 & 0 \\ 0 & 0 & 0 & * & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & * & * \\ * & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} = I_7$$

where we introduced an additional column at the end of the left matrix and an additional row at the end of the second matrix. Firstly, this allows us to satisfy the entire I_7 on the right-hand side; moreover, it will be useful when extending the construction to clauses. It is straightforward to generalize this construction to an arbitrary number of variables. We arrive at the following intermediate result.

Lemma 3 *For any $n \geq 1$, there exists a GCI-instance $A'B' = I_{2n+1}$ with A' having $3n+1$ columns, such that in each solution for the problem, $A'_{1,3n+1}$ is not zero, and for each $k = 0, \ldots, n-1$, the entries $A'_{1,3k+1}$ and $A'_{1,3k+2}$ are not both non-zero. Moreover, for any choice of $v_1, \ldots, v_n \in \{1, 2\}$, there exists a solution of the instance in which $A'_{1,3k+v_i} \neq 0$ for all $k = 0, \ldots, n-1$.*

Next, we extend the instance from Lemma 3 with respect to the clauses. We refer to the clauses as $c_1, \ldots, c_m$. For each clause, we append one further row to A', each of them identical of the form

$$\begin{pmatrix} 0 & \ldots & 0 & * \end{pmatrix}^T .$$

We also append one column to B' for each clause, each of length $3n+1$. For each clause, the entry at row $3n+1$ is set to $*$. If a clause contains a literal of the form x_i (in positive form), we set the entry at row $3i+1$ to $*$. If it contains a literal $\neg x_i$, we set the entry at row $3i+2$ to $*$. In this way, at most 4 entries in the column are fixed to $*$, and we fix all other entries to be 0. Continuing the above example, for the clause $x_0 \vee \neg x_1 \vee x_2$, we obtain a column of the form

$$\begin{pmatrix} * & 0 & 0 & 0 & * & 0 & * & 0 & 0 & * \end{pmatrix}$$

Let A and B denote the matrices extended from A' and B' with the above procedure. We next define C as a square matrix of dimension $2n+1+m$ as follows: The upper left $(2n+1) \times (2n+1)$ submatrix is set to I_{2n+1}. The rest of the first row is set to 0, and the rest of the diagonal is set to 1. All other entries are set to $*$. This concludes the description of a GCI-instance $AB = C$ out of a 3CNF formula ϕ. We exemplify the construction for the formula $(x_0 \vee x_1 \vee \neg x_2) \wedge (\neg x_0 \vee x_1 \vee x_2)$, where the lines mark the boundary of A' and B', respectively.

$$\left(\begin{array}{cccccccccc} a_0 & b_0 & 0 & a_1 & b_1 & 0 & a_2 & b_2 & 0 & * \\ * & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & * & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & * & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * \end{array}\right) \left(\begin{array}{ccccccc|cc} 0 & * & 0 & 0 & 0 & 0 & 0 & * & 0 \\ 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & * \\ 0 & * & * & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & * & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * & 0 \\ 0 & 0 & 0 & 0 & 0 & * & * & 0 & 0 \\ * & 0 & 0 & 0 & 0 & 0 & 0 & * & * \end{array}\right)$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & * \\ * & * & * & * & * & * & * & 1 & * \\ * & * & * & * & * & * & * & * & 1 \end{pmatrix}$$

Lemma 4 *$AB = C$ admits a solution if and only if ϕ is satisfiable.*

Proof "$\Rightarrow$": Let us assume that $AB = C$ has a solution, which also implies a solution $A'B' = I_{2n+1}$ being a subproblem encoded in the instance. Fixing a solution, we assign an assignment of the variables of ϕ as follows: If the entry $A_{1,3i+1}$ is non-zero, we set x_i to true. If the entry $A_{1,3i+2}$ is non-zero, we set x_i to false. If neither is non-zero, we set x_i to false as well (the choice is irrelevant). Note that by Lemma 3, not both $A_{1,3i+1}$ and $A_{1,3i+2}$ can be non-zero, so the assignment is well-defined.

First of all, let γ be the rightmost entry of the first row of A. Because the $(1, 1)$-entry of C is set to 1, it follows that $\gamma\delta = 1$, where δ is the lowest entry of the first column of B. Hence, in the assumed solution, $\gamma \neq 0$.

Now fix a clause c in ϕ and let v denote the column of B assigned to this clause, with column index i. Recall that v consists of (up to) three $*$ entries chosen according to the literals of c, and a $*$ entry at the lowest position. Let λ denote the value of that lowest entry in the assumed solution of $AB = C$. We see that $\lambda \neq 0$, with a similar argument as for γ above, using the (i, i)-entry of C.

Now, the $(1, i)$ entry of C is set to 0 by construction which yields a constraint of the form

$$\mu_1 v_1 + \mu_2 v_2 + \mu_3 v_3 + \underbrace{\gamma\lambda}_{\neq 0} = 0$$

where v_1, v_2, v_3 are entries of v at the $*$ positions, and μ_1, μ_2, μ_3 the corresponding entries of the first row of A. We observe that at least one term $\mu_j v_j$ must be non-zero, hence both entries are non-zero.

This implies that the chosen assignment satisfies the clause: if v_j is at index $3k + 1$ for some k, the clause contains the literal x_k by construction and since $\mu_j \neq 0$, our assignment sets x_k to true. The same argument applies to v_j of the form $3k + 2$. It follows that the assignment satisfies all clauses and hence, ϕ is satisfiable.

"$\Leftarrow$": We pick a satisfying assignment for ϕ and fill the first row of A as follows: if x_i is true, we set $(A_{1,3i+1}, A_{1,3i+2})$ to $(1, 0)$ if x_i is false, we set it as $(0, 1)$. By Lemma 3, there exists a solution for $A'B' = I_{2n+1}$ with this initial values and we choose such a solution, filling the upper $(2n + 1)$ rows of A and the left $(2n + 1)$ columns of B. Note that similar as above, the value γ at $A_{1,3n+1}$ must be non-zero in such a solution. In the remaining m rows of A, by construction, we only need to pick the rightmost entry, and we set it to γ in each of these rows. That determines all entries of A.

To complete B' to B, we need to fix values in the columns of B associated to clauses. In each such column, we pick the lowest entry to be $\frac{1}{\gamma}$, satisfying the constraints of C along the diagonal. Fixing a column i of B, the $(1, i)$-constraint of C reads as

$$\mu_1 v_1 + \mu_2 v_2 + \mu_3 v_3 + \underbrace{\gamma \frac{1}{\gamma}}_{=1} = 0,$$

where v_1, v_2, v_3 are the remaining non-zero entries in i-th column. Because we encoded a satisfying assignment of ϕ in the first row of A, at least one μ_j entry is 1. We set the corresponding entry v_j to -1, and the remaining v_k's to 0. In this way, all constraints are satisfied, and the GCI-instance has a solution. □

Clearly, the GCI-instance of the preceding proof can be computed from ϕ in polynomial time. It follows:

Theorem 1 *CI is NP-complete.*

Proof Lemma 4 shows the reduction of 3SAT to GCI, proving that GCI is NP-complete. As shown in Lemma 2, GCI reduces to CI, proving the claim. □

3 Modules and Interleavings

In what follows, all vector spaces are understood to be $\mathbb{F}$-vector spaces for the fixed base field $\mathbb{F}$. Also, for points $p = (p_x, p_y), q = (q_x, q_y)$ in $\mathbb{R}^2$, we write $p \leq q$ if $p_x \leq q_x$ and $p_y \leq q_y$.

3.1 Persistence Modules

A *(two-parameter) persistence module* M is a collection of $\mathbb{F}$-vector spaces V_p, indexed over $p \in \mathbb{R}^2$ together with linear maps $M_{p\to q}$ whenever $p \leq q$. These maps must have the property that $M_{p\to p}$ is the identity map on M_p and $M_{q\to r} \circ M_{p\to q} = M_{p\to r}$ for $p \leq q \leq r$. Much more succinctly, a persistence module is a functor from the poset category $\mathbb{R}^2$ to the category of vector spaces. A *morphism* between M and N is a collection of linear maps $\{f_p : M_p \to N_p\}$ such that $N_{p\to q} \circ f_p = f_q \circ M_{p\to q}$. We say that f is an *isomorphism* if f_p is an isomorphism for all p, and denote this by $M \cong N$. If we view persistence modules as functors, a morphism is simply a natural transformation between the functors.

The simplest example is the 0-*module* where M_p is the trivial vector space for all $p \in \mathbb{R}^2$. For a more interesting example, define an *interval* in the poset $(\mathbb{R}^2, \leq)$ to be a non-empty subset $S \subset \mathbb{R}^2$ such that whenever $a, c \in S$ and $a \leq b \leq c$, then $b \in S$, and moreover, if $a, c \in S$, there exists a sequence of elements $a = b_1, \ldots, b_\ell = c$ of elements in S such that $b_i \leq b_{i+1}$ or $b_{i+1} \leq b_i$. We associate an *interval module* I^S

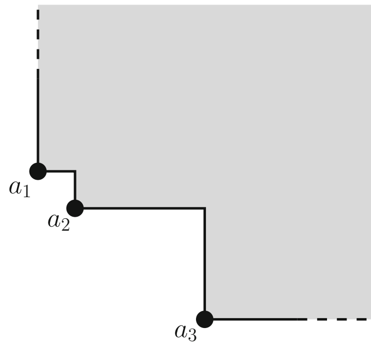


Fig. 1 A staircase of size 3 (shaded area)

to S as follows: for $p \in S$, we set $I_p^S := \mathbb{F}$, and $I_p^S := 0$ otherwise. As map $I_{p\to q}^S$ with $p \leq q$, we attach the identity map if $p, q \in S$, and the 0-map otherwise.

For $a \in \mathbb{R}^2$, let $\langle a \rangle := \{x \in \mathbb{R}^2 \mid a \leq x\}$ be the infinite rectangle with a as lower-left corner. Given k elements $a_1, \ldots, a_k \in \mathbb{R}^2$, the set

$$S := \bigcup_{i=1,\ldots,k} \langle a_i \rangle$$

is called the *staircase* with elements $a_1, \ldots, a_k$. We call k the size of the staircase. See Fig. 1 for an illustration. It is easy to verify that S is an interval for $k \geq 1$. Clearly, if $a_i \leq a_j$, we can remove a_j without changing the staircase, so we assume that the elements forming the staircase are pairwise incomparable. The *staircase module* is the interval module associated to the staircase.

Given two persistence modules M and N, the *direct sum* $M \oplus N$ is the persistence module where $(M \oplus N)_p := M_p \oplus N_p$, and the linear maps are defined componentwise in the obvious way. We call a persistence module M *indecomposable*, if in any decomposition $M = M_1 \oplus M_2$, M_1 or M_2 is the 0-module. For example, it is not difficult to see that interval modules are indecomposable. We call M *interval decomposable* if M admits a decomposition $M \cong M_1 \oplus \ldots \oplus M_\ell$ into (finitely many) interval modules. The decomposition of any persistence module into interval modules is unique up to rearrangement and isomorphism of the summands; see [5, Section 2.1] and the references therein. This implies that there is a well-defined multiset of intervals $B(M)$ given by the decomposition of M into interval modules. The multiset $B(M)$ is called the *barcode* of M. Not every module is interval decomposable; we remark that already rather simple geometric constructions can give rise to complicated indecomposable elements [7].

3.2 Interleavings

Let $\epsilon \in \mathbb{R}$. For a persistence module M, the ϵ-*shift* of M is the module M^ϵ defined by $M_p^\epsilon = M_{p+\epsilon}$ (where $p + \epsilon = (p_x + \epsilon, p_y + \epsilon)$) and $M_{p\to q}^\epsilon = M_{p+\epsilon\to q+\epsilon}$. Note that $(M^\epsilon)^\delta = M^{\epsilon+\delta}$. As an example, staircase modules are closed under shift: the

ϵ-shift of the staircase module associated to $\bigcup\langle a_i\rangle$ is the staircase module associated to $\bigcup\langle a_i - \epsilon\rangle$. We can also define shift on morphisms: for $f: M \to N$, $f^\epsilon: M^\epsilon \to N^\epsilon$ is given by $f_p^\epsilon = f_{p+\epsilon}$. For $\epsilon \geq 0$, there is an obvious morphism $\mathrm{Sh}_M(\epsilon): M \to M^\epsilon$ given by the internal morphisms of M, that is, we have $\mathrm{Sh}_M(\epsilon)_p = M_{p\to p+\epsilon}$. In practice we will often suppress notation and simply write $M \to M^\epsilon$ for this morphism.

With this in mind, we define an *ϵ-interleaving* between M and N for $\epsilon \geq 0$ as a pair (f, g) of morphisms $f: M \to N^\epsilon$ and $g: N \to M^\epsilon$ such that $g^\epsilon \circ f = \mathrm{Sh}_M(2\epsilon)$ and $f^\epsilon \circ g = \mathrm{Sh}_N(2\epsilon)$. Concretely, an ϵ-interleaving between two persistence modules M and N is a collection of maps

$$f_p: M_p \to N_{p+\epsilon} \tag{3}$$
$$g_p: N_p \to M_{p+\epsilon} \tag{4}$$

such that all diagrams that can be composed out of the maps f_*, g_*, and the linear maps of M and N commute. Note that a 0-interleaving simply means that the persistence modules are *isomorphic*. Also, an ϵ-interleaving induces a δ-interleaving for $\epsilon < \delta$ directly by a suitable composition with the linear maps of the modules.

We say that two modules are *ϵ-interleaved* if there exists an ϵ-interleaving between them. We define the *interleaving distance* of two modules M and N as

$$d_I(M, N) := \inf\{\epsilon \geq 0 \mid M \text{ and } N \text{ are } \epsilon\text{-interleaved}\}.$$

Note that d_I defines an extended pseudometric on the space of persistence modules. The distance between two modules might be infinite, and there are non-isomorphic modules with distance 0. The triangle inequality follows from the simple observation that an ϵ-interleaving between M_1 and M_2 and a δ-interleaving between M_2 and M_3 can be composed to an $(\epsilon + \delta)$-interleaving between M_1 and M_3.

3.3 Representation of Persistence Modules

For studying the computational complexity of the interleaving distance, we need to specify a finite representation of persistence modules that allows us to pass such modules as an input to an algorithm.

A *graded matrix representation* of a module M is a 3-tuple (G, R, A), where $G = \{g_1, \ldots, g_n\}$ is a list of n points in $\mathbb{R}^2$, $R = \{r_1, \ldots, r_m\}$ is a list of m points in $\mathbb{R}^2$, (with repetitions allowed), and A is an $(m \times n)$-matrix over the base field $\mathbb{F}$. Equivalently, we can simply think of a matrix A where each row and column is annotated with a grading in $\mathbb{R}^2$.

The algebraic explanation for this representation is as follows: it is known that a persistence module M over $\mathbb{R}^2$ can be equivalently described as a graded $\mathcal{R}$-module over a suitably chosen ring $\mathcal{R}$. Assuming that M is finitely presented, we can consider the free resolution of M

$$\mathcal{R}^m \xrightarrow{\partial^T} \mathcal{R}^n \to M \to 0.$$

A graded matrix representation is simply a way to encode the map ∂ in this resolution.

Let us describe for concreteness how a representation (G, R, A) gives rise to a persistence module. First, let $\mathbb{F}_1, \ldots, \mathbb{F}_n$ be copies of $\mathbb{F}$, and let e_i be the 1-element of $\mathbb{F}_i$. For $p \in \mathbb{R}^2$, we define Gen_p as the direct sum of all $\mathbb{F}_i$ such that $g_i \leq p$. Moreover, every row of A gives rise to a linear combination of the entries $e_1, \ldots, e_n$. Let c_i denote the linear combination in row i. We define Rel_p to be the span of all linear combinations c_i for which $r_i \leq p$. Then, we set

$$M_p := \frac{\mathrm{Gen}_p}{\mathrm{Rel}_p}$$

which is a $\mathbb{F}$-vector space. For $p \leq q$, writing $[x]_p$ for an element of M_p with $x \in \mathrm{Gen}_p$, we define

$$M_{p\to q}([x]_p) := [x]_q.$$

It is easy to check that that $[x]_q$ is well-defined (since $\mathrm{Gen}_p \subseteq \mathrm{Gen}_q$) and independent of the chosen representative in Gen_p (since $\mathrm{Rel}_p \subseteq \mathrm{Rel}_q$). Moreover, it is straightforward to verify that these maps satisfy the properties of a persistence module.

In short, every persistence module that can be expressed by finitely many generators and relations can be brought into graded matrix representation. For instance, a staircase module for $a_1, \ldots, a_n$ of size n where the a_i are ordered by increasing first coordinate can be represented by a matrix with n columns graded by $a_1, \ldots, a_n$, and $n - 1$ rows, where every row corresponds to a pair $(i, i + 1)$ with $1 \leq i \leq n - 1$. In this row, we encode the relation $e_i = e_{i+1}$ and grade it by p_{ij}, which is the (unique) minimal element q in $\mathbb{R}^2$ such that $a_i \leq q$ and $a_{i+1} \leq q$. Hence, the graded matrix representation of a staircase of size n has a size that is polynomial in n.

We also remark that a graded matrix representation is equivalent to *free implicit representations* [15, Sec 5.1] for the special case of $m_0 = 0$.

4 Hardness of Interleaving Distance

We consider the following computational problems:

1-INTERLEAVING: Given two persistence modules M, N in graded matrix representation, decide whether they are 1-interleaved.

c- APPROX- INTERLEAVING- DISTANCE: Given two persistence modules M, N in graded matrix representation, return a real number r such that

$$d_I(M, N) \leq r \leq c \cdot d_I(M, N)$$

Obviously, the problem of computing $d_I(M, N)$ exactly is equivalent to the above definition with $c = 1$.

The main result of this section is the following theorem:

Theorem 2 *Given a CI-instance* (n, P, Q)*, we can compute in polynomial time in n a pair of persistence modules* (M, N) *in graded matrix representation such that*

$$d_I(M, N) = \begin{cases} 1 & \text{if } (n, P, Q) \in CI \\ 3 & \text{if } (n, P, Q) \notin CI \end{cases}.$$

Moreover, both M and N are direct sums of staircase modules and hence interval decomposable.

We will postpone the proof of Theorem 2 to the end of the section and first discuss its consequences.

Theorem 3 1-Interleaving *is NP-complete.*

Proof We first argue that 1-Interleaving is in NP. First, note that to specify a 1-interleaving, it suffices to specify the maps at the points in S, where S is a finite set whose size is polynomial in the size of the graded matrix representation. More precisely, S contains the critical grades of the two modules (that is, the grades specified by G and R), as well as the least common successors of such elements. That ensures that every vector space (in both modules) can be isomorphically pulled back to one of the elements of S, and the interleaving map can be defined using this pull-back. It is enough to consider the points in S to check whether this set of pointwise maps is a valid morphism.

We can furthermore argue that verifying that a pair of such maps yields a 1-interleaving can be checked in a polynomial number of steps. Again, this involves mostly the maps specified above, as well as the corresponding maps shifted by $(1, 1)$, in order to check the compatibility of the two interleaving maps. We omit further details of this step.

Finally, 1-Interleaving is NP-hard: Assuming a polynomial time algorithm A to decide the problem, we can design a polynomial time algorithm for CI just by transforming (n, P, Q) into a pair of modules (M, N) using the algorithm from Theorem 2. If A applied on (M, N) returns true, we return that (n, P, Q) is in CI. Otherwise, we return that (n, P, Q) is not in CI. Correctness follows from Theorem 2, and the algorithm runs in polynomial time, establishing a polynomial time reduction. By Theorem 1, CI is NP-hard, hence, so is 1-Interleaving. □

Theorem 4 c-Approx- Interleaving- Distance *is NP-hard for every* $c < 3$ *(i.e., a polynomial time algorithm for the problem implies P=NP).*

Proof Fixing $c < 3$, assuming a polynomial time algorithm A for c-Approx- Interleaving- Distance yields a polynomial time algorithm for CI: Given the input (n, P, Q), we transform it into (M, N) with Theorem 2. Then, we apply A on (M, N). If the result is less than 3, we return that (n, P, Q) is in CI. Otherwise, we return that (n, P, Q) is not in CI. Correctness follows from Theorem 2, noting that if (n, P, Q) is in CI, algorithm A must return a number in the interval $[1, c]$ and $c < 3$ is assumed. If (n, P, Q) is not in CI, it returns a number ≥ 3. Also, the algorithm runs in polynomial time in n. Therefore, the existence of A yields a polynomial time algorithm for CI, implying P=NP with Theorem 1. □

Since the modules in Theorem 2 are direct sums of staircases, both Theorem 3 and Theorem 4 hold already for the restricted case that the modules are interval decomposable.

4.1 Interleavings of Staircases

The persistence modules constructed for the proof of Theorem 2 will be direct sums of staircases. Before defining them, we establish some properties of the interleaving map between staircases and their direct sums which reveal the connection to the CI problems.

Recall from Sect. 3 that a morphism $M \to N$ can be described more concretely as a collection of maps $M_p \to N_p$ that are compatible with the linear maps in M and N, that M^ϵ is defined by $M_p^\epsilon = M_{p+\epsilon}$, and that an ϵ-interleaving is a pair of morphisms $\phi\colon M \to N^\epsilon$, $\psi\colon N \to M^\epsilon$ satisfying certain conditions. For staircase modules, the set of morphisms is quite limited.

For M and N staircase modules and $\lambda \in \mathbb{F}$, we denote by $1 \mapsto \lambda$ the collection of linear maps ϕ_p such that $\phi_p(1) = \lambda$ for all p such that $M_p = \mathbb{F}$.

Lemma 5 *Let M and N be staircase modules. Every morphism from M to N is of the form $1 \mapsto \lambda$ for some $\lambda \in \mathbb{F}$.*

Proof Assume first that $p \leq q$ and $M_p = \mathbb{F}$. Write $\lambda := \phi_p(1)$. Then, also $M_q = \mathbb{F}$, and $\phi_q(1) = \lambda$ as well, since the linear maps from p to q for M and N are injective maps.

For incomparable p and $q \in \mathbb{R}^2$, we consider the least common successor r of p and q. Using the above property twice, we see at once that $\phi_p(1) = \phi_r(1) = \phi_q(1)$. □

We examine next which values of λ are possible for a concrete pair of staircases. For a staircase S, let S^ϵ denote the staircase where each point is shifted by (ϵ, ϵ). This way, if M is the module associated to S, M^ϵ is the module associated to S^ϵ. As we noted before, the shift of a staircase module is also a staircase module. Define the *directed shift distance* from the staircase S to the staircase T as

$$d_s(S, T) := \min\{\epsilon \geq 0 \mid S \subseteq T^\epsilon\}.$$

One can show that the set on the right-hand side has a minimum value by using the fact that a staircase is generated by a finite set of elements, so d_s is in fact well-defined. Clearly, $d_s(S, T) \neq d_s(T, S)$ in general. The following simple observation is crucial for our arguments. Let M, N denote the staircase modules induced by S and T.

Lemma 6 *If $\epsilon < d_s(S, T)$, the only morphism from M to N^ϵ is $1 \mapsto 0$. If $\epsilon \geq d_s(S, T)$, every choice of $\lambda \in \mathbb{F}$ yields a morphism $1 \mapsto \lambda$ from M to N^ϵ.*

Proof In the first case, by construction, there exists some p such that $M_p = \mathbb{F}$, but $N_{p+\epsilon} = 0$. Hence, 0 is the only choice for λ.

In the second case, $M_p = \mathbb{F}$ implies $N_{p+\epsilon} = \mathbb{F}$ as well. It is easy to check that any choice of λ yields a compatible collection of maps, hence a morphism. □

In particular, there are morphisms $M \to N$ given by arbitrary elements of $\mathbb{F}$ if and only if $S \subseteq T$. As a consequence, we can characterize morphisms of direct sums of staircase modules.

Lemma 7 *Let $M = \oplus_{i=1}^{n} M_i$ and $N = \oplus_{j=1}^{n} N_j$ be direct sums of staircase modules. Then a collection of maps $\phi_p \colon M_p \to N_p$ is a morphism if and only if the restriction to M_i and N_j is a morphism for any $i, j \in \{1, \ldots, n\}$. Therefore, a morphism ϕ is determined by an $(n \times n)$-matrix with entries in $\mathbb{F}$.*

Proof Let $p \leq q$, and consider the following diagram:

$$\begin{tikzcd} M_p \arrow[r, "M_{p\to q}"] \arrow[d, "\phi_p"] & M_q \arrow[d, "\phi_q"] \\ N_p \arrow[r, "N_{p\to q}"'] & N_q \end{tikzcd}$$

We have $M_p = \oplus_{i=1}^{n} (M_i)_p$ and $N_q = \oplus_{j=1}^{n} (N_j)_q$. Thus the diagram above commutes if and only if for all i and j, the restrictions of the two compositions to $(M_i)_p$ and $(N_j)_q$ are the same, since a linear transformation is determined by what happens on basis elements. This is again equivalent to the following diagram commuting for all i and j, where $(\phi_i^j)_p$ is the restriction of ϕ_p to M_i and N_j:

$$\begin{tikzcd} (M_i)_p \arrow[r, "(M_i)_{p\to q}"] \arrow[d, "(\phi_i^j)_p"] & (M_i)_q \arrow[d, "(\phi_i^j)_q"] \\ (N_j)_p \arrow[r, "(N_j)_{p\to q}"'] & (N_j)_q \end{tikzcd}$$

But the collection of ϕ_p forms a morphism if and only if the first diagram commutes for all $p \leq q$, and the restriction of ϕ_p to M_i and N_j forms a morphism if and only if the second diagram commutes for all $p \leq q$. Thus we have proved the desired equivalence. $\square$

Observe that the matrix described in Lemma 6 is simply $\phi_p \colon \oplus_{i=1}^{n} (M_i)_p \to \oplus_{j=1}^{n} (N_j)_p$ written as a matrix in the natural way for any p contained in the support of M_i for all i.

Lemma 8 *Let M, N be direct sums of staircase modules as above and $\phi \colon M \to N^{\epsilon}$ and $\psi \colon N \to M^{\epsilon}$ be morphisms. Then ϕ and ψ form an ϵ-interleaving if and only if their associated $(n \times n)$-matrices are inverse to each other.*

Proof The composition $\psi^{\epsilon} \circ \phi \colon M \to M^{2\epsilon}$ is represented by the matrix BA, as one can see by restricting to a single point contained in all relevant staircases as in the observation above. The morphism $\mathrm{Sh}_M(2\epsilon) \colon M \to M^{2\epsilon}$ is represented by the identity matrix. By definition, (ϕ, ψ) is an interleaving if and only if these are equal and the corresponding statement holds for $\phi^{\epsilon} \circ \psi$, so the statement follows. $\square$

As a consequence, we obtain the following intermediate result.

Theorem 5 *Let (n, P, Q) be a CI-instance and let $S_1, \dots, S_n$, $T_1, \dots, T_n$ be staircases such that*

$$d_S(S_i, T_j) = \begin{cases} 3 & \text{if } (i, j) \in P \\ 1 & \text{if } (i, j) \notin P \end{cases} \quad d_S(T_j, S_i) = \begin{cases} 3 & \text{if } (j, i) \in Q \\ 1 & \text{if } (j, i) \notin Q \end{cases}$$

Write M_i, N_j for the modules associated to S_i, T_j, respectively, and $M := \oplus M_i$ and $N := \oplus N_j$. Then

$$d_I(M, N) = \begin{cases} 1 & \text{if } (n, P, Q) \in CI \\ 3 & \text{if } (n, P, Q) \notin CI \end{cases}$$

Proof Assume first that $(n, P, Q) \in CI$. Let A, B be a solution. We show that A and B define morphisms from M to N^1 and from N to M^1. We restrict to the map from M to N^1, as the other case is symmetric. By Lemma 7, it suffices to show that the map from M_i to N_j^1 is a morphism. This map is represented by the entry A_{ij}. If $(i, j) \in P$, $A_{ij} = 0$ by assumption, and the 0-map is always a morphism. If $(i, j) \notin P$, $d_S(S_i, T_j) = 1$ by construction. Hence, by Lemma 6 any field element yields a morphism. This shows that A and B define a pair of valid morphisms, and by Lemma 8 this pair is an 1-interleaving, as $AB = I_n$. Also with Lemma 6, it can easily be proved that the only morphism $M \to N^\epsilon$ with $\epsilon < 1$ is the 0-map. Hence, $d_I(M, N) = 1$ in this case.

Now assume that $(n, P, Q) \notin CI$. It is clear that M and N as constructed are 3-interleaved: the matrix I_n yields a valid morphism from M to N^3 and from N to M^3 with Lemma 6. Assume for a contradiction that there exists an ϵ-interleaving between M and N represented by matrices A, B, with $\epsilon < 3$. For $(i, j) \in P$, since $d_s(M_i, N_j) = 3 > \epsilon$, Lemma 6 implies that the entry $A_{i,j}$ must be equal to 0. Likewise, $B_{j,i} = 0$ whenever $(j, i) \in Q$. By Lemma 8, $AB = I_n$, and it follows that A and B constitute a solution to the CI-instance (n, P, Q), a contradiction. □

4.2 Construction of the Staircases

To prove Theorem 2, it suffices to construct staircases $S_1, \dots, S_n$, $T_1, \dots, T_n$ with the properties from Theorem 5, in polynomial time.

To describe our construction, we consider to two "base staircases" which we depict in Fig. 2. In what follows, a shift of a point a by $(1, 1)$ means replacing a with the point $a - (1, 1)$. The base staircase S is formed by the points $(-t, t)$ for $t = -4n^2, -4n^2 + 2, \dots, 4n^2$, but with the right side (i.e., the points with negative t) shifted by $(1, 1)$. Likewise, the base staircase T consists of the same points, but with the left side shifted by $(1, 1)$. We observe immediately that the staircase distance of the two base staircases is equal to 1 in either direction. We call the points defining the staircases *corners* from now on.

Now we associate to every entry in P a corner in the left side of S (that is, some $(-t, t)$ with $t > 0$). We also associate with the entry a corner in T, namely the shifted point $(-t - 1, t - 1)$. We do this in a way that between two associated corners of

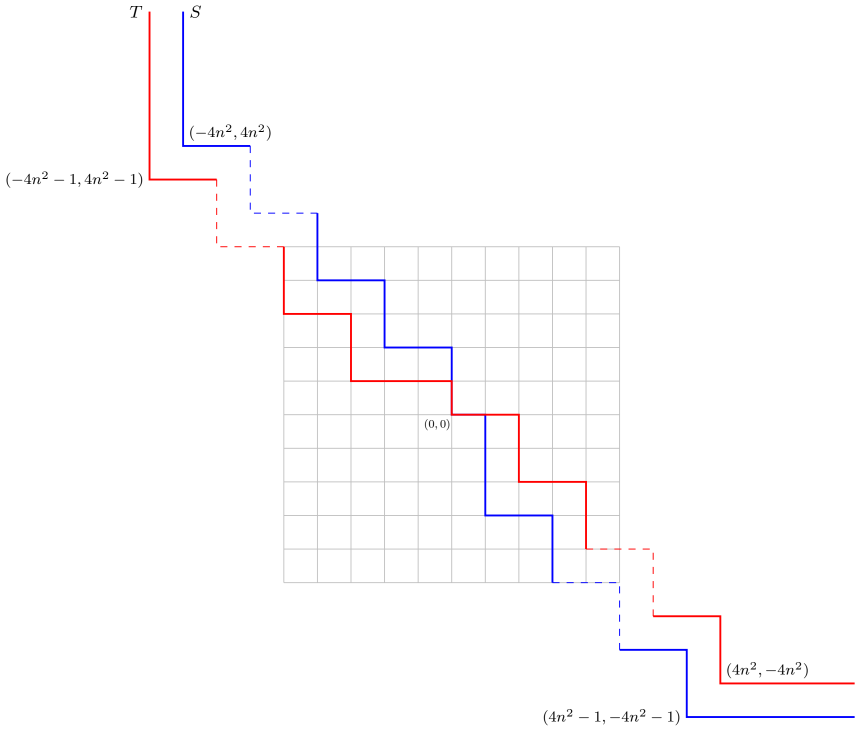


Fig. 2 The base staircases S and T

S, there is at least one corner of the staircase that is not associated. Note that this is always possible because $|P| \leq n^2$ and we have $2n^2$ corners on the left side. We associate corners to entries of Q in the symmetric way, using the right side of the base staircases.

We construct the staircases S_i and T_j out of the base staircases S and T, only shifting associated corners by $(2, 2)$ or $(-2, -2)$ according to P and Q. Specifically, for the staircase S_i, we start with S and for any entry (i, j) in P, we shift the associated corner of S by $(2, 2)$. For every entry (j, i) in Q, we shift the associated corner by $(-2, -2)$. The resulting (partially) shifted version of S defines S_i.

T_j is defined symmetrically: for every $(i, j) \in P$, we shift the associated corner by $(-2, -2)$. For every $(j, i) \in Q$, we shift the associated corner by $(2, 2)$.

We next analyze the staircase distance of S_i and T_j. We observe that, because there is an unassociated corner in-between any two associated corners, the $\pm(2, 2)$ shifts of distinct corners do not interfere with each other. Hence, it suffices to consider the distance of one associated corner of S_i to T_j. Fix the corner c_S of S associated to some entry $(k, \ell) \in P$. Let c_T denote the associated corner of T, that is, $c_T = c_S - (1, 1)$. See Fig. 3 (left) for an illustration. If $k \neq i$ and $\ell \neq j$, neither c_S nor c_T gets shifted, and since $c_T \leq c_S$, the shift required from c_T to reach c_S is 0. If $k = i$ and $\ell \neq j$, then c_S gets shifted by $(2, 2)$, and the required shift is 1 (see second picture of Fig. 3).

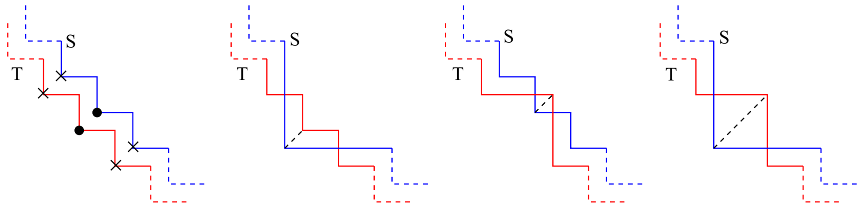


Fig. 3 Left: the associated corners c_S (on staircase S) and c_T (on staircase T) are marked by black circles. The two neighboring corners on both staircases (marked with x) are not associated and hence not shifted in the construction. Second and third picture: the cases (i, ℓ) with $\ell \neq j$ and (k, j) with $k \neq i$. In both cases, the directed staircase distance is 1, as illustrated by the dashed line. Right: the case (i, j). In that case, a shift of 3 is necessary to move the corner of T to S

If $k \neq i$ and $\ell = j$, c_T gets shifted by $(-2, -2)$, the required shift is also 1 (see 3rd picture of Fig. 3). If $k = i$ and $\ell = j$, both c_S and c_T get shifted, and the distance of the shifted c_T to reach c_S increases to 3 (see 4th picture of Fig. 3). This argument implies that the (directed) staircase distance from S_i to T_j is 3 if $(i, j) \in P$, and 1 otherwise. A completely symmetric argument works for $d_S(T_j, S_i)$, inspecting the corners associated to Q.

Finally, it is clear that the size and construction time of each S_i and each T_j is polynomial in n. As remarked at the end of Sect. 3, the staircase module can be brought in graded matrix representation in polynomial time in n, and the same holds for the direct sum of these modules. This finishes the proof of Theorem 2.

With the construction of M and N from Theorem 2 fresh in mind, we can explain the obstacles to obtaining a constant bigger than 3. Exchanging 3 with another constant in Theorem 5 is not a problem; the proof would be exactly the same. The trouble is to construct S_i and T_j satisfying the conditions in Theorem 5 if 3 is replaced by some $\epsilon > 3$. In that case, one would have to force $d_S(S_i, T_j) \geq \epsilon$ for $(i, j) \in P$ and $d_S(T_j, S_i) \geq \epsilon$ for $(j, i) \in Q$, while still keeping $d_S(S_i, T_j) \leq 1$ for $(i, j) \notin P$ and $d_S(T_j, S_i) \leq 1$ for $(j, i) \notin Q$. As we have shown, letting $d_S(S_i, T_j) = 3$ when $(i, j) \in P$ can be done. However, even if $(i, j) \in P$, there might be i', j' such that $(i, j'), (i', j) \notin P$ and $(j', i') \notin Q$, implying

$$\begin{aligned} d_S(S_i, T_{j'}) &\leq 1, \\ d_S(T_{j'}, S_{i'}) &\leq 1, \\ d_S(S_{i'}, T_j) &\leq 1. \end{aligned}$$

which gives $d_S(S_i, T_j) \leq 3 < \epsilon$ by the triangle inequality. This proves that one cannot simply increase the constant in Theorem 5, change the construction of S_i and T_j, and get a better result. That is not to say that using CI problems to improve Theorem 4 is necessarily hopeless, but it would not come as a surprise if a radically new approach is needed, if the theorem can be improved at all.

This is related to questions of stability, more precisely of whether $d_B(B(M), B(N)) \leq 3d_I(M, N)$ is true for staircase decomposable modules, where d_B is the bottleneck distance. We have associated pairs of modules to CI problems in a way such that interleavings correspond to solutions of the CI problems. Matchings between the barcodes of the modules (which is what gives rise to the bottleneck distance) correspond to solutions to the CI problems of a particular simple form, namely with a single non-zero entry in each column and row of each matrix. Claiming that $d_B(B(M), B(N)) \leq 3d_I(M, N)$ is then related to claiming that if a CI problem has a solution, then a "weakening" of the CI problem has a solution of this simple form. We will not go into details about this, other than to say that there are questions that can be formulated purely in terms of CI problems whose answers could have very interesting consequences for the study of interleavings, also beyond the work done in this paper.

5 Indecomposable Modules

Fix a CI problem (n, P, Q) as in the previous section and let M and N be the associated persistence modules. We shall now construct two indecomposable persistence modules $\widehat{M}$ and $\widehat{N}$ such that $\widehat{M}$ and $\widehat{N}$ are ϵ-interleaved if and only if M and N are ϵ-interleaved. In what follows we construct $\widehat{M}$; the construction of $\widehat{N}$ is completely analogous.

Recall that a staircase module can be described by a set of generators, or corners. Let $u = (x, x)$ be a point larger than all the corners defining the staircases making up M and N. Observe the following: $\dim M_u = n$, $M_{u\to p}$ is the identity morphism for any $p \geq u$.

Let $s_i = x + 7 + i/(n+1)$,[1] for $0 \leq i \leq n+1$. Define $\widehat{M}$ at $p \in \mathbb{R}^2$ as follows

$$\widehat{M}_p = \begin{cases} 0 & \text{if } p \geq (s_i, s_{n+1-i}) \text{ for some } 0 \leq i \leq n+1, \\ \mathbb{F} & \text{if } p \in [s_i, s_{i+1}) \times [s_{n-i}, s_{n-i+1}) \text{ for some } 0 \leq i \leq n, \\ M_p & \text{otherwise.} \end{cases}$$

Trivially, $\widehat{M}_{p\to q}$ is the 0 morphism if $\widehat{M}_p = 0$ or $\widehat{M}_q = 0$. For $p \leq q$ such that $M_p = \widehat{M}_p$ and $M_q = \widehat{M}_q$, let $\widehat{M}_{p\to q} = M_{p\to q}$, and for $p, q \in [s_i, s_{i+1}) \times [s_{n-i}, s_{n-i+1})$ let $\widehat{M}_{p\to q} = 1_{\mathbb{F}}$. It remains to consider the case that $M_p = \widehat{M}_p$ and $q \in [s_i, s_{i+1}) \times [s_{n-i}, s_{n-i+1})$ for some i. Observe that all the internal morphisms are fully specified once we define $\widehat{M}_{u\to q}$. Indeed, if $p \geq u$, then $\widehat{M}_{u\to p}$ is the identity, which forces $\widehat{M}_{p\to q} = \widehat{M}_{u\to q}$. For any other p we can always choose an $r \geq p$ such that $r \geq u$ and $M_r = \widehat{M}_r$. The morphism $\widehat{M}_{p\to q}$ is then given by $\widehat{M}_{p\to q} = \widehat{M}_{r\to q} \circ \widehat{M}_{p\to r} = \widehat{M}_{u\to q} \circ \widehat{M}_{p\to r}$. We conclude by specifying the following morphism

$$\widehat{M}_{u\to q} = \begin{cases} \pi_i \text{ (projection onto coordinate } i) & \text{if } 1 \leq i \leq n \\ \pi_1 + \pi_2 + \ldots + \pi_n & \text{if } i = 0. \end{cases}$$

Observe that we have a morphism $\pi^M : M \to \widehat{M}$ given by

$$\pi^M_p = \begin{cases} \text{id} & \text{if } M_p = \widehat{M}_p \\ \widehat{M}_{u\to p} & \text{otherwise} \end{cases}$$

[1] 7 can be replaced with any $\delta > 6$.

Lemma 9 *The persistence module $\widehat{M}$ is indecomposable.*

Proof We recall the following useful trick: if M is *not* indecomposable, say $M \cong M' \oplus M''$, then the projections $M \to M'$ and $M \to M''$ define morphisms which are not given by multiplication with a scalar. Hence, it suffices to show that any endomorphism $\phi\colon \widehat{M} \to \widehat{M}$ is multiplication by a scalar. Furthermore, observe that any endomorphism ϕ of $\widehat{M}$ is completely determined by ϕ_u. Let $e_i \in \mathbb{F}^n$ denote the vector $(0, \ldots, 0, 1, 0, \ldots, 0)$ where the non-zero entry appears at the i-th index. For $0 \leq i \leq n$, ϕ must be such that the following diagram commutes

$$\begin{tikzcd} \widehat{M}_u = \mathbb{F}^n \arrow[r] \arrow[d, "\phi_u"] & \mathbb{F} = \widehat{M}_{(s_i, s_{n-i})} \arrow[d, "\phi_{(s_i, s_{n-i})} = \lambda_i"] \\ \widehat{M}_u = \mathbb{F}^n \arrow[r] & \mathbb{F} = \widehat{M}_{(s_i, s_{n-i})} \end{tikzcd}$$

For $1 \leq i \leq n$ this yields that

$$\pi_i(\phi_u(e_j)) = \begin{cases} \lambda_i \in \mathbb{F} & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}.$$

In particular, we see that $\phi_u(e_i) = \lambda_i e_i$. For $i = 0$ we get

$$\lambda_0 e_i = \lambda_0 \cdot (\pi_1 + \ldots + \pi_n)(e_i) = (\pi_1 + \ldots + \pi_n)(\lambda_i e_i) = \lambda_i e_i.$$

We conclude that $\lambda_i = \lambda_0$ and that $\phi_u = \lambda_0 \cdot \mathrm{id}$. □

Lemma 10 *Fix $1 \leq \epsilon \leq 3$. $\widehat{M}$ and $\widehat{N}$ are ϵ-interleaved if and only if M and N are ϵ-interleaved.*

Proof Assume that $\phi\colon M \to N^\epsilon$ and $\psi\colon N \to M^\epsilon$ form an ϵ-interleaving pair. Define $\widehat{\phi}\colon \widehat{M} \to \widehat{N}^\epsilon$ and $\widehat{\psi}\colon \widehat{N} \to \widehat{M}^\epsilon$ by

$$\widehat{\phi}_p = \begin{cases} \pi^N_{p+\epsilon} \circ \phi_p & \text{if } M_p = \widehat{M}_p \\ 0 & \text{otherwise.} \end{cases}, \quad \widehat{\psi}_p = \begin{cases} \pi^M_{p+\epsilon} \circ \psi_p & \text{if } N_p = \widehat{N}_p \\ 0 & \text{otherwise.} \end{cases}$$

We will show that these two morphisms constitute an ϵ-interleaving pair. Let $p \in \mathbb{R}^2$ and consider the following two cases:

1. Assume that $N_{p+\epsilon} = \widehat{N}_{p+\epsilon}$. Under this assumption, we have that $\pi^N_{p+\epsilon} = \mathrm{id}$, and thus $\phi_p = \widehat{\phi}_p$. Using that ψ and ϕ form an ϵ-interleaving pair, and that $\pi^M_p = \mathrm{id}$, we get:

$$\widehat{\psi}_{p+\epsilon} \circ \widehat{\phi}_p = \pi^M_{p+2\epsilon} \circ \psi_{p+\epsilon} \circ \phi_p = \pi^M_{p+2\epsilon} \circ M_{p \to p+2\epsilon} = \widehat{M}_{p \to p+2\epsilon} \circ \pi^M_p = \widehat{M}_{p \to p+2\epsilon}.$$

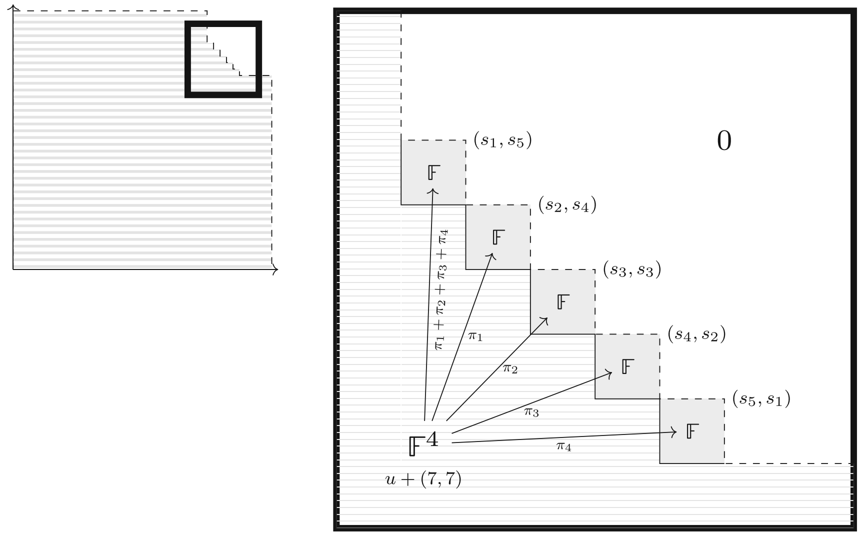


Fig. 4 Left: $\widehat{M}$ coincides with M on the restriction to the shaded subset of $\mathbb{R}^2$. Right: this shows the modification done to M in order to obtain an indecomposable persistence module $\widehat{M}$ in the case $n = 4$

2. Assume that $N_{p+\epsilon} \neq \widehat{N}_{p+\epsilon}$. Since $\epsilon \geq 1$, it follows by construction that $\widehat{M}_{p+2\epsilon} = 0$. Hence, the interleaving condition is trivially satisfied.

Symmetrically we get that $\widehat{\phi}_{p+\epsilon} \circ \widehat{\phi}_p = \widehat{N}_{p \to p+2\epsilon}$. Hence, $\widehat{M}$ and $\widehat{N}$ are ϵ-interleaved.

Conversely, assume that $\widehat{\phi}$ and $\widehat{\psi}$ define an interleaving pair between $\widehat{M}$ and $\widehat{N}$. Define $\phi\colon M \to N^\epsilon$ and $\psi\colon N \to M^\epsilon$ by

$$\phi_p = \begin{cases} \widehat{\phi}_u & \text{if } p \geq u \\ \widehat{\phi}_p & \text{otherwise} \end{cases}, \quad \psi_p = \begin{cases} \widehat{\psi}_u & \text{if } p \geq u \\ \widehat{\psi}_p & \text{otherwise} \end{cases}.$$

By construction, $\widehat{M}_p = M_p$ and $\widehat{N}_p = N_p$ for all $p < u + (7, 7)$. This implies that $\widehat{\phi}_p = \widehat{\phi}_u$ and $\widehat{\psi}_p = \widehat{\psi}_u$ for all $p < u + (7 - \epsilon, 7 - \epsilon)$. Hence, for any $p \leq u$ we must have that

$$\psi_{p+\epsilon} \circ \phi_p = \widehat{\psi}_{p+\epsilon} \circ \widehat{\phi}_p = \widehat{M}_{p \to p+2\epsilon} = M_{p \to p+2\epsilon}.$$

Similarly we get that $\phi_{p+\epsilon} \circ \psi_p = N_{p \to p+2\epsilon}$ for all such p. In particular, by considering the case $p = u$, we see that ϕ_u and ψ_u are mutually inverse matrices. It follows readily that the interleaving condition is satisfied for all $p \not\leq u$. □

With the two previous results at hand, we can state the following corollary of Theorem 4.

Corollary 1 1- INTERLEAVING *is NP-complete and* c-APPROX- INTERLEAVING- DISTANCE *is NP-hard for* $c < 3$*, even if the input modules are restricted to indecomposable modules.*

Proof We only prove hardness of 1- INTERLEAVING, the remaining statements follow with the same methods. Given a CI-instance (n, P, Q), we use the construction from Sect. 4 to construct two persistence modules M and N. Then we transform them into the indecomposable modules $\widehat{M}$ and $\widehat{N}$ as above. Note that this transformation can be performed in polynomial time in n by introducing up to n relations at the lower-left corners of the $(n+1)$ rectangles in Fig. 4. Hence, an algorithm to decide 1-INTERLEAVING for the case of indecomposable modules would solve CI in polynomial time. □

6 One-Sided Stability

The results of the previous sections also apply in the setting of one-sided stability. Here we give a brief introduction to the topic; see [1] for a thorough introduction.

Let $f: M \to N$ be a morphism. The linear map $M_{p\to q}$ induces a linear map $\ker(f_p) \to \ker(f_q)$ by restriction, and $N_{p\to q}$ induces a linear map $\operatorname{coker}(f_p) \to \operatorname{coker}(f_q)$ by taking a quotient, as one can readily verify. We say that f has *ϵ-trivial kernel* if the map $\ker(f_p) \to \ker(f_{p+\epsilon})$ is the 0-map for all $p \in \mathbb{R}^2$. Likewise, we say that f has *ϵ-trivial cokernel* if $\operatorname{coker}(f_p) \to \operatorname{coker}(f_{p+\epsilon})$ is the 0-map for all $p \in \mathbb{R}^2$. If f has 0-trivial kernel (cokernel), then we say that f is *injective* (*surjective*). The following lemma follows readily from the definition of an ϵ-interleaving.

Lemma 11 *If $f: M \to N^\epsilon$ is an ϵ-interleaving morphism (i.e., it forms an ϵ-interleaving with some $g: N \to M^\epsilon$), then f has 2ϵ-trivial kernel and cokernel.*

In fact, Bauer and Lesnick [1] show that in the case of persistence modules over $\mathbb{R}$, M and N are ϵ-interleaved if and only if there exists a morphism $f: M \to N^\epsilon$ with 2ϵ-trivial kernel and cokernel. They also observe that this equivalence does not generalize to two parameters. However, it *is* true (and the proof is very similar to the one given below) that if there exists a morphism $f: M \to N^\epsilon$ with ϵ-trivial kernel and cokernel, then M and N are ϵ-interleaved. Hence, there is a close connection between interleavings and morphisms with kernels and cokernels of bounded size also in the multi-parameter landscape.

Lemma 12 *For any injective $f: M \to N^\epsilon$ with 2ϵ-trivial cokernel, there exists a morphism $g: N \to M^\epsilon$ such that f and g constitute an ϵ-interleaving pair.*

Proof We have the following commutative square for all $p \in \mathbb{R}^2$:

$$\begin{tikzcd} M_p \arrow[r, "M_{p\to p+2\epsilon}"] \arrow[d, "f_p"] & M_{p+2\epsilon} \arrow[d, "f_{p+2\epsilon}"] \\ N_{p+\epsilon} \arrow[r, "N_{p+\epsilon\to p+3\epsilon}"'] & N_{p+3\epsilon} \end{tikzcd}$$

Let $n \in N_{p+\epsilon}$. Since f has 2ϵ-trivial cokernel and f is injective, there exists a unique $m \in M_{p+2\epsilon}$ such that $f_{p+2\epsilon}(m) = N_{p+\epsilon\to p+3\epsilon}(n)$. Define $g_p: N_{p+\epsilon} \to M_{p+2\epsilon}$ by $g_p(n) = m$. Doing this for all $p \in \mathbb{R}^2$ defines a morphism $g: N^\epsilon \to M^{2\epsilon}$ and we leave it to the reader to verify that f and $g^{-\epsilon}$ define an ϵ-interleaving pair. □

For fixed parameters $s, t \in [0, \infty]$, we consider the following computational problem:

s-t-TRIVIAL- MORPHISM: Given two persistence modules M, N in graded matrix representation, decide whether there exists a morphism $f: M \to N$ with s-trivial kernel and t-trivial cokernel.

Choosing $s = t = 0$ simply asks whether the modules are isomorphic, which can be decided in polynomial time [6]. On the other extreme, $s = t = \infty$ imposes no conditions on the morphism, which turns the decision problem to be trivially true, using the 0-morphism. We show

Theorem 6 s-t-TRIVIAL- MORPHISM *is NP-complete for every* $(s, t) \notin \{(0, 0), (\infty, \infty)\}$.

The case (s, t) is computationally equivalent to the case (cs, ct) with $c > 0$, since we can scale all grades occurring in M and N by a factor of c. So, it suffices to prove hardness of 2-t-TRIVIAL- MORPHISM, s-2-TRIVIAL- MORPHISM (we will see that the choice of 2 will be convenient in the argument), ∞-0-TRIVIAL- MORPHISM and 0-∞-TRIVIAL- MORPHISM.

Note that for any choice of s and t, s-t-TRIVIAL- MORPHISM is in NP. The argument is similar to the first part of the proof of Theorem 3: a morphism can be specified in polynomial size with respect to the module sizes, and we can check the triviality conditions of the kernel and cokernel by considering ranks of matrices.

For the hardness, we first focus on the case $(s, 2)$; hence, we want to decide the existence of a morphism with s-trivial kernel and 2-trivial cokernel. The following simple observation is the key insight of the proof.

Lemma 13 *Let M, N be as in Theorem 2. Any morphism $f: M \to N^1$ with 2-trivial cokernel is injective.*

Proof Recall that both M and N are direct sums of n staircase modules. Let p be any point such that $\dim M_p = \dim N_p = n$, and observe that $M_{p \to q} = \mathrm{id}_{\mathbb{F}}$, $N_{p \to q} = \mathrm{id}_{\mathbb{F}}$ and $f_p = f_q$ for all $q \geq p$. In particular, if $q = p + (2, 2)$, the induced map $\mathrm{coker}(f_p) \to \mathrm{coker}(f_q)$ is the identity, and since f has a 2-trivial cokernel by assumption, the map is also the 0-map. Hence $\mathrm{coker}(f_p)$ is trivial, implying that the map f_p is surjective, and hence also injective, and the same holds for f_q with $q \geq p$.

Now consider f_r for an arbitrary $r \in \mathbb{R}^2$. Let $q \geq r$ be a point satisfying $q \geq p$. Since the internal morphisms of M are all injective and f_p is injective, so is f_r. $\square$

In other words, for M and N^1 as above, the answer to s-2-TRIVIAL- MORPHISM is independent of s. Moreover, it follows:

Corollary 2 *With M, N as above, there exists a morphism $f: M \to N^1$ with 2-trivial cokernel and s-trivial kernel if and only if M and N are 1-interleaved.*

Proof If such a morphism exists, Lemma 13 guarantees that the morphism is in fact injective with 2-trivial cokernel. Lemma 12 with $\epsilon = 1$ guarantees that the modules are 1-interleaved.

Vice versa, if M and N are 1-interleaved, there is a morphism f with 2-trivial kernel and cokernel by Lemma 11. Again using Lemma 13 guarantees that f is injective, hence has a 0-trivial kernel. $\square$

Corollary 3 *s-2-TRIVIAL- MORPHISM is NP-hard for all $s \in [0, \infty]$.*

Proof Given a CI-instance, we transform it into modules M and N as in Sect. 4. Assuming a polynomial time algorithm for s-2-TRIVIAL- MORPHISM, we apply it on (M, N^1). If the algorithm returns that a morphism exists, we know by Corollary 2 that M and N are 1-interleaved and therefore, the CI-instance has a solution. If no morphism exists, M and N are not 1-interleaved and therefore, the CI-instance has no solution. We can thus solve the CI problem in polynomial time. $\square$

6.1 Dual Staircases

We will prove that 2-t-TRIVIAL- MORPHISM is NP-hard by a reduction from s-2-TRIVIAL- MORPHISM. First we need some notation. For a staircase S, let S° denote the interior of S, and for a staircase module M_l supported on a staircase S, we let M_l° denote the interval module supported on S°. Observe that there is a canonical injection $M_l^\circ \hookrightarrow M$ (given by $m \mapsto m$). It is also easy to see that $d_s(S, T) = d_s(S^\circ, T^\circ)$. Here d_s for interiors of staircases is defined in the obvious way. The reason why we look at interiors is technical: We eventually end up with a dual module $(M^\circ)^*$, and taking interiors makes sure the changes in this dual module happen at given points instead of "immediately after" the points, which is needed for a graded matrix representation of the module.

Lemma 14 *Let M and N be staircase decomposable modules. There exists an injection $f: M \to N$ with ϵ-trivial cokernel if and only if there exists an injection $f^\circ: M^\circ \to N^\circ$ with ϵ-trivial cokernel.*

Proof Let $M = \oplus_i M_i$ and $N = \oplus_j N_j$. Observe that $S \subseteq T$ if and only if $S^\circ \subseteq T^\circ$. Therefore, any morphism $M_i^\circ \to N_j^\circ$ extends to a morphism $M_j \to N_j$ in the obvious way. Conversely, any morphism $M \to N$ restricts to a morphism $M^\circ \to N^\circ$. It is not hard to see that extension and restriction are inverse functions. In particular, there is a one-to-one correspondence between morphisms $f: M \to N$ and $f^\circ: M^\circ \to N^\circ$.

Suppose f° is injective. For any point p, there exists a $\delta > 0$ such that $M_{p\to p+\delta}$ and $N_{p\to p+\delta}$ are isomorphisms, which also gives $f^\circ_{p+\delta} = f_{p+\delta}$. Since f° is injective, $f^\circ_{p+\delta} = f_{p+\delta}$ is, and by using the isomorphisms, we get that f_p is injective, too. Since p was arbitrary, we conclude that f is injective. The converse can be proved by using the dual fact that for any p, there exists a γ such that $M^\circ_{p-\gamma\to p}$ and $N^\circ_{p-\gamma\to p}$ are isomorphisms.

Suppose that coker f is not ϵ-trivial, so there is a p and an $m \in N_p$ such that $N_{p\to p+\epsilon}(m)$ is not in the image of $f_{p+\epsilon}$. Similarly to how we picked δ above, we can pick δ and γ with $\delta \le \gamma$ in a way that makes the following diagram commute, with equalities and isomorphisms as shown.

$$\begin{array}{ccccccc}
 & & M^\circ_{p+\delta} & \xrightarrow{\qquad\qquad\qquad} & & & M^\circ_{p+\epsilon+\gamma} \\
 & & \big\downarrow{=} & & & & \big\downarrow{=} \\
M_p & \xrightarrow{\cong} & M_{p+\delta} & \longrightarrow & M_{p+\epsilon} & \xrightarrow{\cong} & M_{p+\epsilon+\gamma} \\
\big\downarrow{f_p} & & \big\downarrow{f_{p+\delta}=f^\circ_{p+\delta}} & & \big\downarrow{f_{p+\epsilon}} & & \big\downarrow{f_{p+\epsilon+\gamma}=f^\circ_{p+\epsilon+\gamma}} \\
N_p & \xrightarrow{\cong} & N_{p+\delta} & \longrightarrow & N_{p+\epsilon} & \xrightarrow{\cong} & N_{p+\epsilon+\gamma} \\
 & & \big\downarrow{=} & & & & \big\downarrow{=} \\
 & & N^\circ_{p+\delta} & \xrightarrow{\qquad\qquad\qquad} & & & N^\circ_{p+\epsilon+\gamma}
\end{array}$$

All the horizontal maps are internal morphisms. We know that $N_{p\to p+\epsilon}(m) \in N_{p+\epsilon}$ is not in the image of $f_{p+\epsilon}$. Let $m' \in N^\circ_{p+\epsilon+\gamma}$ be the image of m along the maps in the above diagram. Then m' is in the image of $N^\circ_{p+\delta\to p+\epsilon+\gamma}$, but not in the image of $f^\circ_{p+\epsilon+\gamma}$. Since $(\epsilon+\gamma)-\delta \geq \epsilon$, this shows that f° is not ϵ-trivial. Again, the argument can be dualized to show the converse. □

For an interval $I \subseteq \mathbb{R}^2$, define the dual interval I^* as follows: $(x, y) \in I^*$ if and only if $(-x, -y) \in I$. And for an interval module M_I supported on I, let M_I^* denote be the interval module supported on I^*. If $M = \oplus_i M_i$ is a sum of interval modules M_i, then $M^* = \oplus_i (M_i)^*$. This is equivalent to considering M as a module indexed by $\mathbb{R}^2$ with the partial order reversed.

Let $M^\circ = \oplus_i M_i^\circ$ and $N^\circ = \oplus_j N_j^\circ$, where M_i° and N_j° are interval modules supported on interiors of staircases, and let $f^\circ\colon M^\circ \to N^\circ$. Observe that we can represent f° by a collection of matrices $\{A_p\}_{p\in\mathbb{R}^2}$, where A_p is the matrix representation of f_p° with respect to the bases given by the non-trivial elements of $\{(M_i^\circ)_p\}_i$ and $\{(N_j^\circ)_p\}_j$. Similarly, for any $p \leq q$, we can represent the linear maps $M^\circ_{p\to q}$ and $N^\circ_{p\to q}$ by matrices with respect to the obvious bases.

Importantly, representing f_p° by matrices A_p as above, we get a dual morphism $(f^\circ)^*\colon N^* \to M^*$ given by the matrices $\{(A_{-p})^T\}_{p\in\mathbb{R}^2}$. This induces a bijection between the set of morphisms from M° to N° and the set of morphisms from $(N^\circ)^*$ to $(M^\circ)^*$.

Lemma 15 *f° is an injection with ϵ-trivial cokernel if and only if $(f^\circ)^*$ is a surjection with ϵ-trivial kernel.*

Proof The first part is straightforward: the matrix A_p represents a surjective linear map if and only if A_p^T represents an injective linear map. Since $(f^\circ)^*_p = f^\circ_{-p}$, the result follows readily.

For the second part, let p be any point in $\mathbb{R}^2$, and let X be the matrix representation of the morphism $N^\circ_{p\to p+\epsilon}$ with respect to the basis given by the N_j°'s. Then, by construction, X^T is a matrix representation for $(N^\circ)^*_{-p-\epsilon\to -p}$ (with respect to the dual bases). Using the elementary fact that $\mathrm{col}(X) \subseteq \mathrm{col}(A_{p+\epsilon})$ if and only if $\ker(A^T_{p+\epsilon}) \subseteq \ker(X^T)$, where $\mathrm{col}(X)$ denotes the column space of X, we conclude

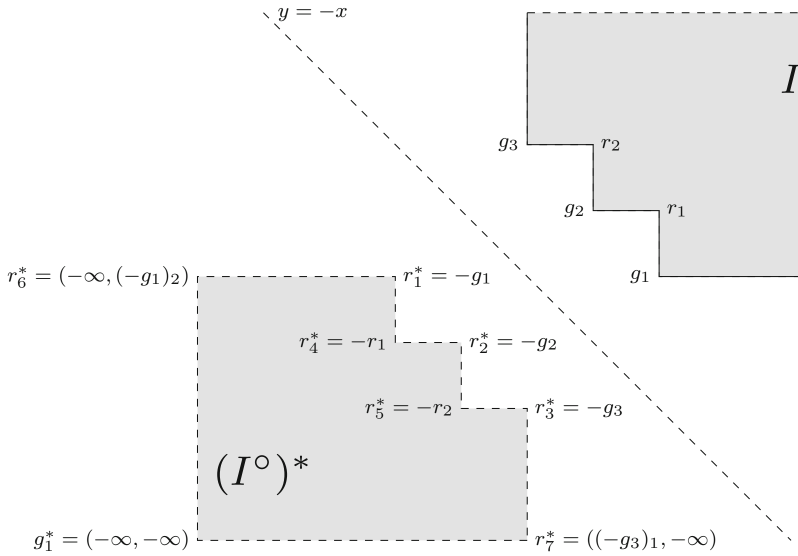


Fig. 5 The staircase module k_I supported on the interval I admits a graded matrix representation with $G = \{g_1, g_2, g_3\}$ and $R = \{r_1, r_2\}$. The module $(k_{I^\circ})^* = k_{(I^\circ)^*}$ admits a (generalized) graded matrix representation with $G^* = \{(-\infty, -\infty)\}$ and $R^* = \{-g_1, -g_2, -g_3, -r_1, -r_2, (-\infty, (-g_1)_2), ((-g_3)_1, -\infty)\}$. In the proof of Corollary 4, we may replace ∞ with $z \gg 0$ to obtain a proper graded matrix representation

that $\mathrm{im}(N^\circ_{p\to p+\epsilon}) \subseteq \mathrm{im}(f^\circ_{p+\epsilon})$ if and only if $\ker((f^\circ)^*_{-p-\epsilon}) \subseteq \ker((N^\circ)^*_{-p-\epsilon\to p})$. As p was arbitrary, this concludes the proof. □

Corollary 4 2-t-TRIVIAL- MORPHISM *is NP-hard for all* $t \in [0, \infty]$.

Proof This follows from the previous two lemmas and Corollary 3. There is however a technical obstacle arising from the fact that $(M^\circ)^*$ and $(N^\circ)^*$ have their generators at grade $(-\infty, -\infty)$. This problem is easy to solve, either by altering the graded matrix representation to allow such a generator, or by placing all generators at a sufficiently small value $p \in \mathbb{R}^2$ that is smaller than all corners of the staircase, see Fig. 5 for an illustration. Introducing such a minimal grade does not invalidate any of the given arguments—we omit the technical details. □

6.2 Surjective Morphisms

After Corollaries 3 and 4, all we have left to prove Theorem 6 is the cases ∞-0-TRIVIAL- MORPHISM and 0-∞-TRIVIAL- MORPHISM. Recall that these correspond to asking for a surjection in the first case and an injection in the second.

Lemma 16 ∞-0-TRIVIAL- MORPHISM *and* 0-∞-TRIVIAL- MORPHISM *are both NP-hard.*

Proof We will only prove the first case; the second follows by dualizing the arguments in an appropriate way, for instance by using dual staircases as above.

Recall that we have assumed $\mathbb{F}$ to be finite. Let q denote the number of elements in $\mathbb{F}$, and assume that ϕ is a 3CNF formula with n variables $\{x_1, \ldots, x_n\}$ and m clauses $\{c_1, \ldots, c_m\}$. We shall construct modules $M = A \oplus B \oplus (\oplus_{i=1}^n \oplus_{r=1}^q M_i^r)$ and $N = N_1 \oplus N_2$, where M_i^r, A, B, N_1 and N_2 are staircase modules, in such a way that there exists a surjection $M \to N$ if and only if ϕ is satisfiable. Importantly, we know from Lemma 7 that any morphism between staircase decomposable modules can be represented by a matrix with entries in $\mathbb{F}$. We only stated the result in the case where each module is built from the same number of staircases, but the same argument shows that a morphism $M \to N$ in this case is described by a $2 \times (nq + 2)$-matrix, which we shall assume is ordered in the following way

$$\begin{array}{cc} \begin{array}{cccccccc} A & B & M_1^1 & M_1^2 & \cdots & M_2^1 & \cdots & M_n^q \end{array} & \\ \begin{pmatrix} * & * & * & * & \cdots & * & \cdots & * \\ * & * & * & * & \cdots & * & \cdots & * \end{pmatrix} & \begin{array}{c} N_1 \\ N_2 \end{array} \end{array}$$

Furthermore, recall that any staircase module is defined by a set of generators, i.e., a set of incomparable points defining the "corners" of the staircase. It is not hard to see that a morphism $M \to N$ is surjective if and only if it is surjective at the all the corners points of N_1 and N_2.

Let

$$D = \{A, B, N_1, N_2\} \bigcup_{i,r} \{M_i^r\}$$

and let $S \subseteq \mathbb{R}^2$ be a set of pairwise incomparable points. Any function $G\colon S \to P(D)$, where $P(D)$ is the power set of D, specifies the modules in the decomposition of M and N by enforcing that $X \in D$ has a corner point at $s \in S$ if and only if $X \in G(s)$. In what follows we shall define such a function G in four steps, and define the staircase modules in D accordingly.

Let $S = \{a, b, g_i^r, g_i^{r,s}, h_j^{y,z,w}\}$ be a set of distinct incomparable points in $\mathbb{R}^2$, where i, j, r, s, y, z, w run through indices which will be defined as we define G. In the initial step, we define $G(a) = \{A, N_1\}$ and $G(b) = \{B, N_2\}$. The addition of these corners enforce that the matrix (in the ordering given above) must be of the form

$$\begin{pmatrix} 1 & 0 & * & \ldots & * \\ 0 & 1 & * & \ldots & * \end{pmatrix}.$$

This can be seen as follows: since a and b are incomparable, and $a \in A$ while $a \notin N_2$, we must have that $N_2 \not\subseteq A$. Lemma 6 allows us to conclude that the only morphism from $A \to N_2$ is the trivial one. Similarly we see that the morphism $B \to N_1$ must be the trivial one. Furthermore, since $M_a = A_a$ and $N_a = (N_2)_a$, surjectivity at a implies that $A \to N_1$ must be non-zero, which gives the non-zero entry in the first column. We can multiply any column in the matrix with a non-zero element without changing the validity or surjectivity of the morphism, so we can assume that this element is 1. Similarly we get a 1 in the second row of the second column.

We proceed our inductive step by defining $G(g_i^r) = \{A, M_i^r, N_1, N_2\}$ for all $1 \leq i \leq n$ and $1 \leq r \leq q$. Restricting the matrix to the columns corresponding to A and M_i^r we get

$$\begin{pmatrix} 1 & * \\ 0 & * \end{pmatrix}.$$

For the morphism to be surjective at the point g_i^r, this matrix must be of full rank. Therefore, we can write it as

$$\begin{pmatrix} 1 & d_i^r \\ 0 & 1 \end{pmatrix},$$

where we again have used the fact that we can scale columns by non-zero constants. In other words, any surjection $M \to N$ must be of the form

$$\begin{pmatrix} 1 & 0 & d_1^1 & \dots & d_1^q & d_2^1 & \dots & d_2^q & \dots & d_n^1 & \dots & d_n^q \\ 0 & 1 & 1 & \dots & 1 & 1 & \dots & 1 & \dots & 1 & \dots & 1 \end{pmatrix}.$$

Continuing, let $G(g_i^{r,s}) = \{M_i^r, M_i^s, N_1, N_2\}$, for all $1 \leq i \leq n$ and $1 \leq r < s \leq q$. Restricting the matrix to the columns corresponding to M_i^r and M_i^s yields the matrix

$$\begin{pmatrix} d_i^r & d_i^s \\ 1 & 1 \end{pmatrix}.$$

For the matrix to be surjective at $g_i^{r,s}$, also this matrix must be of full rank. In particular, it must be the case that $d_i^r \neq d_i^s$, and therefore exactly one of $d_i^1, \dots, d_i^q$ equals 0. We will interpret $d_i^1 = 0$ as choosing x_i to be false, and $d_i^1 \neq 0$ as choosing x_i to be true.

What remains is to encode the clauses of ϕ. For a clause c_j, let $x_{\alpha_{j,1}}, x_{\alpha_{j,2}}, x_{\alpha_{j,3}}$ be the variables such that either the variable itself or its negation occurs in c_j, with $\alpha_{j,1} < \alpha_{j,2} < \alpha_{j,3}$. For $1 \leq i \leq 3$, let $X_j^i = \{1\}$ if $x_{\alpha_{j,i}}$ occurs in c_j; if instead its negation occurs, let $X_j^i = \{2, \dots, q\}$. For example, if $c_j = x_1 \vee \neg x_2 \vee \neg x_4$, then $\alpha_{j,1} = 1$, $\alpha_{j,2} = 2$ and $\alpha_{j,3} = 4$, and $X_j^1 = \{1\}$, $X_j^2 = \{2, \dots, q\}$ and $X_j^3 = \{2, \dots, q\}$. Define $G(h_j^{y,z,w}) = \{B, M_{\alpha_{j,1}}^r, M_{\alpha_{j,2}}^s, M_{\alpha_{j,3}}^t, N_1, N_2\}$, for all $1 \leq j \leq m$ and $y \in X_j^1$, $z \in X_j^2$, $w \in X_j^3$.

This time, the following submatrix must have rank 2 for all $h_j^{y,z,w}$ with j, y, z, w as above.

$$\begin{pmatrix} 0 & d_{\alpha_{j,1}}^y & d_{\alpha_{j,2}}^z & d_{\alpha_{j,3}}^w \\ 1 & 1 & 1 & 1 \end{pmatrix} \tag{5}$$

At this stage, we have concluded the construction of the modules and no further restrictions will be imposed on the matrix. In particular, the above shows that there exists a surjection $M \to N$ if and only if there is an assignment $d_i^r \in \mathbb{F}$ such that the following is satisfied:

- $\mathbb{F} = \{d_i^1, \ldots, d_i^q\}$ for all $1 \leq i \leq n$.
- The matrix of (5) has full rank for every $h_j^{y,z,w}$.

We show that this is equivalent to ϕ being satisfiable.

"$\Rightarrow$": Assume that ϕ is satisfiable and pick a satisfying assignment. If x_i is set to false, then define $d_i^1 = 0$. If x_i is set to true, then define $d_i^2 = 0$. In both cases, we assign the remaining variables values such that $\mathbb{F} = \{d_i^1, \ldots, d_i^q\}$ for all $1 \leq i \leq n$. Consider the clause c_j as above, and assume that $x_{\alpha_{j,l}}$ is assigned a truth value such that the literal associated to $x_{\alpha_{j,l}}$ in c_j evaluates to true. Then $d_{\alpha_{j,l}}^y \neq 0$ for all $y \in X_j^l$, implying that the matrix of (5) has rank 2 for all $h_j^{y,z,w}$.

"$\Leftarrow$": Assume an assignment of the variables d_i^r satisfying the two bullet points above, and set x_i to be false if $d_i^1 = 0$, and true otherwise. Consider the clause c_j as above, and observe that there exists an index $y \in X_j^l$ such that $d_{\alpha_{j,l}}^y = 0$ if and only if the literal in c_j associated to $x_{\alpha_{j,l}}$ evaluates to false. In particular, c_j evaluates to true if and only if at least one of $d_{\alpha_{j,1}}^y$, $d_{\alpha_{j,2}}^z$ and $d_{\alpha_{j,3}}^w$ is non-zero for every $(x, y, z) \in X_j^1 \times X_j^2 \times X_j^3$. This is equivalent to the matrix of (5) having full rank for every $h_j^{y,z,w}$.

In the end, we have a reduction from 3SAT to ∞-0-TRIVIAL- MORPHISM. To complete the proof, we must show that the instance of ∞-0-TRIVIAL- MORPHISM can be constructed in polynomial time in the input size of the instance of 3SAT. As we have assumed q to be fixed and finite, it suffices to observe that M is defined by $nq + 2$ staircase modules, while N is a sum of 2 staircase modules, and that each of these are generated by at most $2 + nq + n\binom{q}{2} + m(q-1)^3$ generators. We remark that the generators can be chosen along the antidiagonal $x = -y$ in $\mathbb{R}^2$. $\square$

An interesting point is that the number of generators of the staircases in the proof increases with the size of $\mathbb{F}$. Hence, the proof strategy only applies in the setting of a finite field (with a constant number of elements).

We conclude this section by remarking that 0-∞-TRIVIAL- MORPHISM is equivalent to the problem of deciding if a module M' is a *submodule* of another persistence module M. Interestingly, it can be checked in polynomial time if M' is a *summand* of M [6, Theorem 3.5].

7 A Distance Induced by a Noise System

As a last application of our methods, we show that a particular distance induced by a *noise system* is NP-hard to approximate within a factor of 2.

A noise system, as introduced by Scolamiero et al. [19], induces a pseudometric on (tame) persistence modules. In this section, we shall briefly consider one particular noise system and we refer the reader to [19] for an in-depth treatment of the more general theory.

We say that $f: M \to N$ is a μ-*equivalence* if f has μ_1-trivial kernel and μ_2-trivial cokernel, and $\mu_1 + \mu_2 \leq \mu$. From this definition, we can define the following distance between two persistence modules M and N

$$d_{\text{noise}}(M, N)$$
$$= \inf\{\mu \mid \exists M \xleftarrow{f} X \xrightarrow{g} N,\ f \text{ an } \epsilon\text{-equivalence, } g \text{ a } \delta\text{-equivalence and } \epsilon + \delta \leq \mu\}$$

The reader may verify that this distance coincides with the distance induced by the noise system $\{S_\epsilon\}$ where S_ϵ consists of all persistence modules M with the property that $M_{p \to p+(\epsilon,\epsilon)}$ is trivial for all p. In particular, d_{noise} is indeed an extended pseudometric [19, Proposition 8.7].

Like for the interleaving distance, we can define the computational problem of c-approximating d for a constant $c \geq 1$.

c-APPROX-d_{noise}: Given two persistence modules M, N in graded matrix representation, return a real number r such that

$$d_{\text{noise}}(M, N) \leq r \leq c \cdot d_{\text{noise}}(M, N)$$

Theorem 7 *c-APPROX-d is NP-hard for $c < 2$.*

Proof Let (n, P, Q) be a CI-instance and construct M and N as in Theorem 2. We will show the following implications:

$$d_I(M, N) = 1 \Rightarrow d_{\text{noise}}(M, N^1) \leq 2$$
$$d_I(M, N) = 3 \Rightarrow d_{\text{noise}}(M, N^1) \geq 4.$$

This allows us to conclude that an algorithm c-approximating $d_{\text{noise}}(M, N)$ for $c < 2$ will return a number < 4 if $d_I(M, N) = 1$ and a number ≥ 4 if $d_I(M, N) = 3$. This constitutes a polynomial time reduction from CI to 2-APPROX-d_{noise} and the result follows from Theorem 1.

First assume that $d_I(M, N) = 1$. Let $X = M$ with $f : X \to M$ the identity morphism. Lemma 11 shows that the interleaving morphism $g: M \to N^1$ has 2-trivial cokernel, and from Lemma 13 we know that it is injective. Hence g is a 2-equivalence and thus $d_{\text{noise}}(M, N^1) \leq 2$.

Now assume that $d_{\text{noise}}(M, N^1) < 4$. By definition this gives a diagram $M \xleftarrow{f} X \xrightarrow{g} N^1$ where f is an ϵ-equivalence, g is a δ-equivalence, and $\epsilon + \delta < 4$. We may assume that both f and g are injective. To see this, consider $x \in \ker(f_p)$. Because f is an ϵ-equivalence, $\ker(f)$ is ϵ-trivial, so $X_{p \to p+(\epsilon,\epsilon)}(x) = 0$. This gives

$$0 = g_{p+(\epsilon,\epsilon)} \circ X_{p \to p+(\epsilon,\epsilon)}(x) = N^1_{p \to p+(\epsilon,\epsilon)} \circ g_p(x).$$

Since $N^1_{p \to p+(\epsilon,\epsilon)}$ is injective, we conclude that $g_p(x) = 0$. This shows that $\ker(f) \subseteq \ker(g)$, and by symmetry, that $\ker(f) = \ker(g)$. Replacing X with $\tilde{X} = X/\ker(f)$

induces injective morphisms $M \xleftarrow{\tilde{f}} \tilde{X} \xrightarrow{\tilde{g}} N^1$ with the properties that $\tilde{f}$ is an ϵ-equivalence and that $\tilde{g}$ is a δ-equivalence. Hence f and g may be assumed to be injective. Under this assumption, we get the following two inequalities from Lemma 12

$$d_I(M, X^{\epsilon/2}) \le \epsilon/2$$
$$d_I(N^1, X^{\delta/2}) \le \delta/2.$$

Observe that $d_I(N^1, X^{\delta/2}) = d_I(N^{1-\delta/2}, X) = d_I(N^{1+(\epsilon-\delta)/2}, X^{\epsilon/2})$. Together with the first inequality, this gives $d_I(M, N^{1+(\epsilon-\delta)/2}) \le (\epsilon+\delta)/2$, and thus

$$\begin{aligned} d_I(M, N) &\le d_I(M, N^{1+(\epsilon-\delta)/2}) + d_I(N^{1+(\epsilon-\delta)/2}, N) \\ &\le (\epsilon+\delta)/2 + (1 + (\epsilon-\delta)/2) = 1 + \epsilon. \end{aligned}$$

To conclude the proof, we will show that $\delta \ge 2$, as this implies $1+\epsilon < 1+4-\delta \le 3$. Assuming that $n \ge 1$, let p be such that $\dim N_p > 0$ and $\dim M_{p+(r,r)} = 0$ for all $r < 1$. Such a point exists for the following reason: let M_i be any indecomposable summand of M and let N_j be any indecomposable summand of N. Then M_i is a staircase module for which the underlying staircase is obtained by moving certain corners of the staircase S in Fig. 2. Likewise, the staircase supporting N_j is obtained by moving certain corners of T. However, by construction, and as shown in Fig. 3, a number of corners are left unmoved. Hence, we may simply choose p to be any corner point of T with negative 1st coordinate which is left unmoved in the construction of N_j.

Let $q = p - (1, 1)$. Then $\dim N^1_q > 0$ and $\dim M_{q+(r,r)} = 0$ for all $r < 2$. But since f is an injection, the space $X_{q+(r,r)}$ must also be trivial for any $r < 2$. It follows that $\delta \ge 2$. □

8 Conclusion

Using the link between persistence modules indexed over $\mathbb{R}^2$ and CI problems introduced in [4], we settle the computational complexity of a series of problems. Most notably, we show that computing the interleaving distance is NP-hard, as is approximating it to any constant factor < 3. Moreover, we investigated the problem of deciding one-sided stability. Except for checking isomorphism, which is known to be polynomial, we show that all non-trivial cases are NP-hard. This includes checking whether a module is a submodule of another. Our assumption that we are working over a finite field stays in the background for most of the paper, but we rely heavily on this assumption for proving the submodule problem. Lastly, we showed that approximating a distance d arising from a noise system up to a constant less than 2 is also NP-hard.

Throughout, we use persistence modules decomposing into very simple modules called staircase modules. These have the big advantage that the morphisms between them have very simple descriptions in terms of matrices. While this simplification might appear to throw the complexity out with the bathwater, our results clearly show that this is not the case.

The question of whether c-APPROX- INTERLEAVING- DISTANCE is NP-hard for $c \geq 3$ is still open, and it is not clear whether one can prove this with CI problems or not. Even if this should not be possible, we believe that a better understanding of CI problems would lead to a better understanding of persistence modules and interleavings.

Acknowledgements We thank the anonymous referees for valuable suggestions, including the connection to noise systems discussed in Sect. 7. Magnus Bakke Botnan has been partially supported by the DFG Collaborative Research Center SFB/TR 109 "Discretization in Geometry and Dynamics". Michael Kerber is supported by Austrian Science Fund (FWF) Grant Number P 29984-N35.

References

1. Bauer, U., Lesnick, M.: Induced matchings of barcodes and the algebraic stability of persistence. In: Proceedings of the thirtieth annual symposium on Computational geometry, p. 355. ACM (2014)
2. Bendich, P., Marron, J.S., Miller, E., Pieloch, A., Skwerer, S.: Persistent homology analysis of brain artery trees. The annals of applied statistics **10 1**, 198–218 (2016)
3. Bjerkevik, H.B.: Stability of higher-dimensional interval decomposable persistence modules. arXiv preprint arXiv:1609.02086 (2016)
4. Bjerkevik, H.B., Botnan, M.B.: Computational complexity of the interleaving distance. In: 34th International Symposium on Computational Geometry, SoCG 2018, pp. 13:1–13:15 (2018)
5. Botnan, M.B., Lesnick, M.: Algebraic stability of zigzag persistence modules. Algebraic & Geometric Topology **18**, 3133–3204 (2018)
6. Brooksbank, P.A., Luks, E.M.: Testing isomorphism of modules. Journal of Algebra **320**(11), 4020–4029 (2008)
7. Buchet, M., Escolar, E.G.: Realizations of indecomposable persistence modules of arbitrarily large dimension. In: 34th International Symposium on Computational Geometry, SoCG 2018, pp. 15:1–15:13 (2018)
8. Chazal, F., De Silva, V., Glisse, M., Oudot, S.: The structure and stability of persistence modules. Springer, New York (2016)
9. Chazal, F., Guibas, L.J., Oudot, S.Y., Skraba, P.: Persistence-based clustering in riemannian manifolds. J. ACM **60**(6), 41:1–41:38 (2013)
10. Dey, T.K., Xin, C.: Computing bottleneck distance for 2-d interval decomposable modules. In: 34th International Symposium on Computational Geometry, SoCG 2018, pp. 32:1–32:15 (2018)
11. Edelsbrunner, H., Harer, J.: Computational Topology: An Introduction. American Mathematical Society, Providence, RI, USA (2010)
12. Ivanyos, G., Karpinski, M., Saxena, N.: Deterministic polynomial time algorithms for matrix completion problems. SIAM journal on computing **39**(8), 3736–3751 (2010)
13. Kerber, M., Nigmetov, A.: Geometry helps to compare persistence diagrams. ACM Journal of Experimental Algorithmics **22** (2017)
14. Lesnick, M.: The theory of the interleaving distance on multidimensional persistence modules. Foundations of Computational Mathematics **15**(3), 613–650 (2015)
15. Lesnick, M., Wright, M.: Interactive visualization of 2-d persistence modules. arXiv preprint arXiv:1512.00180 (2015)
16. Pranav, P., Edelsbrunner, H., van de Weygaert, R., Vegter, G., Kerber, M., Jones, B., Wintraecken, M.: The topology of the cosmic web in terms of persistent betti numbers. Monthly Notices of the Royal Astronomical Society **465**(4), 4281–4310 (2016)

17. Reininghaus, J., Huber, S., Bauer, U., Kwitt, R.: A stable multi-scale kernel for topological machine learning. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7–12, 2015, pp. 4741–4748 (2015)
18. Rybakken, E., Baas, N., Dunn, B.: Decoding of neural data using cohomological feature extraction. Neural computation **31**(1), 68–93 (2019)
19. Scolamiero, M., Chachólski, W., Lundman, A., Ramanujam, R., Öberg, S.: Multidimensional persistence and noise. Foundations of Computational Mathematics **17**(6), 1367–1406 (2017)

Appendix

A conjecture

Håvard Bakke Bjerkevik

A conjecture

The idea of representing interleavings as pairs of matrices is a theme running through all the papers. In the first, it is applied in the setting of stability, and in the other two, the setting is complexity. In Section 6 of Paper I and in the end of Section 4 of Paper III, we allude to a connection between stability of staircase decomposable modules and the complexity of approximating d_I. The purpose of this note is to elaborate on this connection.

The connection takes the form of a conjecture. It ties the papers in the thesis neatly together, as a corollary of the conjecture is a statement of stability that would strengthen the main theorem in Paper I to the point where it cannot be improved, while it also puts the finger on the problem of improving the constant in the main theorem in Paper III. Still, I must admit that perhaps my main motivation for including this note is simply that I think it is a really neat problem regardless of applications. Like many good problems, it is easy to state, but, at least in my experience, surprisingly hard to solve.

1 The conjecture and its connection to interleavings

Let $\mathcal{P} = (n, P, Q)$ be a CI problem and $G_{\mathcal{P}}$ the directed graph on $\{a_1, \dots, a_n, b_1, \dots, b_n\}$ with an edge from a_i to b_j if $P_{i,j} \neq 0$ and an edge from b_i to a_j if $Q_{i,j} \neq 0$. Let G be the undirected graph on the same set of vertices with an edge between a_i and b_j if they are contained in the same directed cycle of length at most 4.

Conjecture 1. *If (n, P, Q) has a solution, then G has a matching.*

The following lemma and corollary explains the link to interleavings. We assume that M and N are staircase decomposable modules.

Lemma 1. *If (n, P, Q) is associated to the pair (M, N) and G has a matching, then there is a 3ϵ-matching between $B(M)$ and $B(N)$.*

Proof. Suppose G has a matching, and a_i is matched with b_j. Then there is an edge $a_i \to b_j$ and a path $b_j \to a_{i'} \to b_{j'} \to a_i$, or a path $a_i \to b_{j'} \to a_{i'} \to b_j$ and an edge $b_j \to a_i$ in $G_{\mathcal{P}}$. In the first case, there is a nonzero morphism $M_i \to N_j(\epsilon)$ and nonzero morphisms $N_j \to M_{i'}(\epsilon) \to N_{j'}(2\epsilon) \to M_i(3\epsilon)$ whose composition is also nonzero. Thus, the matching in G describes a 3ϵ-matching between $B(M)$ and $B(N)$. □

The following is a corollary of the conjecture and lemma above.

Corollary 2. *If M and N are ϵ-interleaved, there is a 3ϵ-matching between $B(M)$ and $B(N)$.*

This is a statement about stability: It says that for staircase decomposable modules M, N,

$$d_B(B(M), B(N)) \leq 3d_I(M, N)$$

holds. Presumably, this implies the same inequality for rectangle decomposable modules, though the presence of small rectangles makes this case slightly different.

Corollary 3. *The interleaving distance between staircase decomposable modules can be 3-approximated in polynomial time.*

Proof. One can compute $d_B(B(M), B(N))$ in polynomial time, and Corollary 2 shows that this is a 3-approximation of the interleaving distance, as $d_B(B(M), B(N)) \in [d_I(M, N), 3d_I(M, N)]$. □

The next natural question would then be what the situation looks like for general multi-parameter modules. Could it be possible to 3-approximate the interleaving distance between them in polynomial time? If, on the other hand, one can find counterexamples to the conjecture, that could open the door to strengthening the results in Paper III.

2 Discussion

One way of viewing the conjecture is the following. Assume you have a bipartite directed graph on $n+n$ nodes. On one hand, you can view the two sets of nodes as close if they allow a bijection pairing up nodes that are contained in the same short cycle. On the other hand, you can view them as close if there is an invertible $n \times n$ matrix A such that A and A^{-1} are nonzero only in the places given by the graph. Are these notions of closeness roughly the same?

Now, matchings and invertable matrices are simple enough concepts, but here combinatorics and algebra interact in a way that seems to cause both purely combinatorial and purely algebraic arguments to fail. Whether or not this problem is relevant for TDA, I find this phenomenon very interesting on a purely mathematical basis. If the conjecture is true, it would in my opinion be a fascinating theorem.

The parallel to persistence is of course that the first notion of closeness corresponds to matchings and d_B, while the second corresponds to interleavings through CI problems. The counterexample in Paper I where $d_B > d_I$ shows that the equality in the algebraic stability theorem is an accident of the geometry of $\mathbb{R}$ rather than a universal statement about the relation between d_B and d_I. The conjecture asks if a weaker relation still exists when we remove all geometric restrictions, and the only assumption is that the modules decompose into summands such that the morphisms between them are nice enough.

In Paper II, we observed that solvability of CI problems sometimes depend on the characteristic of the field we are working with. Thus, it is not clear whether the truth of the conjecture depends on the field or not.