



Distributing Persistent Homology via Spectral Sequences

Álvaro Torras-Casas¹ 

Received: 11 August 2020 / Revised: 27 October 2022 / Accepted: 12 November 2022 /
Published online: 20 September 2023
© The Author(s) 2023

Abstract

We set up the theory for a distributed algorithm for computing persistent homology. For this purpose we develop linear algebra of persistence modules. We present bases of persistence modules, together with an operation \boxplus that leads to a method for obtaining images, kernels and cokernels of tame persistence morphisms. Our focus is on developing efficient methods for the computation of homology of chains of persistence modules. Later we give a brief, self-contained presentation of the Mayer–Vietoris spectral sequence. Then we study the Persistent Mayer–Vietoris spectral sequence and present a solution to the extension problem. This solution is given by finding coefficients that indicate gluings between bars on the same dimension. Finally, we review PERMAVISS, an algorithm that computes all pages in the spectral sequence and solves the extension problem. This procedure distributes computations on sub-complexes, while focusing on merging homological information. Additionally, some computational bounds are found which confirm the distribution of the method.

Keywords Spectral sequences · Distributed persistent homology · Mayer–Vietoris

Mathematics Subject Classification 55-04 · 55N35 · 55T99

1 Introduction

1.1 Motivation

Persistent homology has existed for about two decades [15]. This tool introduced the field of Topological Data Analysis which, very soon, was applied to a multitude of

Editor in Charge: Kenneth Clarkson

Álvaro Torras-Casas
TorrasCasasA@cardiff.ac.uk

¹ School of Mathematics, Cardiff University, Abacws, Senghennydd Road, Cardiff CF24 4AG, Wales, UK

problems, see [5, 16]. Among others, persistent homology has been applied to study the geometric structure of sets of points lying in \mathbb{R}^n [12, 15], coverage in sensor networks [25], pattern detection [23], classification and recovery of signals [24] and it has also had an impact on shape recognition using machine learning techniques, see [1, 13]. All these applications motivate the need for fast algorithms for computing persistent homology. The usual algorithm used for these computations was introduced in [15], with some later additions to speed up such as those of [7, 8, 26]. In [21] persistent homology is proven to be computable in matrix multiplication time. However, since these matrices become large very quickly, the computations are generally very expensive, both in terms of computational time and in memory required.

In practice, computing the persistent homology of a given filtered complex is equivalent to computing its matrices of differentials and perform successive Gaussian eliminations; see [14, 15]. In recent years, some methods have been developed for the parallelization of persistent homology. The first approach was introduced in [14] as the *spectral sequence algorithm*, and was successfully implemented in [2]. This consists in dividing the original matrix M into groups of rows, and sending these to different processors. These processors will, in turn, perform a local Gaussian Elimination and share the necessary information between them, see [2]. On the other hand, a more topological approach is presented in [18]. It uses the *blow-up complex* introduced in [33]. This approach first takes a cover \mathcal{C} of a filtered simplicial complex K , and uses the result that the persistent homology of K is isomorphic to that of the blow-up complex $K^{\mathcal{C}}$. This proceeds by computing the sparsified persistent homology for each cover, and then using this information to reduce the differential of $K^{\mathcal{C}}$ efficiently. Both of these parallelization methods have provided substantial speedups compared to the standard method presented in [15].

The relation between homology classes and a cover is known as *Localized Homology*, which was introduced in [33]. It would be useful to have a method that leads to the speedups from [2, 18], while keeping Localized Homology information. Further, such covers should have no restrictions, such as those used in the *mapper algorithm*, see [27]. This last point limits substantially the use of the blowup-complex, since the number of simplices increases very quickly when we allow the intersections to grow. In fact, in the extreme case where a complex K is covered by n copies of K , the blowup complex $K^{\mathcal{C}}$ has size $2^n|K|$.

1.2 Literature Review

Since distribution is an important issue in persistent homology, it is worth exploring which classical tools of algebraic topology could be used in this context. A well-known tool for distributing homology computations is the Mayer–Vietoris spectral sequence [3, Sect. 8], see [9] for a quick introduction to spectral sequences and [20] as a general reference. Since the category of persistence modules and persistence morphisms is an abelian category, the process of computing a spectral sequence should be more or less straightforward. However, implementing this in practice is a difficult task. Furthermore, this approach has been proposed in [19], although without a solution to the extension problem. Later, spectral sequences were used for distributing computations

of cohomology groups in [11], and recently in [31] and [32] spectral sequences are used for distributing persistent homology computations. However, all of [11, 31, 32] assume that the nerve of the cover is one-dimensional.

The first problem when dealing with spectral sequences is that we need to be able to compute images, kernels and quotients as well as their representatives. This question has already been studied in [10], where the authors give a very efficient algorithm. However, there are couple of problems that come up when using [10] in spectral sequences. First, in [10] the authors consider a persistence morphism induced by a simplicial morphism $f: X \rightarrow Y$. However, when working with the Mayer–Vietoris spectral sequence we consider maps in the second, third and higher pages which are not induced by a simplicial morphism at all. Furthermore, even when working with the first page differentials, we cannot adapt the work from [10]; as a simplex from an intersection is sent to several copies along lower degree intersections. Second, a key assumption in [10] is that the filtrations in X and Y are both *general* in the sense that a simplex in either X or Y is born at a time. However, in spectral sequences generality *hardly ever* holds. Indeed, this follows from the fact that a simplex might be contained in various overlapping covers.

Thus, if we want to compute images, kernels and cokernels, we will need to be able to overcome these two difficulties first. Also, notice that a good solution should lead to the representatives, as these are needed for the spectral sequence. It is worth to mention that in [28] such images, kernels and cokernels were studied in terms of Smith Normal Forms of presentations associated to persistence modules. Further, as mentioned in [28], such work was developed with the aim of computing persistence spectral sequences [19] in mind. However, adapting [28] to an algorithm which computes spectral sequences is still a challenge.

The other difficulty with spectral sequences is the extension problem, which we explain in Sect. 5.1. Within the context of persistent homology, the extension problem first appeared in [17, Sect. 6]. There the authors give an approximate result that holds in the case of acyclic coverings. This allows them to compare the persistent homology to the lower row of the infinity page in the spectral sequence. This leads to an ε -interleaving between the global persistent homology and that of the filtered nerve. Later, the extension problem appeared in the PhD thesis of Yoon [31], and also in the recent joint work with Ghrist [32]. In Sect. 4.2.3 from Yoon's Thesis, there is a detailed solution for the extension problem in the case when the nerve of the cover is one-dimensional.

1.3 Original Contribution

In this paper, we set the theoretical foundations for a distributed method on the input data. In order to do this, we use the algebraic power of the Mayer–Vietoris spectral sequence. Since the aim is to build up an explicit algorithm, we need to develop linear algebra of persistence modules, as done through Sect. 3. There, we define *barcode bases* and also we develop an operation \boxplus that allows to determine whether a set of *barcode vectors* is linearly independent or not. Using this machinery, we are able to encapsulate all the information related to a persistence morphism in a matrix

that depends on the choice of two barcode bases. This allows defining a Gaussian elimination outlined in the `box_gauss_reduce` algorithm—see Algorithm 1—which addresses the two issues raised above with regards to [10].

Next in Sect. 4, we give a detailed review of the Mayer–Vietoris spectral sequence in the homology case. This is followed by Sect. 5.1, where we give a solution to the extension problem. The solution is given by a careful consideration of the total complex homology, together with the use of barcode basis machinery developed in Sect. 3. In Sect. 5.2 we introduce PERMAVISS, an algorithm for computing the persistence Mayer–Vietoris spectral sequence and solving the extension problem. The advantage of this procedure is that all the simplicial information is enclosed within local matrices. This has one powerful consequence; this method consists in computing local Gaussian eliminations plus computing `image_kernel` on matrices whose order is that of homology classes. In particular, given enough processors and a ‘good’ cover of our data, one has that the complexity is about

$$\mathcal{O}(X^3) + \mathcal{O}(H^3),$$

where X is the order of the maximal local complex and H is the overall number of nontrivial persistence bars on the whole dataset.¹ For more details on this, we refer the reader to Sect. 5.4.

By using the ideas in this text we developed PERMAVISS, a Python library that computes the Persistence Mayer–Vietoris spectral sequence. In the results from [29], one can see that nontrivial higher differentials come up and also the extension problem cannot be solved in a trivial way in general. This supports the idea that the spectral sequence adds more information on top of persistent homology. Finally, we outline future directions, both for the study of the Persistence Mayer–Vietoris spectral sequence and future versions of PERMAVISS.

2 Preliminaries

2.1 Simplicial Complexes

Definition 2.1 Given a set X , a *simplicial complex* K is a subset of the power set $K \subseteq P(X)$ such that if $\sigma \in K$, then for all subsets $\tau \subseteq \sigma$ we have that $\tau \in K$. An element $\sigma \in K$ will be called an *n-simplex* whenever $|\sigma| = n + 1$, whereas a subset $\tau \subseteq \sigma$ will be called a *face*. Thus, if a simplex is contained in K all its faces must also be contained in K . Given a simplicial complex K , we denote by K_n the set containing all the n -simplices from K . Given a pair of simplicial complexes K and L , if $L \subseteq K$, then we say that L is a *subcomplex* of K . Also, given a mapping $f : K \rightarrow L$ between two simplicial complexes K and L , we call f a *simplicial morphism* whenever

¹ Notice that in a filtration indexed by integers where we introduce a simplex at a time there are no trivial bars and the number of these is about $(\# \text{simplices})/2$. Here we are referring to the case where the filtration is over the real numbers; where many simplices are introduced at the same time. In this case the number of nontrivial persistent homology bars should be much smaller than the number of simplices.

$f(K_n) \subseteq \bigcup_{l=0}^n L_l$ for all $n \geq 0$. The category composed of simplicial complexes and simplicial morphisms will be denoted by **SpCpx**.

We represent the simplices from K as equivalence classes of tuples from X which are equal up to even permutations, see [22, Sect. 5]. Let \mathbb{F} be a field. For each $n \geq 0$ we define the free vector space $S_n(K) := \mathbb{F}[K_n]$. We also consider linear maps $d_n: S_n(K) \rightarrow S_{n-1}(K)$ usually called *differentials*, defined by

$$d_n([v_0, \dots, v_n]) = \sum_{i=0}^n (-1)^i [v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_n]. \tag{1}$$

Setting $S_n(K) = 0$ for all $n < 0$, we put all of these in a sequence

$$0 \xleftarrow{0} S_0(K) \xleftarrow{d_1} S_1(K) \xleftarrow{d_2} S_2(K) \xleftarrow{d_3} \dots \tag{2}$$

It follows from formula (1) that the composition of two consecutive differentials vanishes: $d_n \circ d_{n-1} = 0$ for all $n \geq 0$. In this case we say that (2) is a *chain complex*. As a consequence, we have that $\text{Im}(d_{n+1}) \subseteq \text{Ker}(d_n)$, and we can define the *homology* with coefficients in \mathbb{F} to be $H_n(K; \mathbb{F}) = \text{Ker}(d_n) / \text{Im}(d_{n+1})$ for all $n \geq 0$. In general, \mathbb{F} will be understood by the context and the notation $H_n(K)$ might be used instead. On the other hand, we consider the *augmentation map* $\varepsilon: S_0(K) \rightarrow \mathbb{F}$ defined by the assignment $s \mapsto 1_{\mathbb{F}}$, for any simplex $s \in S_0(K)$. Then, we define the *reduced homology* by $\tilde{H}_0(K; \mathbb{F}) = \text{Ker}(\varepsilon) / \text{Im}(d_1)$ and $\tilde{H}_n(K; \mathbb{F}) = H_n(K; \mathbb{F})$ for all $n > 0$. Consider $\tilde{S}_*(K)$, obtained by augmenting (2) by ε and a copy of \mathbb{F} in degree -1 :

$$0 \xleftarrow{0} \mathbb{F} \xleftarrow{\varepsilon} S_0(K) \xleftarrow{d_1} S_1(K) \xleftarrow{d_2} S_2(K) \xleftarrow{d_3} \dots$$

Then, computing homology on $\tilde{S}_*(K)$ we obtain the reduced homology groups.

Definition 2.2 (*standard m -simplex*) Given $m > 0$, we define $\Delta^m = P(\{0, 1, \dots, m\})$, which will be called the *standard m -simplex*. This leads to a chain complex $\tilde{S}_*(\Delta^m)$

$$\begin{aligned} 0 \xleftarrow{0} \mathbb{F} \xleftarrow{\varepsilon} S_0(\Delta^m) \xleftarrow{d_1} S_1(\Delta^m) \\ \xleftarrow{d_2} S_2(\Delta^m) \xleftarrow{d_3} \dots \xleftarrow{d_n} S_n(\Delta^m) \xleftarrow{\quad} 0. \end{aligned}$$

By [22, Thm. 8.3], $\tilde{S}_*(\Delta^m)$ is *exact*, that is, $\tilde{H}_n(\Delta^m) = 0$ for all $n \geq 0$.

Definition 2.3 Let K be a simplicial complex. A finite set $\mathcal{U} = \{U_i\}_{i=0}^m$ of subcomplexes from K , is a *cover* of K whenever $K = \bigcup_{i=0}^m U_i$. Let $\sigma \in \Delta^m$ and denote $U_\sigma = \bigcap_{i \in \sigma} U_i$. The *nerve* of \mathcal{U} is defined by

$$N^{\mathcal{U}} = \{\sigma : U_\sigma \neq \emptyset\} \subseteq \Delta^m.$$

In particular, given a simplex $\sigma \in N^{\mathcal{U}}$, there is an injection $f^\sigma : \Delta^{|\sigma|} \hookrightarrow N^{\mathcal{U}}$ which induces an injection of chain complexes $f_*^\sigma : \tilde{\mathcal{S}}_*(\Delta^{|\sigma|}) \hookrightarrow \tilde{\mathcal{S}}_*(N^{\mathcal{U}})$ whose image $f_*^\sigma(\tilde{\mathcal{S}}_*(\Delta^{|\sigma|}))$ is exact.

Definition 2.4 (*Čech chain complex*) Let K be a simplicial complex and let $\mathcal{U} = \{U_i\}_{i=0}^m$ be a cover of K . Given $s \in K$, there exists $\sigma(s) \in N^{\mathcal{U}}$ with maximal cardinality $|\sigma(s)|$, so that $s \in U_{\sigma(s)}$. Then, we define the (n, \mathcal{U}) -Čech chain complex by

$$\check{C}_*(n, \mathcal{U}; \mathbb{F}) = \bigoplus_{s \in K_n} f_*^{\sigma(s)}(\tilde{\mathcal{S}}_*(\Delta^{|\sigma(s)|})).$$

For $k \geq -1$, we use the notation $(\tau)_s$ with $s \in K_n$ and $\tau \in \tilde{\mathcal{S}}_k(\Delta^{|\sigma(s)|})$, to denote an element in $\check{C}_k(n, \mathcal{U}; \mathbb{F})$ that is zero everywhere except for τ in the component indexed by s . Then the image of the k -Čech differential is defined by the assignment $\delta_k^{\mathcal{U}}((\tau)_s) = (d_k^{N^{\mathcal{U}}} \tau)_s$, where $d_k^{N^{\mathcal{U}}}$ denotes the k th differential of $\tilde{\mathcal{S}}_*(N^{\mathcal{U}})$. By definition, the Čech complex is a chain complex and is exact. Also, $\check{C}_{-1}(n, \mathcal{U}; \mathbb{F}) \simeq S_n(K)$ follows easily. On the other hand, for each $k \geq 0$ we define an isomorphism

$$\psi_k : \check{C}_k(n, \mathcal{U}; \mathbb{F}) \simeq \bigoplus_{\sigma \in N_k^{\mathcal{U}}} S_n(U_\sigma)$$

by sending $(\tau)_s$ to $(s)_\tau$ for any pair of simplices $s \in K_n$ and $\tau \in f_k^{\sigma(s)} \Delta^{|\sigma(s)|}$. In particular, we can rewrite the (n, \mathcal{U}) -Čech chain complex as a sequence

$$0 \longleftarrow S_n(K) \xleftarrow{\delta_0} \bigoplus_{\sigma \in \Delta_0^m} S_n(U_\sigma) \xleftarrow{\delta_1} \bigoplus_{\sigma \in \Delta_1^m} S_n(U_\sigma) \xleftarrow{\delta_2} \bigoplus_{\sigma \in \Delta_2^m} S_n(U_\sigma) \longleftarrow \dots \quad (3)$$

where, for any pair $\sigma \in N_k^{\mathcal{U}}$ and $s \in (U_\sigma)_n$, the differentials δ_i are defined as follows:

$$\delta_k((s)_\sigma) = \psi_k \circ \check{\delta}_k \circ \psi_k^{-1}((s)_\sigma) = \psi_k((d_k^{N^{\mathcal{U}}} \sigma)_s) = (\{d_k^{N^{\mathcal{U}}}(\sigma)\}_\tau \cdot s)_{\tau \in N_{k-1}^{\mathcal{U}}},$$

and where $\{d_k^{N^{\mathcal{U}}}(\sigma)\}_\tau \in \mathbb{F}$ is the coefficient of $d_k^{N^{\mathcal{U}}}(\sigma)$ in the simplex $\tau \in N_{k-1}^{\mathcal{U}}$.

Remark 2.5 Alternatively, the Čech chain complex can be defined straight away as the sequence (3), and prove exactness by using cosheaf theory. Namely, given a simplicial complex K , we consider the topology where the open sets are subcomplexes. Then, for each integer $n \geq 0$, consider the simplicial *precosheaf* as an assignment $\mathcal{S}_n : V \mapsto S_n(V)$ for each subcomplex $V \subseteq K$. This precosheaf is in fact a *flabby cosheaf*. Then, using 2.5, 4.3, and 4.4 from [4, Sect. VI], one has exactness of the Čech chain complex.

2.2 Persistence Modules

Let \mathbf{R} be the category of real numbers as a poset, where $\text{hom}_{\mathbf{R}}(s, t)$ contains a single morphism whenever $s \leq t$, and is empty otherwise. Let \mathbb{F} be a field and let \mathbf{Vect} denote the category of \mathbb{F} -vector spaces.

Definition 2.6 A *filtered simplicial complex* is a functor $K : \mathbf{R} \rightarrow \mathbf{SpCpx}$, with $K_s \subseteq K_t$ for all $s \leq t$. Define $\text{PH}_n(K)$, the *n-persistent homology* of K as the composition $H_n(K) : \mathbf{R} \rightarrow \mathbf{Vect}$.

Definition 2.7 A *persistence module* \mathbb{V} is a functor $\mathbb{V} : \mathbf{R} \rightarrow \mathbf{Vect}$. That is, to any $r \in \mathbf{R}$, \mathbb{V} assigns a vector space in \mathbf{Vect} which is denoted either by $\mathbb{V}(r)$ or \mathbb{V}^r . Additionally, to any pair $s \leq t$, there is a linear morphism $\mathbb{V}(s \leq t) : \mathbb{V}^s \rightarrow \mathbb{V}^t$. These morphisms satisfy $\mathbb{V}(s \leq s) = \text{Id}_{\mathbb{V}^s}$ for any $s \in \mathbf{R}$ and the relation $\mathbb{V}(r \leq t) = \mathbb{V}(s \leq t) \circ \mathbb{V}(r \leq s)$ for all $r \leq s \leq t$ in \mathbf{R} . Given two persistence modules \mathbb{V} and \mathbb{W} , a *persistence morphism* is a natural transformation $f : \mathbb{V} \rightarrow \mathbb{W}$. Thus, for any pair $s \leq t$, there is a commutative square

$$\begin{array}{ccc}
 \mathbb{V}^s & \xrightarrow{\mathbb{V}(s \leq t)} & \mathbb{V}^t \\
 f^s \downarrow & & \downarrow f^t \\
 \mathbb{W}^s & \xrightarrow{\mathbb{W}(s \leq t)} & \mathbb{W}^t.
 \end{array}$$

We denote by \mathbf{PMod} the category of persistence modules and persistence morphisms.

By *naturality* of f we refer to the commutative square above. A persistence morphism $f : \mathbb{V} \rightarrow \mathbb{W}$ is an *isomorphism* whenever f_t is an isomorphism for all $t \in \mathbf{R}$, which we denote by $\mathbb{V} \simeq \mathbb{W}$.

Example 2.8 Let $s \leq t$ from \mathbf{R} . We define the *interval module* $\mathbb{F}_{[s,t]}$ by $\mathbb{F}_{[s,t]}(r) = \mathbb{F}$ for all $r \in [s, t)$ and $\mathbb{F}_{[s,t]}(r) = 0$ otherwise. The morphisms $\mathbb{F}_{[s,t]}(a \leq b)$ are the identity for any two $a, b \in [s, t)$ and 0 otherwise.

Given $\mathbb{F}_{[s,t]}$, s and t are the *birth* and *death* values respectively. If $\mathbb{V}(r)$ is finite dimensional for all $r \in \mathbf{R}$, then there is an isomorphism $\mathbb{V} \simeq \bigoplus_{i \in J} \mathbb{F}_{[s_i, t_i]}$, as shown in [6]. This is the *barcode decomposition* of \mathbb{V} .

Definition 2.9 $\mathbb{V} \in \mathbf{PMod}$ is *tame* iff it has a finite barcode decomposition: $\mathbb{V} \simeq \bigoplus_{i=1}^N \mathbb{F}_{[a_i, b_i]}$.

Throughout this work all persistence modules are assumed to be tame.

Definition 2.10 A *chain of persistence modules* is a sequence of persistence modules \mathbb{V}_k and differentials $\delta_k : \mathbb{V}_k \rightarrow \mathbb{V}_{k-1}$ such that $\delta_{k-1} \circ \delta_k = 0$ for all $k \in \mathbb{Z}$. Elements from $\text{PVect}(\mathbb{V}_k)$, for all $k \in \mathbb{Z}$, are sometimes called *chains*.

3 Homology of Persistence Modules

3.1 Barcode Bases

In this section, we introduce barcode bases and the operation \boxplus . This framework allows to introduce the matrix associated to a persistence morphism $f : \mathbb{V} \rightarrow \mathbb{W}$ and a choice of bases. This theory is applied to develop algorithms for computing images and kernels of persistence morphisms as well as quotients of persistence modules. In addition, we evaluate the respective computational complexities. Additionally, we illustrate how to compute homology in the category of persistence modules.

Definition 3.1 (*barcode basis*) A barcode basis \mathcal{A} of a tame persistence module \mathbb{V} is a choice of an isomorphism, $\alpha : \bigoplus_{i=1}^N \mathbb{F}_{[a_i, b_i]} \rightarrow \mathbb{V}$. A direct summand of α is a restricted morphism $\alpha_i : \mathbb{F}_{[a_i, b_i]} \rightarrow \mathbb{V}$ which we call a *barcode generator*. Often, we denote a barcode basis \mathcal{A} by the set of barcode generators $\mathcal{A} = \{\alpha_i\}_{i=1}^N$.

Within the context of Definition 3.1, we make some notational remarks.

- Given a barcode generator $\alpha_i \in \mathcal{A}$, we write $\alpha_i \sim [a_i, b_i)$ to indicate that the domain of α_i is $\mathbb{F}_{[a_i, b_i)}$ and say that α_i is *associated* to the interval $[a_i, b_i)$.
- Given $\alpha_i \in \mathcal{A}$ with $\alpha_i \sim [a_i, b_i)$, we have linear transformations $\alpha_i(r) : \mathbb{F}_{[a_i, b_i)}(r) \rightarrow \mathbb{V}(r)$ for all $r \in \mathbf{R}$. In particular, since $\mathbb{F}_{[a_i, b_i)}(r)$ is either 0 or \mathbb{F} , $\alpha_i(r)$ is uniquely determined by $\alpha_i(r)(1_{\mathbb{F}}) \in \mathbb{V}(r)$ for $r \in [a_i, b_i)$. In addition, notice that $\alpha_i(r) \neq 0$ for all $r \in [a_i, b_i)$ since otherwise α would not be injective.
- For any given $r \in \mathbf{R}$, we define a subset of \mathcal{A} ,

$$\mathcal{A}^r = \{\alpha_i : 1 \leq i \leq N, \alpha_i(r) \neq 0\}.$$

In this case, if $\alpha_i \in \mathcal{A}^r$ and $\alpha_i \sim [a_i, b_i)$, then $a_i \leq r < b_i$ by naturality of α_i . Also, evaluating all the elements from \mathcal{A}^r on $1_{\mathbb{F}}$ leads to a basis $\mathcal{A}^r(1_{\mathbb{F}})$ for $\mathbb{V}(r)$, where

$$\mathcal{A}^r(1_{\mathbb{F}}) = \{\alpha_i(r)(1_{\mathbb{F}}) : \alpha_i \in \mathcal{A}^r\}.$$

Proposition 3.2 *Given a persistence morphism $\alpha : \bigoplus_{i=1}^N \mathbb{F}_{[a_i, b_i)} \rightarrow \mathbb{V}$, consider $\mathcal{A} = \{\alpha_i\}_{i=1}^N$. Then, \mathcal{A} is a barcode basis for \mathbb{V} if and only if $\mathcal{A}^r(1_{\mathbb{F}})$ is a basis for $\mathbb{V}(r)$ for all $r \in \mathbf{R}$.*

Proof Since $\mathbf{Vect}_{\mathbb{F}}$ is an abelian category and \mathbf{R} is a small category, α is an isomorphism if and only if $\alpha(r)$ is an isomorphism for all $r \in \mathbf{R}$. That is, the kernel $\text{Ker}(\alpha) \hookrightarrow \bigoplus_{i=1}^N \mathbb{F}_{[a_i, b_i)}$ vanishes iff $\text{Ker}(\alpha)(r) = 0$ for all $r \in \mathbf{R}$. A similar argument is done for surjectivity. Then, $\alpha(r)$ is an isomorphism iff $\mathcal{A}^r(1_{\mathbb{F}})$ is a base for $\mathbb{V}(r)$ and the result follows. □

Next, we use barcode bases to understand persistence morphisms $f : \mathbb{V} \rightarrow \mathbb{W}$. In particular, fixing a pair of barcode bases \mathcal{A} and \mathcal{B} for \mathbb{V} and \mathbb{W} respectively, we show that there is a unique matrix F associated to f , see Corollary 3.12. However, this makes no sense unless we are able to perform additions on barcode generators, which, as shown in the following example, cannot be done in a straight-away manner.

Example 3.3 Consider $\mathbb{V} \simeq \mathbb{F}_{[0,2)} \oplus \mathbb{F}_{[1,3)}$ together with the canonical basis \mathcal{A} given by generators $\alpha_1 \sim [0, 2)$ and $\alpha_2 \sim [1, 3)$. Then, it is not possible to add α_1 and α_2 since at some points the domains differ; for example at filtration value 0 we have that $\alpha_1(0)$ has \mathbb{F} as domain but $\alpha_2(0)$ has 0 as domain.

To fix this, consider the following set of pairs:

$$\text{PVect}(\mathbb{V}) = \{(\gamma, (a_\gamma, b_\gamma)) \mid \gamma: \mathbb{F}_{[a_\gamma, b_\gamma)} \rightarrow \mathbb{V} \text{ where } \gamma(r) \neq 0 \text{ iff } r \in [a_\gamma, b_\gamma)\}.$$

Given $(\gamma, (a_\gamma, b_\gamma)) \in \text{PVect}(\mathbb{V})$, we call the first component, γ , a *persistence vector* and say that γ is associated to the pair (a_γ, b_γ) , which we denote as $\gamma \sim [a_\gamma, b_\gamma)$. Notice that in our definition of $\text{PVect}(\mathbb{V})$ we include $(Z_r, (r, r))$ where $Z_r: 0 \rightarrow \mathbb{V}$ is the zero morphism; here we can distinguish Z_r and Z_s since they are associated to different pairs for $r \neq s$. We define $\mathcal{Z} \subseteq \text{PVect}(\mathbb{V})$ to be the subset of zero element pairs $(Z_r, (r, r))$ for all $r \in \mathbf{R}$. We show that $\text{PVect}(\mathbb{V})$ has many properties analogous to those of vector spaces.

Definition 3.4 (*barcode sum*) We define the *barcode sum* as the assignment

$$\boxplus: \text{PVect}(\mathbb{V}) \times \text{PVect}(\mathbb{V}) \longrightarrow \text{PVect}(\mathbb{V})$$

which sends $((\gamma, (a_\gamma, b_\gamma)), (\tau, (a_\tau, b_\tau)))$ to the pair $(\gamma \boxplus \tau, (\max(a_\gamma, a_\tau), B_{\gamma\tau}))$ where we set

$$S_{\gamma\tau} = \begin{cases} \min(b_\gamma, b_\tau) & \text{if } b_\gamma \neq b_\tau, \\ \sup\{r \in [\max(a_\gamma, a_\tau), b_\gamma) : \gamma(r) + \tau(r) \neq 0\} & \text{if } b_\gamma = b_\tau, \end{cases}$$

and

$$B_{\gamma\tau} = \begin{cases} \max(b_\gamma, b_\tau) & \text{if } b_\gamma \neq b_\tau, \\ S_{\gamma\tau} & \text{if } b_\gamma = b_\tau. \end{cases}$$

We define $\gamma \boxplus \tau$ for each $r \in [\max(a_\gamma, a_\tau), B_{\gamma\tau})$ by

$$\gamma \boxplus \tau(r) = \begin{cases} \gamma(r) + \tau(r) & \text{for } r \in [\max(a_\gamma, a_\tau), S_{\gamma\tau}), \\ \gamma(r) & \text{for } r \in [b_\tau, b_\gamma) \text{ if } b_\tau < b_\gamma, \\ \tau(r) & \text{for } r \in [b_\gamma, b_\tau) \text{ if } b_\gamma < b_\tau. \end{cases}$$

One can check that $\gamma \boxplus \tau: \mathbb{F}_{[\max(a_\gamma, a_\tau), B_{\gamma\tau})} \rightarrow \mathbb{V}$ is a well-defined persistence morphism.

By definition, \boxplus is commutative and $\gamma \boxplus \tau(r) \neq 0$ iff $r \in [\max(a_\gamma, a_\tau), B_{\gamma\tau})$. For brevity, given $(\gamma, (a_\gamma, b_\gamma)) \in \text{PVect}(\mathbb{V})$, we refer only to the first component γ . By abuse of notation, we say “given a persistence vector $\gamma \in \text{PVect}(\mathbb{V})$ ” or “given a subset of persistence vectors $\mathcal{S} \subseteq \text{PVect}(\mathbb{V})$ ”. Also, elements $Z_r \in \mathcal{Z}$ behave nontrivially with respect to \boxplus ; for example, given a persistence vector $\gamma \sim [a_\gamma, b_\gamma)$ and considering $c > b_\gamma$ we have that $\gamma \boxplus Z_c = Z_c$.

Proposition 3.5 \boxplus is associative in $\text{PVect}(\mathbb{V})$.

Proof Consider three persistence vectors $\gamma \sim [a_\gamma, b_\gamma]$, $\tau \sim [a_\tau, b_\tau]$, and $\rho \sim [a_\rho, b_\rho]$ from $\text{PVect}(\mathbb{V})$. We show $L = R$ for $L := (\gamma \boxplus \tau) \boxplus \rho$ and $R := \gamma \boxplus (\tau \boxplus \rho)$. Notice that both L and R will share the same startpoint $A = \max(a_\gamma, a_\tau, a_\rho)$; thus L and R are associated to a pair of intervals $[A, B_L]$ and $[A, B_R]$ respectively. Additionally, we have that $L(A) = \gamma(A) + \tau(A) + \rho(A) = R(A)$. Thus, by naturality, it follows that

$$L(r) = \mathbb{V}(A \leq r)(L(A)) = \mathbb{V}(A \leq r)(R(A)) = R(r),$$

for all $r \in \mathbf{R}$ with $A \leq r$. Since $L(r) \neq 0$ iff $r \in [A, B_L]$ and $R(r) \neq 0$ iff $r \in [A, B_R]$ we must have $B_L = B_R$ and the equality $L = R$ holds. \square

Definition 3.6 (scalar multiplication) Define $\lambda: \mathbb{F} \times \text{PVect}(\mathbb{V}) \rightarrow \text{PVect}(\mathbb{V})$ to send (c, γ) , with $\gamma \sim [a_\gamma, b_\gamma]$, to either $c\gamma$, if $c \neq 0$ (where $c\gamma(r) = c \cdot \gamma(r)$ for all $r \in \mathbf{R}$), or to Z_{a_γ} , if $c = 0$.

Now we are ready to introduce the key for characterizing barcode bases.

Definition 3.7 Let $\mathcal{T} \subseteq \text{PVect}(\mathbb{V}) \setminus \mathcal{Z}$. We say that \mathcal{T} is linearly independent or the elements from \mathcal{T} are linearly independent iff for any nonempty subset $\mathcal{S} \subseteq \mathcal{T}$ and any coefficients $k_\gamma \in \mathbb{F} \setminus \{0\}$ with $\gamma \in \mathcal{S}$, the sum $\boxplus_{\gamma \in \mathcal{S}} k_\gamma \gamma$ is associated to $[\max_{\gamma \in \mathcal{S}}(a_\gamma), \max_{\gamma \in \mathcal{S}}(b_\gamma)]$, where $\gamma \sim [a_\gamma, b_\gamma]$ for all $\gamma \in \mathcal{S}$.

Example 3.8 Suppose that $\{\alpha_1 \sim [0, 2), \alpha_2 \sim [0, 1)\}$ is a barcode basis of $\mathbb{V} \simeq \mathbb{F}_{[0,2)} \oplus \mathbb{F}_{[0,1)}$. Then, $\{\alpha_1, \alpha_2\}$ is linearly independent as will follow from Proposition 3.11. On the contrary, α_1 and $\alpha_1 \boxplus \alpha_2$ are not linearly independent since $(-\alpha_1) \boxplus (\alpha_1 \boxplus \alpha_2) = \alpha_2$ is associated to $[0, 1)$ but $-\alpha_1 \sim [0, 2)$ and $\alpha_1 \boxplus \alpha_2 \sim [0, 2)$.

In Proposition 3.11 we show that a barcode base is linearly independent. However, we would like that a barcode base also generates the set $\text{PVect}(\mathbb{V})$. For this, we need to introduce a further ingredient.

Definition 3.9 (barcode cuts) Let $s \in \mathbf{R}$. We define $\mathbf{1}_s: \text{PVect}(\mathbb{V}) \rightarrow \text{PVect}(\mathbb{V})$ as $\mathbf{1}_s(\alpha) = \alpha \boxplus Z_s$, for all $\alpha \in \text{PVect}(\mathbb{V})$.

Notice that λ and $\{\mathbf{1}_s\}_{s \in \mathbf{R}}$ are compatible, in the sense that $\mathbf{1}_r(c\gamma) = c\mathbf{1}_r(\gamma)$ for all $\gamma \in \text{PVect}(\mathbb{V})$, all $r \in \mathbf{R}$ and all $c \in \mathbb{F}$. Also, given $\gamma, \tau \in \text{PVect}(\mathbb{V})$ and $s, r \in \mathbf{R}$, it follows that $\mathbf{1}_s(\gamma) \boxplus \mathbf{1}_r(\tau) = \mathbf{1}_{\max(s,r)}(\gamma \boxplus \tau)$. Thus, persistence vectors on \mathbb{V} correspond to a tuple $(\text{PVect}(\mathbb{V}), \boxplus, \lambda, \{\mathbf{1}_s\}_{s \in \mathbf{R}})$.

Definition 3.10 Given $\mathcal{T} \subseteq \text{PVect}(\mathbb{V})$, we say that \mathcal{T} generates $\text{PVect}(\mathbb{V})$ iff for any $\gamma \in \text{PVect}(\mathbb{V}) \setminus \mathcal{Z}$, there exists $\mathcal{S} \subseteq \mathcal{T}$ together with coefficients $k_\gamma \in \mathbb{F} \setminus \{0\}$ for all $\gamma \in \mathcal{S}$, and some $s \in \mathbf{R}$ such that

$$\alpha = \mathbf{1}_s \left(\boxplus_{\gamma \in \mathcal{S}} k_\gamma \gamma \right).$$

Proposition 3.11 $\mathcal{A} \subseteq \text{PVect}(\mathbb{V})$ is a barcode basis for \mathbb{V} iff it generates $\text{PVect}(\mathbb{V})$ and is linearly independent.

Proof Assume first that \mathcal{A} is a barcode basis. By Proposition 3.2, $\mathcal{A}^r(1_{\mathbb{F}})$ is a basis for $\mathbb{V}(r)$ for all $r \in \mathbb{F}$. Then, for any $\gamma \in \text{PVect}(\mathbb{V})$ with $\gamma \sim [a_\gamma, b_\gamma)$, we must have $\gamma(a_\gamma)(1_{\mathbb{F}}) = \sum_{\alpha \in \mathcal{A}^{a_\gamma}} k_\alpha \alpha(a_\gamma)(1_{\mathbb{F}})$ for some coefficients $k_\alpha \in \mathbb{F}$ and all $\alpha \in \mathcal{A}^{a_\gamma}$. By naturality of γ , this implies $\gamma = \mathbf{1}_{a_\gamma}(\bigoplus_{\alpha \in \mathcal{A}^{a_\gamma}} k_\alpha \alpha)$, which proves that \mathcal{A} generates $\text{PVect}(\mathbb{V})$. On the other hand, assume that \mathcal{A} is not linearly independent. Then there exist some nonempty subset $\mathcal{S} \subseteq \mathcal{A}$ together with coefficients $k_\alpha \in \mathbb{F} \setminus \{0\}$ such that $\bigoplus_{\alpha \in \mathcal{S}} k_\alpha \alpha$ is associated to an interval $[\max_{\alpha \in \mathcal{S}}(a_\alpha), B)$ with $B < \max_{\alpha \in \mathcal{S}}(b_\alpha)$. However, this implies that $\mathcal{A}^B(1_{\mathbb{F}})$ is not linearly independent in $\mathbb{V}(B)$, since $\sum_{\alpha \in \mathcal{S} \cap \mathcal{A}^B} k_\alpha \alpha(B)(1_{\mathbb{F}}) = 0$, but $k_\alpha \in \mathbb{F} \setminus \{0\}$ for all $\alpha \in \mathcal{S} \cap \mathcal{A}^B \neq \emptyset$, reaching a contradiction. Thus, \mathcal{A} must be linearly independent.

Now, suppose that \mathcal{A} generates $\text{PVect}(\mathbb{V})$ and is linearly independent. We prove that \mathcal{A} is a barcode base by using Proposition 3.2. That is, we only need to show that $\mathcal{A}^r(1_{\mathbb{F}})$ is a basis for $\mathbb{V}(r)$ for all $r \in \mathbb{R}$. Thus, let us show that $\mathcal{A}^r(1_{\mathbb{F}})$ generates $\mathbb{V}(r)$ for all $r \in \mathbb{R}$. Let $g \in \mathbb{V}(r)$ with $g \neq 0$ and define the persistence vector $\gamma: \mathbb{F}_{[r, s)} \rightarrow \mathbb{V}$ by setting $\gamma(r)(1_{\mathbb{F}}) = g$, where $s = \sup \{a : \mathbb{V}(r \leq a)(g) \neq 0\}$. Thus, by generation of \mathcal{A} , there exists some subset $\mathcal{S} \subseteq \mathcal{A}$ together with some coefficients $k_\alpha \in \mathbb{F}$ such that $\gamma = \mathbf{1}_r(\bigoplus_{\alpha \in \mathcal{S}} k_\alpha \alpha)$. In particular, $g = \gamma(r)(1_{\mathbb{F}}) = \sum_{\alpha \in \mathcal{S}} k_\alpha \alpha(r)(1_{\mathbb{F}})$ and the claim follows. To show that $\mathcal{A}^r(1_{\mathbb{F}})$ is linearly independent, we consider any non-empty subset $\mathcal{S} \subseteq \mathcal{A}^r$ together with coefficients $k_\alpha \in \mathbb{F} \setminus \{0\}$ for all $\alpha \in \mathcal{S}$. Then $\Gamma = \bigoplus_{\alpha \in \mathcal{S}} k_\alpha \alpha$ is associated to $[\max_{\alpha \in \mathcal{S}}(a_\alpha), \max_{\alpha \in \mathcal{S}}(b_\alpha))$ which must contain r , and so $\Gamma(r)(1_{\mathbb{F}}) = \sum_{\alpha \in \mathcal{S}} k_\alpha \alpha(r)(1_{\mathbb{F}}) \neq 0$. Altogether $\mathcal{A}^r(1_{\mathbb{F}})$ is a basis for $\mathbb{V}(r)$. \square

Let $f: \mathbb{V} \rightarrow \mathbb{W}$ be a persistence morphism and consider two bases \mathcal{A} and \mathcal{B} for \mathbb{V} and \mathbb{W} respectively. Given $\gamma \in \text{PVect}(\mathbb{V})$ such that $\gamma \sim [a_\gamma, b_\gamma)$, we define $f(\gamma)$ as the persistence vector $f(\gamma): \mathbb{F}_{[a_\gamma, b_{f(\gamma)})} \rightarrow \mathbb{W}$ where $b_{f(\gamma)} = \sup \{r \in [a_\gamma, b_\gamma) : f(r) \circ \gamma(r) \neq 0\}$ and $f(\gamma)(r) := f(r) \circ \gamma(r)$ for all $r \in [a_\gamma, b_{f(\gamma)})$. Now, for each $\alpha \in \mathcal{A}$ with $\alpha \sim [a_\alpha, b_\alpha)$, as \mathcal{B} is a barcode base, there exist some subset $\mathcal{S} \subseteq \mathcal{B}$ together with coefficients $k_{\beta, \alpha} \in \mathbb{F} \setminus \{0\}$ for all $\beta \in \mathcal{S}$ such that

$$f(\alpha) = \mathbf{1}_{a_\alpha} \left(\bigoplus_{\beta \in \mathcal{S}} k_{\beta, \alpha} \beta \right).$$

It follows that $\mathcal{S} \subset \mathcal{B}^{a_\alpha}$ since adding elements from $\mathcal{B} \setminus \mathcal{B}^{a_\alpha}$ would have no effect or would cut the startpoint to a value greater than a_α . Also, notice that if $\beta(b_\alpha) \neq 0$ then $\beta \notin \mathcal{S}$, since otherwise f would not be natural as a persistence morphism. Thus, \mathcal{S} must be a subset of

$$\mathcal{B}(\alpha) := \{ \beta : \beta \in \mathcal{B}, \beta(a_\alpha) \neq 0, \beta(b_\alpha) = 0 \} \subseteq \mathcal{B}^{a_\alpha} \subseteq \mathcal{B}.$$

By pointwise-linearity and naturality of f , for any $\mathcal{S} \subseteq \mathcal{A}$ the equality

$$f \left(\bigoplus_{\alpha \in \mathcal{S}} k_\alpha \alpha \right) = \bigoplus_{\alpha \in \mathcal{S}} k_\alpha f(\alpha)$$

holds, where $k_\alpha \in \mathbb{F}$ for all $\alpha \in \mathcal{S}$.

Corollary 3.12 *Let \mathbb{V} and \mathbb{W} be a pair of persistence modules together with their respective barcode bases \mathcal{A} and \mathcal{B} . Given a persistence morphism $f: \mathbb{V} \rightarrow \mathbb{W}$, there is a unique associated matrix $F = (k_{\beta,\alpha})_{\beta \in \mathcal{B}, \alpha \in \mathcal{A}}$ which is well defined in the sense that whenever $k_{\beta,\alpha} \neq 0$ then $\beta \in \mathcal{B}(\alpha)$. Conversely, assume that F is well defined, then there exists a unique persistence morphism $f: \mathbb{V} \rightarrow \mathbb{W}$ whose associated matrix is F .*

Proof By the reasoning above, we only need to prove the converse statement. First, for each $\alpha \in \mathcal{A}$ such that $\alpha \sim [a_\alpha, b_\alpha]$, we define $f(\alpha) := \mathbf{1}_{a_\alpha}(\bigoplus_{\beta \in \mathcal{B}(\alpha)} k_{\beta,\alpha} \beta)$. By linear independence of \mathcal{B} , $f(\alpha)$ is associated to the interval $[a_\alpha, B]$ for $B = \max_{\beta \in \mathcal{S}}(b_\beta)$ with \mathcal{S} being the set $\{\beta \in \mathcal{B}(\alpha) : k_{\beta,\alpha} \neq 0\}$. We can extend the definition of f by the linear formula $f(\bigoplus_{\alpha \in \mathcal{A}} c_\alpha \alpha) = \bigoplus_{\alpha \in \mathcal{A}} c_\alpha f(\alpha)$ for any coefficients $c_\alpha \in \mathbb{F}$ for all $\alpha \in \mathcal{A}$. This implies the claim as f is then natural and pointwise linear. \square

We end this section by introducing different orders of barcode bases. These orders are important to introduce Gaussian eliminations in the barcode basis context.

Definition 3.13 Let \mathbb{V} be a persistence module with barcode base \mathcal{A} . Also, let $\alpha_i, \alpha_j \in \mathcal{A}$ with $\alpha_i \sim [a_i, b_i]$ and $\alpha_j \sim [a_j, b_j]$. We consider two orders in \mathcal{A} :

The standard order: $\alpha_i < \alpha_j$ if either $a_i < a_j$ or $a_i = a_j$ and $b_i > b_j$.

The endpoint order: $\alpha_i < \alpha_j$ if either $b_i < b_j$ or $b_i = b_j$ and $a_i < a_j$.

If \mathbb{V} is tame, it is straightforward to extend these orders to total orders for \mathcal{A} .

3.2 Computing Kernels and Images

Consider two finite barcode bases $\mathcal{A} = \{\alpha_i\}_{i=1}^n$ and $\mathcal{B} = \{\beta_j\}_{j=1}^m$ for \mathbb{V} and \mathbb{W} , respectively. Additionally, suppose that \mathcal{A} is ordered according to the standard order while \mathcal{B} is ordered using the endpoint order. We assume such orders are total; e.g., even if $\alpha_r, \alpha_s \sim [a, b]$ for $r \neq s$, either $\alpha_r < \alpha_s$ or $\alpha_s < \alpha_r$ holds. Then, we consider $f(\mathcal{A})_{\mathcal{B}} = (f(\alpha_1), \dots, f(\alpha_n))$, the matrix of f in the bases \mathcal{A} and \mathcal{B} . In Sect. 3.3, we transform $f(\mathcal{A})_{\mathcal{B}}$ performing left to right column additions until obtaining a reduced matrix, i.e., with unique column pivots,

$$\mathcal{I}_{\mathcal{B}} = \left(f(\alpha_1) \left| f(\alpha_2) \boxplus k_{2,1} f(\alpha_1) \right. \left| \dots \right. \left| f(\alpha_n) \boxplus \bigoplus_{j=1}^{n-1} k_{n,j} f(\alpha_j) \right. \right) \tag{4}$$

for suitable $k_{i,j} \in \mathbb{F}$ and $1 \leq j < i \leq n$. This $\mathcal{I}_{\mathcal{B}}$ has the property that its non-zero columns form a basis $\tilde{\mathcal{I}}$ for $\text{Im}(f)$.

Definition 3.14 Given $\mathcal{S} \subseteq \mathcal{B}$, consider a vector $V = (k_\beta)_{\beta \in \mathcal{B}}$ such that $k_\beta \neq 0$ iff $\beta \in \mathcal{S}$. The *pivot* of V is the greatest element from \mathcal{S} in the endpoint order. We also refer to the *pivot* of \mathcal{S} or the pivot of $\bigoplus_{\beta \in \mathcal{S}} k_\beta \beta$.

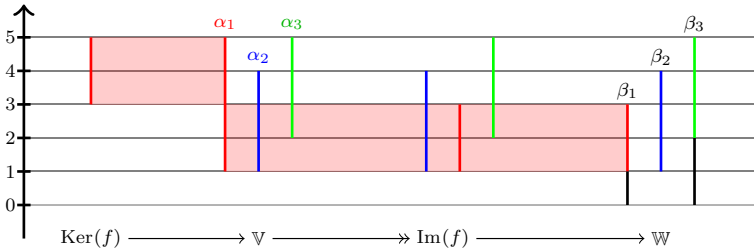


Fig. 1 Decomposition of barcodes in image, kernel, domain and codomain of $f : \mathbb{V} \rightarrow \mathbb{W}$. The colors correspond to the different generators associated to $\tilde{\mathcal{I}}$ and $\tilde{\mathcal{K}}$

Consider again the matrix \mathcal{I}_B from (4). By linearity, the j th column from \mathcal{I} is $f(\alpha_j \boxplus \boxplus_{i=1}^{j-1} k_{j,i} \alpha_i)$; thus, its preimage is $p_j = \alpha_j \boxplus \boxplus_{i=1}^{j-1} k_{j,i} \alpha_i$ and we define the set of preimages of \mathcal{I} by $\mathcal{PT} = \{p_j\}_{j=1}^n$. Given $p_j \in \mathcal{PT}$, notice that $p_j \sim [a_j, b_j]$ while $f(p_j) \sim [a_j, c_j]$ for filtration values $c_j \leq b_j$. Thus, we must have that $\mathbf{1}_{c_j}(p_j) \in \ker(f)$ for all $1 \leq j \leq n$. We consider $\mathcal{GK} = \{\mathbf{1}_{c_j}(p_j)\}_{1 \leq j \leq n}$ which generates $\ker(f)$, as it is shown later in Proposition 3.17. Then, we order \mathcal{GK} by choosing a permutation $\sigma : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ such that it is consistent with the standard order. Using the order from \mathcal{GK} we consider the matrix $\mathcal{GK}_A = (\mathbf{1}_{c_{\sigma(1)}}(p_{\sigma(1)}), \dots, \mathbf{1}_{c_{\sigma(n)}}(p_{\sigma(n)}))$ where the rows correspond to generators from \mathcal{A} with the endpoint order. Reducing columns we find some coefficients $q_{i,j} \in \mathbb{F}$ so that the resulting matrix has unique pivots:

$$\mathcal{K}_A = \left(\mathbf{1}_{c_{\sigma(1)}}(p_{\sigma(1)}) \mid \mathbf{1}_{c_{\sigma(2)}}(p_{\sigma(1)} \boxplus q_{2,2} p_{\sigma(2)}) \mid \dots \mid \mathbf{1}_{c_{\sigma(n)}} \left(p_{\sigma(n)} \boxplus \boxplus_{i=1}^{n-1} q_{j,i} p_{\sigma(i)} \right) \right).$$

Taking the non-zero columns from \mathcal{K}_A leads to a basis $\tilde{\mathcal{K}}$ for $\ker(f)$. In the following we present an algorithm obtaining such bases for $\text{Im}(f)$ and $\ker(f)$. First we go through an illustrative example:

Example 3.15 Consider two persistence modules (see Fig. 1)

$$\mathbb{V} \simeq \mathbb{F}_{[1,5]} \oplus \mathbb{F}_{[1,4]} \oplus \mathbb{F}_{[2,5]}, \quad \mathbb{W} \simeq \mathbb{F}_{[0,3]} \oplus \mathbb{F}_{[1,4]} \oplus \mathbb{F}_{[0,5]},$$

with canonical barcode bases $(\alpha_1, \alpha_2, \alpha_3)$ and $(\beta_1, \beta_2, \beta_3)$ ordered respectively using the standard and endpoint orders. Let $f : \mathbb{V} \rightarrow \mathbb{W}$ be given by the $|\mathcal{B}| \times |\mathcal{A}|$ matrix $f(\mathcal{A})_B$:

$$f(\mathcal{A})_B = \begin{pmatrix} & \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & 1 & 0 & 0 \\ \beta_2 & 1 & 1 & 1 \\ \beta_3 & 0 & 0 & 1 \end{pmatrix} \longrightarrow \mathcal{I}_B = \begin{pmatrix} & \alpha_1 & \alpha_2 \boxplus (-\alpha_1) & \alpha_3 \\ \beta_1 & 1 & -1 & 0 \\ \beta_2 & 1 & 0 & 1 \\ \beta_3 & 0 & 0 & 1 \end{pmatrix}.$$

Then notice that the first two columns from $f(\mathcal{A})_B$ share the same pivot β_2 , while the third's column pivot is β_3 . We subtract the first column to the second, leading to

the matrix $\mathcal{I}_{\mathcal{B}}$ above which has unique pivots for each column. From $\mathcal{I}_{\mathcal{B}}$ we obtain

$$f(\alpha_1) = \mathbf{1}_1(\beta_1 \boxplus \beta_2), \quad f(\alpha_2 \boxplus (-\alpha_1)) = -\mathbf{1}_1(\beta_1), \quad f(\alpha_3) = \mathbf{1}_2(\beta_2 \boxplus \beta_3).$$

In particular, we obtain a basis for the image

$$\tilde{\mathcal{I}} = \{\mathbf{1}_1(\beta_1 \boxplus \beta_2), -\mathbf{1}_1(\beta_1), \mathbf{1}_2(\beta_2 \boxplus \beta_3)\},$$

which leads to the barcode decomposition $\text{Im}(f) \simeq \mathbb{F}_{[1,4]} \oplus \mathbb{F}_{[1,3]} \oplus \mathbb{F}_{[2,5]}$. At the same time, we obtain a corresponding set of preimages $\mathcal{PT} = \{\alpha_1, \alpha_2 \boxplus (-\alpha_1), \alpha_3\}$. From this we deduce the set of kernel generators \mathcal{GK} and order it by the standard barcode order $\mathcal{GK} = \{\mathbf{1}_3(\alpha_2 \boxplus (-\alpha_1)), \mathbf{1}_4(\alpha_1), \mathbf{1}_5(\alpha_3)\}$. Thus, we consider the matrix $\mathcal{GK}_{\mathcal{A}}$ for the kernels, where the rows correspond to the endpoint order on \mathcal{A} , and reduce it:

$$\mathcal{GK}_{\mathcal{A}} = \left(\begin{array}{c|ccc} & \mathbf{1}_3(\alpha_2 \boxplus (-\alpha_1)) & \mathbf{1}_4(\alpha_1) & \mathbf{1}_5(\alpha_3) \\ \hline \alpha_2 & & 1 & 0 & 0 \\ \alpha_1 & & -1 & 1 & 0 \\ \alpha_3 & & 0 & 0 & 1 \end{array} \right) \longrightarrow \mathcal{K}_{\mathcal{A}} = \left(\begin{array}{c|ccc} & \mathbf{1}_3(\alpha_2 \boxplus (-\alpha_1)) & \mathbf{1}_4(\alpha_2) & \mathbf{1}_5(\alpha_3) \\ \hline \alpha_2 & & 1 & 1 & 0 \\ \alpha_1 & & -1 & 0 & 0 \\ \alpha_3 & & 0 & 0 & 1 \end{array} \right)$$

Notice that the second and third columns from \mathcal{K} are trivial since $\mathbf{1}_4(\alpha_2) = Z_4$ and $\mathbf{1}_5(\alpha_3) = Z_5$. Since \mathcal{K} contains a single nontrivial element we have obtained a basis for the kernel $\tilde{\mathcal{K}} = \{\mathbf{1}_3(\alpha_2 \boxplus (-\alpha_1))\}$. Thus, $\ker(f) \simeq \mathbb{F}_{[3,5]}$.

Finally, we explain why uniqueness of pivots leads to linear independence.

Proposition 3.16 *Let \mathbb{V} be a persistence module with a barcode basis \mathcal{A} ordered using the endpoint order. Consider a subset of persistence vectors $\mathcal{M} \subset \text{PVect}(\mathbb{V})$ and suppose that their pivots on \mathcal{A} are all different. Then \mathcal{M} is linearly independent.*

Proof Let $m \in \mathcal{M}, m \sim [a_m, b_m)$, and write it in terms of \mathcal{A} as $m = \mathbf{1}_{a_m}(\boxplus_{\alpha \in \mathcal{S}_m} k_{\alpha} \alpha)$, where $\mathcal{S}_m \subseteq \mathcal{A}$ and $k_{\alpha} \in \mathbb{F} \setminus \{0\}$ for all $\alpha \in \mathcal{S}_m$. Let $\alpha_m \in \mathcal{S}_m$ be the pivot of m . In particular, by the definition of pivot we must have $\alpha_m \sim [c_m, b_m)$ for some filtration value $c_m \leq a_m$. Now, consider a nonempty subset $\mathcal{R} \subseteq \mathcal{M}$ together with coefficients $q_m \in \mathbb{F} \setminus \{0\}$ for all $m \in \mathcal{R}$ and take the sum $V = \boxplus_{m \in \mathcal{R}} q_m m$. We claim that $V \sim [\max_{m \in \mathcal{R}}(a_m), \max_{m \in \mathcal{R}}(b_m))$. Consider the element $P \in \mathcal{R}$ whose pivot α_P is the highest according to the endpoint order. Consequently, we have that $b_P = \max_{m \in \mathcal{R}}(b_m)$. Notice that P is unique, since otherwise there would be two elements from \mathcal{M} with the same pivots, contradicting our assumption. This implies that V is written in terms of \mathcal{A} with a nonzero coefficient for α_P . By linear independence of \mathcal{A} the claim follows. \square

3.3 The `box_gauss_reduce` Algorithm

Suppose that $f : \mathbb{V} \rightarrow \mathbb{W}$ is a morphism between two tame persistence modules. Let \mathcal{A} , with the standard order, and \mathcal{B} , with the endpoint order, be barcode bases for \mathbb{V} and \mathbb{W} respectively. Suppose also that we know $f(\mathcal{A})_{\mathcal{B}}$, the matrix associated to f with respect to \mathcal{A} and \mathcal{B} . Sending this information to Algorithm 1, called `box_gauss_reduce`², we obtain a set of persistence vectors $\mathcal{I} \subseteq \text{PVect}(f(\mathbb{V}))$ together with the coordinates, in \mathcal{A} , of their respective preimages stored in $\mathcal{PI}_{\mathcal{A}}$. Then, using $\mathcal{PI}_{\mathcal{A}}$, we obtain a set of generators \mathcal{GK} for $\ker(f)$, order it according to the standard order and define a matrix $\mathcal{GK}_{\mathcal{A}}$ which expresses the elements from \mathcal{GK} in terms of \mathcal{A} . Reducing $\mathcal{GK}_{\mathcal{A}}$ by using Algorithm 1 again, we end up with \mathcal{K} . Taking the nonzero elements from \mathcal{I} and \mathcal{K} leads to the barcode bases $\tilde{\mathcal{I}}$ and $\tilde{\mathcal{K}}$. An outline of this procedure is shown in Algorithm 2, which we call `image_kernel`.

Algorithm 1 `box_gauss_reduce`

Input: $\mathcal{A}, \mathcal{B}, f(\mathcal{A})_{\mathcal{B}}$, where \mathcal{B} follows the endpoint order
Output: $\mathcal{I}, \mathcal{PI}_{\mathcal{A}}$.

- 1: Set $M = \left(\frac{\text{Id}_{|\mathcal{A}|}}{f(\mathcal{A})_{\mathcal{B}}} \right)$ and store pivots of columns in a list called `lpivots`.
- 2: With the same order, store birth values of elements from \mathcal{A} into `lbirths`.
- 3: With the same order, store death values of elements from \mathcal{B} into `ldeaths`.
- 4: **for** $p = |\mathcal{A}| + |\mathcal{B}| \geq \dots \geq |\mathcal{A}| + 1$, with p decreasing, **do**
- 5: Find all $1 \leq i \leq |\mathcal{A}|$ such that `lpivots[i] == p`; store into a new ordered list called `lpiv_idx`.
- 6: **while** $L = \text{length}(\text{lpiv_idx}) > 1$ **do**
- 7: Find $1 \leq m < L$ such that `lbirths[lpiv_idx[m]]` is minimal. For multiple choices, take the smallest m .
- 8: **for** $j = L \geq \dots \geq m + 1$ **do**
- 9: `idx ← lpiv_idx.pop()`
- 10: `lbirths[idx] ← max(lbirths[idx], lbirths[lpiv_idx[m]])`
- 11: $M[:, \text{idx}] \leftarrow M[:, \text{idx}] - \left(\frac{M[p, \text{idx}]}{M[p, \text{lpiv_idx}[m]]} \right) M[:, \text{lpiv_idx}[m]]$
- 12: **for** $k = |\mathcal{A}| + 1 \leq \dots \leq p$ **do**
- 13: If `ldeaths[k] <= lbirths[idx]`, set $M[k, \text{idx}] \leftarrow 0$.
- 14: **end for**
- 15: `lpivots[idx] ← pivot from M[:, idx]`.
- 16: **end for**
- 17: **end while**
- 18: **end for**
- 19: Denote by $\mathcal{PI}_{\mathcal{A}}$ and $\mathcal{I}_{\mathcal{B}}$ the matrices such that $M = \left(\frac{\mathcal{PI}_{\mathcal{A}}}{\mathcal{I}_{\mathcal{B}}} \right)$.
- 20: Compute $S^i = \{\beta_j \in \mathcal{B} : \mathcal{I}_{\mathcal{B}}[j, i] \neq 0\}$ for all $1 \leq i \leq |\mathcal{A}|$.
- 21: Set $\mathcal{I} = \{\mathbf{1}_{a_i} (\boxplus_{\beta_j \in S^i} M[j, i]\beta_j)\}_{1 \leq i \leq |\mathcal{A}|}$, where $a_i = \text{lbirths}[i]$.
- 22: **return** $\mathcal{I}, \mathcal{PI}_{\mathcal{A}}$

Proposition 3.17 *Algorithm 2 computes $\tilde{\mathcal{K}}$ and $\tilde{\mathcal{I}}$ bases for the kernel and image of f .*

² Here we use the Numpy notation for matrices, where for a matrix M , the (i, j) -entry is denoted by $M[i, j]$ and the j th column is denoted by $M[:, j]$. Also, we use indexing starting at 1 instead of 0 to be consistent with the rest of the article. In addition, we make use of the `pop()` function, which simultaneously returns and deletes the last element from a list.

Algorithm 2 `image_kernel`

Input: $\mathcal{A}, \mathcal{B}, f(\mathcal{A})_{\mathcal{B}}$; \mathcal{A} and \mathcal{B} follow the startpoint and endpoint orders resp.

Output: $\tilde{\mathcal{K}}, \tilde{\mathcal{I}}, \mathcal{P}\tilde{\mathcal{I}}$

- 1: $\mathcal{I}, \mathcal{P}\mathcal{I}_{\mathcal{A}} \leftarrow \text{box_gauss_reduce}(\mathcal{A}, \mathcal{B}, f(\mathcal{A})_{\mathcal{B}})$.
- 2: Compute $\mathcal{T}^i = \{\alpha_j \in \mathcal{A} : \mathcal{P}\mathcal{I}_{\mathcal{A}}[j, i] \neq 0\}$ for all $1 \leq i \leq |\mathcal{A}|$.
- 3: Set $\mathcal{P}\mathcal{I} = \{\boxplus_{\alpha_j \in \mathcal{T}^i} \mathcal{P}\mathcal{I}_{\mathcal{A}}[j, i] \alpha_j\}_{1 \leq i \leq |\mathcal{A}|}$.
- 4: Set $\mathcal{G}\mathcal{K}$ to contain all $\mathbf{1}_{c_i}(p_i)$ for all $p_i \in \mathcal{P}\mathcal{I}$ with $f(p_i) \sim [a_i, c_i]$.
- 5: Order $\mathcal{G}\mathcal{K}$ with standard order, while \mathcal{A} with the endpoint order.
- 6: Write the matrix of coordinates, $\mathcal{G}\mathcal{K}_{\mathcal{A}}$, of $\mathcal{G}\mathcal{K}$ in terms of \mathcal{A} .
- 7: $\mathcal{K}, _ \leftarrow \text{box_gauss_reduce}(\mathcal{G}\mathcal{K}, \mathcal{A}, \mathcal{G}\mathcal{K}_{\mathcal{A}})$
- 8: We get rid of zero elements to obtain $\tilde{\mathcal{K}}$ and $\tilde{\mathcal{I}}$ from \mathcal{K} and \mathcal{I} .
- 9: **return** $\tilde{\mathcal{K}}, \tilde{\mathcal{I}}$ and $\mathcal{P}\tilde{\mathcal{I}}$ (where $\mathcal{P}\tilde{\mathcal{I}}$ are the preimages of $\tilde{\mathcal{I}}$ taken from $\mathcal{P}\mathcal{I}$).

Proof First of all, by Proposition 3.16, we know that $\tilde{\mathcal{I}}$ and $\tilde{\mathcal{K}}$ are linearly independent, as these are both sets of persistence vectors with different pivots. Thus, all we need to show is that both sets generate $\text{PVect}(\text{Im}(f))$ and $\text{PVect}(\text{Ker}(f))$, respectively.

Let us prove that $\tilde{\mathcal{I}}$ generates $\text{PVect}(\text{Im}(f))$. First, we will show that $\mathcal{P}\mathcal{I}$ generates $\text{PVect}(\mathbb{V})$. Consider $\gamma \in \text{PVect}(\mathbb{V})$ and write $\gamma = \mathbf{1}_a(\boxplus_{i \in I} k_i \alpha_i)$ for coefficients $k_i \in \mathbb{F} \setminus \{0\}$ with $i \in I$ for some subset $I \subseteq \{1, 2, \dots, |\mathcal{A}|\}$. Then, consider the maximum index m from I and compute $\tilde{\gamma} = \gamma \boxplus (-k_m p_m)$. Here it is key to recall that the preimage $p_m \in \mathcal{P}\mathcal{I}$ is written as $\alpha_m \boxplus \boxplus_{i=1}^{m-1} k_{m,i} \alpha_i$ for coefficients $k_{m,i} \in \mathbb{F}$ for $1 \leq i < m$. Now, $\tilde{\gamma} = \mathbf{1}_a(\boxplus_{i \in J} \tilde{k}_i \alpha_i)$ for coefficients $\tilde{k}_i \in \mathbb{F} \setminus \{0\}$ with $i \in J$ for some subset $J \subseteq \{1, 2, \dots, m - 1\}$. Repeating this argument, eventually, we write γ in terms of $\mathcal{G}\mathcal{I}$. This implies that $f(\gamma)$ can be expressed in terms of $\tilde{\mathcal{I}}$. Thus, $\tilde{\mathcal{I}}$ generates $\text{PVect}(\text{Im}(f))$.

Now, let us show that $\tilde{\mathcal{K}}$ generates $\text{PVect}(\text{Ker}(f))$. In fact, it will be enough to show that $\mathcal{G}\mathcal{K}$ generates $\text{PVect}(\text{ker}(f))$. This is because $\tilde{\mathcal{K}}$ is obtained from reducing $\mathcal{G}\mathcal{K}_{\mathcal{A}}$ in a similar manner as $\tilde{\mathcal{I}}$ was obtained by reducing $f(\mathcal{A})_{\mathcal{B}}$. Consequently, by replicating the argument which proved that $\mathcal{P}\mathcal{I}$ generates $\text{PVect}(\mathbb{V})$, it follows that $\tilde{\mathcal{K}}$ generates $\text{PVect}(\text{ker}(f))$. So let us prove our claim. Suppose that $\gamma : \mathbb{F}_{[a,b]} \rightarrow \mathbb{V}$ lies in the kernel; i.e., $f(\gamma) = Z_a$. As $\mathcal{P}\mathcal{I}$ generates $\text{PVect}(\mathbb{V})$, we have that $\gamma = \mathbf{1}_a(\boxplus_{i \in I} k_i p_i)$ for coefficients $k_i \in \mathbb{F} \setminus \{0\}$ with $i \in I$ for some subset $I \subseteq \{1, 2, \dots, |\mathcal{A}|\}$. Applying f , we obtain the equality $f(\gamma) = Z_a = \boxplus_{i \in I} k_i f(p_i)$ and notice that $f(p_i)(a) = 0$ for all $i \in I$; otherwise linear independence of $\tilde{\mathcal{I}}$, and in particular that of $\mathcal{I}^a(1_{\mathbb{F}})$, would be contradicted. However, if $f(p_i)(a) = 0$ for all $i \in I$, then $\mathbf{1}_{c_i}(p_i) \in \mathcal{G}\mathcal{K}$ for some $c_i \leq a$ and all $i \in I$. Altogether we obtain that γ must be generated by $\mathcal{G}\mathcal{K}$. \square

Notice that Algorithm 1 for the cases from Proposition 3.17 is simply a Gaussian elimination where the procedure differs from the standard method. In the former case, the input \mathcal{A} in Algorithm 1 is given ordered in the startpoint order. However, this hypotheses is not assumed in Sect. 3.4. Next, we give a computational bound.

Proposition 3.18 *Algorithm 2 takes at most $\mathcal{O}(N^2|\mathcal{A}|)$ time, where $N = \max(|\mathcal{A}|, |\mathcal{B}|)$.*

Proof Let us start by measuring the computational complexity of Algorithm 1, `box_gauss_reduce`. First, the for loop from line 4 iterates $|\mathcal{B}|$ times. Next, for each iteration of line 4 the `pop()` function from line 9 cannot be executed more than L times, where $L \leq |\mathcal{A}|$. Then, lines 9 to 15 have a computational cost of

about $|\mathcal{B}|$. Altogether, the computational complexity of `box_gauss_reduce` is about $\mathcal{O}(|\mathcal{A}||\mathcal{B}|^2)$. Now, let us compute the complexity of Algorithm 2. Executing line 1 should take $\mathcal{O}(|\mathcal{A}||\mathcal{B}|^2)$ time, while executing line 7 should take about $\mathcal{O}(|\mathcal{A}|^3)$ time. This leads to the expected result. \square

3.4 Computing Quotients

Suppose that we have inclusions $\mathbb{H} \subseteq \mathbb{G} \subseteq \mathbb{V}$ of finite dimensional persistence modules, together with barcode bases \mathcal{H}, \mathcal{G} , and \mathcal{A} respectively. Suppose that \mathcal{H} and \mathcal{G} are ordered using the standard order, while \mathcal{A} is ordered using the endpoint order. Consider the inclusions $\iota^{\mathbb{H}}: \mathbb{H} \hookrightarrow \mathbb{V}$ and $\iota^{\mathbb{G}}: \mathbb{G} \hookrightarrow \mathbb{V}$ together with their respective associated matrices $\iota^{\mathbb{H}}(\mathcal{H})_{\mathcal{A}} \in \mathcal{M}_{|\mathcal{A}| \times |\mathcal{H}|}(\mathbb{F})$ and $\iota^{\mathbb{G}}(\mathcal{G})_{\mathcal{A}} \in \mathcal{M}_{|\mathcal{A}| \times |\mathcal{G}|}(\mathbb{F})$. Without loss of generality, we assume that $\iota^{\mathbb{H}}(\mathcal{H})_{\mathcal{A}}$ is already reduced and $\iota^{\mathbb{H}}(\mathcal{H})$ is a barcode base for $\iota^{\mathbb{H}}(\mathbb{H})$. Given all this data, the aim is to find a barcode base for \mathbb{G}/\mathbb{H} .

Let $\mathbb{H} \oplus \mathbb{G}$ together with a barcode base given by the pair $(\mathcal{H} | \mathcal{G})$; here we extend the orders from \mathcal{H} and \mathcal{G} with the rule $h < g$ for any pair of generators $h \in \mathcal{H}$ and $g \in \mathcal{G}$; of course, this might break the standard persistence vector order. Then, we consider $\iota = \iota^{\mathbb{H}} + \iota^{\mathbb{G}}: \mathbb{H} \oplus \mathbb{G} \rightarrow \mathbb{V}$ which will have the associated block matrix $(\iota^{\mathbb{H}}(\mathcal{H})_{\mathcal{A}} | \iota^{\mathbb{G}}(\mathcal{G})_{\mathcal{A}})$. We send the triple $((\mathcal{H} | \mathcal{G}), \mathcal{A}, (\iota^{\mathbb{H}}(\mathcal{H})_{\mathcal{A}} | \iota^{\mathbb{G}}(\mathcal{G})_{\mathcal{A}}))$ to the `box_gauss_reduce` algorithm and obtain the output $\mathcal{I}, \mathcal{PT}_{\mathcal{A}}$. We focus on the subset $\mathcal{I}[\mathcal{G}]$ containing the last $|\mathcal{G}|$ elements from \mathcal{I} .

Recall that the `box_gauss_reduce` algorithm adds columns from $\iota^{\mathbb{H}}(\mathcal{H})_{\mathcal{A}}$ and $\iota^{\mathbb{G}}(\mathcal{G})_{\mathcal{A}}$ to eventually obtain a matrix $\mathcal{I}[\mathcal{G}]_{\mathcal{A}}$ from which we deduce the elements in $\mathcal{I}[\mathcal{G}]$. Further, using $\mathcal{PT}_{\mathcal{A}}$, we are able to know exactly which combinations of columns were performed in the reduction procedure. Thus, each $\Gamma \in \mathcal{I}[\mathcal{G}]$ can be written as $\Gamma = \Gamma^{\mathbb{H}} \boxplus \Gamma^{\mathbb{G}}$ where $\Gamma^{\mathbb{H}}$ and $\Gamma^{\mathbb{G}}$ denote the respective linear combinations of elements from $\iota^{\mathbb{H}}(\mathcal{H})$ and $\iota^{\mathbb{G}}(\mathcal{G})$. Given $\Gamma \in \mathcal{I}[\mathcal{G}]$, we use the notation $\Gamma^{\mathbb{G}} \sim [a_{\Gamma}, c_{\Gamma})$ and $\Gamma \sim [b_{\Gamma}, c_{\Gamma})$ for the corresponding associated intervals; in particular, notice that $a_{\Gamma} \leq b_{\Gamma}$. Then, we define the persistence vector

$$\overline{\Gamma^{\mathbb{G}}}: \mathbb{F}_{[a_{\Gamma}, c_{\Gamma})} \rightarrow \frac{\mathbb{V}}{\mathbb{H}},$$

which is defined by $\overline{\Gamma^{\mathbb{G}}}(r) = p_{\mathbb{H}}(r) \circ \Gamma^{\mathbb{G}}(r)$ for all $r \in [a_{\Gamma}, c_{\Gamma})$, where we use the projection $p_{\mathbb{H}}: \mathbb{V} \rightarrow \mathbb{V}/\mathbb{H}$. We claim that $\overline{\Gamma^{\mathbb{G}}}$ is well defined, i.e., $\overline{\Gamma^{\mathbb{G}}}(r) \neq \overline{0}$ iff $r \in [a_{\Gamma}, c_{\Gamma})$. First, notice that $\overline{\Gamma^{\mathbb{G}}}(c_{\Gamma}) = \overline{0}$ since by definition $\Gamma(c_{\Gamma}) = 0$, which implies $\Gamma^{\mathbb{G}}(c_{\Gamma}) = -\Gamma^{\mathbb{H}}(c_{\Gamma})$. Next, we need to show that $\overline{\Gamma^{\mathbb{G}}}(r) \neq 0$ for all $r \in [a_{\Gamma}, c_{\Gamma})$. In fact, we prove the stronger statement that

$$\tilde{\mathcal{Q}} = \{\overline{\Gamma^{\mathbb{G}}}: \Gamma \in \mathcal{I}[\mathcal{G}] \text{ such that } a_{\Gamma} < c_{\Gamma}\}$$

is linearly independent. Take a subset $\mathcal{S} \subseteq \mathcal{I}[\mathcal{G}]$ such that $\{\overline{\Gamma^{\mathbb{G}}}\}_{\Gamma \in \mathcal{S}}$ is a nonempty subset of $\tilde{\mathcal{Q}}$. Also, take some coefficients $k_{\Gamma} \in \mathbb{F} \setminus \{0\}$ for all $\Gamma \in \mathcal{S}$. We want to show that $\overline{V^{\mathbb{G}}} := \boxplus_{\Gamma \in \mathcal{S}} k_{\Gamma} \overline{\Gamma^{\mathbb{G}}}$ is associated to the interval $[A, C)$, where we use the notation $A = \max_{\Gamma \in \mathcal{S}}(a_{\Gamma})$ and $C = \max_{\Gamma \in \mathcal{S}}(c_{\Gamma})$. By contradiction, suppose that

$\overline{V^G}$ is associated to $[A, r)$ for some value $r \in [A, C)$. This implies that $\mathbf{1}_r(V^G)$ is in $\text{PVect}(\mathbb{H})$, where we define $V^G := \bigsqcup_{\Gamma \in \mathcal{S}} k_\Gamma \Gamma^G$.

Next, take the greatest $\Gamma \in \mathcal{S}$ whose endpoint is C . Thus, there must exist $R \in \mathbb{R}$ such that $r \leq R < C$ and $\mathbf{1}_R(\Upsilon) = Z_R$ for all $\Upsilon \in \mathcal{S}$ such that $\Upsilon > \Gamma$. Equivalently, $\mathbf{1}_R(\Upsilon^G) \in \text{PVect}(\mathbb{H})$ for all $\Upsilon \in \mathcal{S}$ such that $\Upsilon > \Gamma$. Since $\mathbf{1}_R(V^G)$ is in $\text{PVect}(\mathbb{H})$, we conclude that $\mathbf{1}_R(\Gamma^G)$ can be written in terms of generators from $\iota^{\mathbb{H}}(\mathcal{H})^R$ and elements $\tilde{\Gamma}^G$ with $\tilde{\Gamma} < \Gamma$; where, from the third bullet point after Definition 3.1, $\iota^{\mathbb{H}}(\mathcal{H})^R$ denotes the set of generators $\gamma \in \iota^{\mathbb{H}}(\mathcal{H})$ such that $\gamma(R) \neq 0$. Now, consider the matrix $\mathcal{I}[\mathcal{G}]_A^G$ whose columns correspond to the coordinates of Υ^G in terms of \mathcal{A} for all $\Upsilon \in \mathcal{I}[\mathcal{G}]$. Then, consider the block matrix $M = (\iota^{\mathbb{H}}(\mathcal{H})_A^R \mid \mathcal{I}[\mathcal{G}]_A^G)$. In matrix terms, our hypotheses on $\mathbf{1}_R(\Gamma^G)$ means that its corresponding column from M can be reduced by left to right column additions on M up to a pivot whose associated interval endpoint is smaller or equal to R .

Now, denote $N = (\iota^{\mathbb{H}}(\mathcal{H})_A^R \mid \iota^{\mathbb{G}}(\mathcal{G})_A)$. It is not difficult to notice that M is the result of applying left-to-right column additions to N . Consequently, denoting by C_Γ the column from N that corresponds to Γ^G , the column C_Γ can be reduced by left columns in N up to a pivot whose associated interval endpoint is smaller or equal to R . There are two options:

- Assume $\Gamma \sim [b_\Gamma, C)$, with $b_\Gamma < C$. By hypotheses, C_Γ is a combination of previous columns up to a pivot with death value $\leq R$. Thus, following the instructions from Algorithm 1, we should reduce C_Γ (at least) to a pivot with death value $\leq R$. However, in such a case, Γ would be associated to an interval with endpoint $\leq R$. But $R < C$, reaching a contradiction.
- Assume $\Gamma = Z_C$. There exists a pivot of index p in the reduction process where we first add to C_Γ a column \tilde{C} with start value strictly bigger than R ; as $\Gamma = Z_C$ and $a_\Gamma \leq r \leq R < C$. By step 7 in Algorithm 1 and by our hypotheses on C_Γ , such a pivot must have an endpoint smaller or equal to R . However, the column \tilde{C} has a startpoint strictly bigger than R . This contradicts Corollary 3.12 as well as step 13 in Algorithm 1.

It can be shown that $\tilde{\mathcal{Q}}$ generates \mathbb{G}/\mathbb{H} by a similar reasoning as used in Proposition 3.17. Consequently, $\tilde{\mathcal{Q}}$ is a barcode base for the quotient.

3.5 Homology of Persistence Modules

Consider a chain of tame persistence modules:

$$0 \longleftarrow V_0 \xleftarrow{d_1} V_1 \xleftarrow{d_2} V_2 \longleftarrow \dots \xleftarrow{d_n} V_n,$$

where each term has basis \mathcal{B}_j for $0 \leq j \leq n$. Then applying `image_kernel` we obtain bases \mathcal{I}_{j-1} and \mathcal{K}_j for the image and kernel of d_j for all $0 \leq j \leq n$. Proceeding as on the previous section, we send triples $((\mathcal{I}_j \mid \mathcal{K}_j), \mathcal{B}_j, ((\mathcal{I}_j)_{\mathcal{B}_j} \mid (\mathcal{K}_j)_{\mathcal{B}_j}))$ to `box_gauss_reduce` for all $0 \leq j \leq n$. This leads to bases \mathcal{Q}_j for the quotients $\text{Ker}(d_j)/\text{Im}(d_{j+1})$ for all $0 \leq j \leq n$.

4 A Review on the Mayer–Vietoris Spectral Sequence

In this section, we give an introduction to the Mayer–Vietoris spectral sequence. These ideas come mainly from [3, 20]. Here, we outline a minimal, self-contained explanation of the procedure. Also, this is used in Sect. 5. For simplicity we focus on ordinary homology over a field \mathbb{F} . Later, in Sect. 5, we go back to the case of persistent homology over a field.

4.1 The Mayer–Vietoris Long Exact Sequence

Consider a torus \mathbb{T}^2 covered by two cylinders U and V , as illustrated in Fig. 2. Naively, one could think that $H_n(\mathbb{T}^2) \cong H_n(U) \oplus H_n(V)$ for all $n \geq 0$. However, this does not hold in dimensions 0 and 2:

$$H_0(\mathbb{T}^2) = \mathbb{F} \not\cong \mathbb{F} \oplus \mathbb{F} = H_0(U) \oplus H_0(V), \quad H_2(\mathbb{T}^2) = \mathbb{F} \not\cong 0 = H_2(U) \oplus H_2(V).$$

To amend this, one has to look at the information given by the intersection $U \cap V$. This information comes as *identifications* and *new loops*. For example, U and V are connected through $U \cap V$. Also, the loop going around each cylinder U and V is identified in $U \cap V$. These identifications are performed by taking the quotient

$$I_n := \text{coker} (H_n(U \cap V) \rightarrow H_n(U) \oplus H_n(V))$$

for all $n \geq 0$. Where the previous morphism is the Čech differential $\delta_1^n : S_n(U \cap V) \rightarrow S_n(U) \oplus S_n(V)$. Additionally, the 1-loops in the intersection merge to the same loop when included in each cylinder U or V . This situation creates a 2-loop or “void”, see Fig. 2. Thus we have the n -loops detected by the kernel

$$L_n := \text{Ker} (H_{n-1}(U \cap V) \rightarrow H_{n-1}(U) \oplus H_{n-1}(V))$$

for all $n \geq 0$. Notice that n -loops are found by $n - 1$ information on the intersection. Putting all together, we have that

$$H_0(\mathbb{T}^2) \cong I_0 \cong \mathbb{F}, \quad H_1(\mathbb{T}^2) \cong I_1 \oplus L_1 \cong \mathbb{F} \oplus \mathbb{F}, \quad H_2(\mathbb{T}^2) \cong L_2 \cong \mathbb{F}.$$

On a more theoretical level, what we have presented here is commonly known as the Mayer–Vietoris Theorem. That is, $H_n(U \cup V)$ is a filtered object,

$$\{0\} = F^{-1}(H_n(U \cup V)) \subset F^0(H_n(U \cup V)) \subset F^1(H_n(U \cup V)) = H_n(U \cup V),$$

and there are expressions for the different ratios between consecutive filtrations,

$$F^0(H_n(U \cup V)) = I_n, \quad \frac{F^1(H_n(U \cup V))}{F^0(H_n(U \cup V))} = L_n.$$

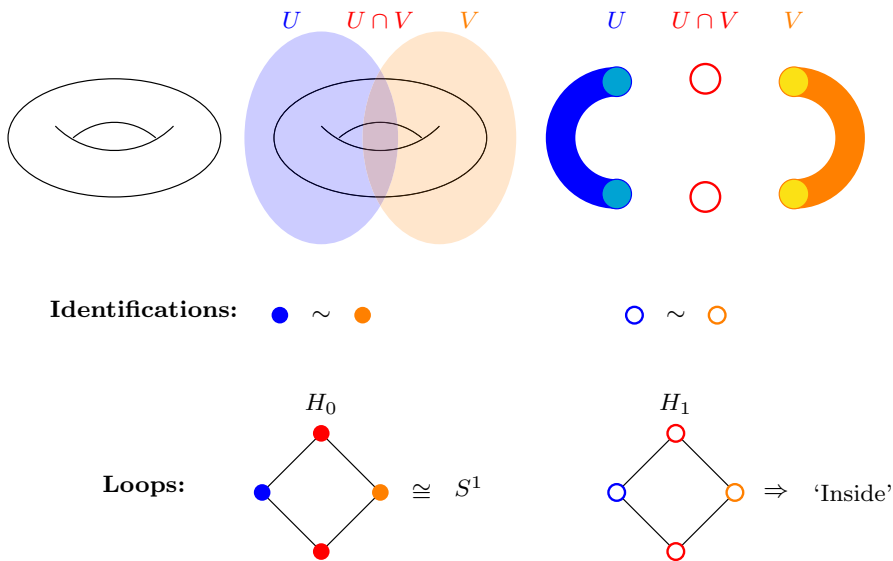


Fig. 2 Torus covered by a pair of cylinders U and V

In particular, as we are working with vector spaces, $H_n(U \cup V) \cong I_n \oplus L_n$ for all $n \geq 0$.

The above discussion gives rise to the *total* chain complex,

$$\text{Tot}_n(\mathcal{S}_*) = S_n(V) \oplus S_n(U) \oplus S_{n-1}(U \cap V),$$

with morphism $d_n^{\text{Tot}} = (d, d, d - \delta_1)$ for all $n \geq 0$. Notice that the first two morphisms do not change components, whereas the third encodes the “merging” of information. This last morphism is represented by red arrows on the diagram:

$$\begin{array}{ccccc}
 \text{Tot}_{n+1}(\mathcal{S}_*) & \cong & S_{n+1}(U) \oplus S_{n+1}(V) & \oplus & S_n(U \cap V) \\
 \downarrow d_{n+1}^{\text{Tot}} & & \downarrow d_{n+1} & \swarrow \delta_1 & \downarrow d_n \\
 \text{Tot}_n(\mathcal{S}_*) & \cong & S_n(U) \oplus S_n(V) & \oplus & S_{n-1}(U \cap V) \\
 \downarrow d_n^{\text{Tot}} & & \downarrow d_n & \swarrow \delta_1 & \downarrow d_{n-1} \\
 \text{Tot}_{n-1}(\mathcal{S}_*) & \cong & S_{n-1}(U) \oplus S_{n-1}(V) & \oplus & S_{n-2}(U \cap V)
 \end{array}$$

where the rectangle of red arrows is commutative. In particular, this implies that $d_n^{\text{Tot}} \circ d_{n+1}^{\text{Tot}} = 0$ for all $n \geq 0$. Computing the homology with respect to the total differentials and using the previous characterization of I_n and L_n , one obtains

$$H_n(\text{Tot}_*(\mathcal{S}_*)) \cong I_n \oplus L_n \cong H_n(K).$$

This result is generalized in Proposition 4.1.

4.2 The Mayer–Vietoris Spectral Sequence

Consider a simplicial complex K with a covering $\mathcal{U} = \{U_i\}_{i=0}^m$ by subcomplexes. We can extend the intuition from the previous subsection, by recalling the definition of the (n, \mathcal{U}) -Čech chain complex given on the preliminaries. Stacking all these sequences on top of each other, and also multiplying differentials in odd rows by -1 , we obtain a diagram:

$$\begin{array}{ccccccc}
 & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
 0 & \longleftarrow & S_2(K) & \xleftarrow{\delta_0} & \bigoplus_{\sigma \in \Delta_0^m} S_2(U_\sigma) & \xleftarrow{\delta_1} & \bigoplus_{\sigma \in \Delta_1^m} S_2(U_\sigma) & \xleftarrow{\delta_2} & \bigoplus_{\sigma \in \Delta_2^m} S_2(U_\sigma) & \longleftarrow \dots \\
 & & \downarrow d & & \downarrow d & & \downarrow d & & \downarrow d & \\
 0 & \longleftarrow & S_1(K) & \xleftarrow{-\delta_0} & \bigoplus_{\sigma \in \Delta_0^m} S_1(U_\sigma) & \xleftarrow{-\delta_1} & \bigoplus_{\sigma \in \Delta_1^m} S_1(U_\sigma) & \xleftarrow{-\delta_2} & \bigoplus_{\sigma \in \Delta_2^m} S_1(U_\sigma) & \longleftarrow \dots \\
 & & \downarrow d & & \downarrow d & & \downarrow d & & \downarrow d & \\
 0 & \longleftarrow & S_0(K) & \xleftarrow{\delta_0} & \bigoplus_{\sigma \in \Delta_0^m} S_0(U_\sigma) & \xleftarrow{\delta_1} & \bigoplus_{\sigma \in \Delta_1^m} S_0(U_\sigma) & \xleftarrow{\delta_2} & \bigoplus_{\sigma \in \Delta_2^m} S_0(U_\sigma) & \longleftarrow \dots \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
 & & 0 & & 0 & & 0 & & 0 &
 \end{array}$$

This leads to a *double complex* $(\mathcal{S}_{*,*}, \bar{\delta}, d)$ defined as

$$\mathcal{S}_{p,q} := \bigoplus_{\sigma \in \Delta_p^m} S_q(U_\sigma)$$

for all $p, q \geq 0$, and also $\mathcal{S}_{p,q} := 0$ otherwise. We denote $\bar{\delta} = (-1)^q \delta$, the Čech differential multiplied by a -1 on odd rows. The reason for this change of sign is because we want $\mathcal{S}_{*,*}$ to be a double complex, in the sense that the following equalities hold:

$$\bar{\delta} \circ \bar{\delta} = 0, \quad d \circ d = 0, \quad \bar{\delta} \circ d + d \circ \bar{\delta} = 0. \tag{5}$$

Since $\mathcal{S}_{*,*}$ is a double complex, we can study the associated chain complex $\mathcal{S}_*^{\text{Tot}}$, commonly known as the *total complex*. This is formed by taking the sums of anti-diagonals

$$\mathcal{S}_n^{\text{Tot}} := \bigoplus_{p+q=n} \mathcal{S}_{p,q}$$

for $n \geq 0$. The differentials on the total complex are defined by $d^{\text{Tot}} = d + \bar{\delta}$, which satisfy $d^{\text{Tot}} \circ d^{\text{Tot}} = 0$ from (5), see Fig. 3 for a depiction of this. Later, in Proposition 4.1, we prove that $H_n(K) \cong H_n(\mathcal{S}_*^{\text{Tot}})$ for all $n \geq 0$. The problem still

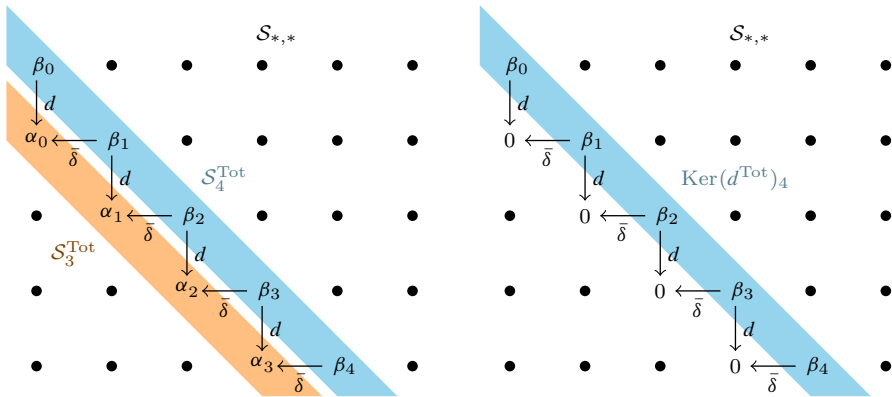


Fig. 3 $S_{*,*}$ represented as a lattice for convenience. On the left, the total complex S^{Tot} associated to $S_{*,*}$. Here $(\beta_0, \dots, \beta_4) \in S_4^{\text{Tot}}$ maps to $(\alpha_0, \dots, \alpha_3) \in S_3^{\text{Tot}}$, where $d(\beta_i) + \bar{\delta}(\beta_{i+1}) = \alpha_i$ for all $0 \leq i \leq 3$. On the right, the kernel $\text{Ker}(d^{\text{Tot}})_4$, where $d(\beta_i) + \bar{\delta}(\beta_{i+1}) = 0$ for all $0 \leq i \leq 3$

remains difficult, since computing $H_n(S_*^{\text{Tot}})$ directly might be even harder than computing $H_n(K)$. The key is that the Mayer–Vietoris spectral sequence allows us to break apart the calculation of $H_n(S_*^{\text{Tot}})$ into small, computable steps.

Let us start by computing the kernel $\text{Ker}(d_n^{\text{Tot}})$, which is depicted in Fig. 3. Recall that in this section we are working with vector spaces and linear maps. Let $s = (s_{k,n-k})_{0 \leq k \leq n} \in S_n^{\text{Tot}}$ be in $\text{Ker}(d_n^{\text{Tot}})$. Then, the equations $d(s_{k,n-k}) = -\bar{\delta}(s_{k+1,n-k-1})$ hold for all $0 \leq k < n$. This leads to subspaces $\text{GK}_{p,q} \subseteq S_{p,q}$ composed of elements $s_{p,q} \in S_{p,q}$ such that $d(s_{p,q}) = 0$ and such that there exists a sequence $s_{p-r,q+r} \in S_{p-r,q+r}$ with $d(s_{p-r,q+r}) = -\bar{\delta}(s_{p-r+1,q+r-1})$ for all $0 < r \leq p$. Notice that $\text{GK}_{p,q}$ is a subspace of $S_{p,q}$ since both d and $\bar{\delta}$ are linear. This is depicted in Fig. 4. There are (non-canonical) isomorphisms,

$$\text{Ker}(d_n^{\text{Tot}}) \cong \bigoplus_{p+q=n} \text{GK}_{p,q}. \tag{6}$$

It turns out that (6) only holds when we are working with vector spaces. Later, we work with a more general case where we have to solve nontrivial *extension problems*.

By (6), recovering all $\text{GK}_{p,q}$ leads to the kernel of d_*^{Tot} . However, computing $\text{GK}_{p,q}$ still requires a large set of equations to be checked. A step-by-step way of computing these is by adding one equation at a time. For this, we define the subspaces $\text{GZ}_{p,q}^r \subseteq S_{p,q}$ where we add the first r equations progressively. That is, we start setting $\text{GZ}_{p,q}^0 = S_{p,q}$. Then we define $\text{GZ}_{p,q}^1$ to be elements $s_{p,q} \in S_{p,q}$ such that $d(s_{p,q}) = 0$, or equivalently $\text{GZ}_{p,q}^1 = \text{Ker}(d)_{p,q}$. For $r \geq 2$, we define $\text{GZ}_{p,q}^r$ to be formed by elements $s_{p,q} \in \text{Ker}(d)_{p,q}$ such that there exists a sequence $s_{p-k,q+k} \in S_{p-k,q+k}$ with $d(s_{p-k,q+k}) = -\bar{\delta}(s_{p-k+1,q+k-1})$ for all $1 \leq k < r$. Altogether,

$$\text{GK}_{p,q} = \text{GZ}_{p,q}^{p+1} \subseteq \text{GZ}_{p,q}^p \subseteq \dots \subseteq \text{GZ}_{p,q}^0 = S_{p,q}$$

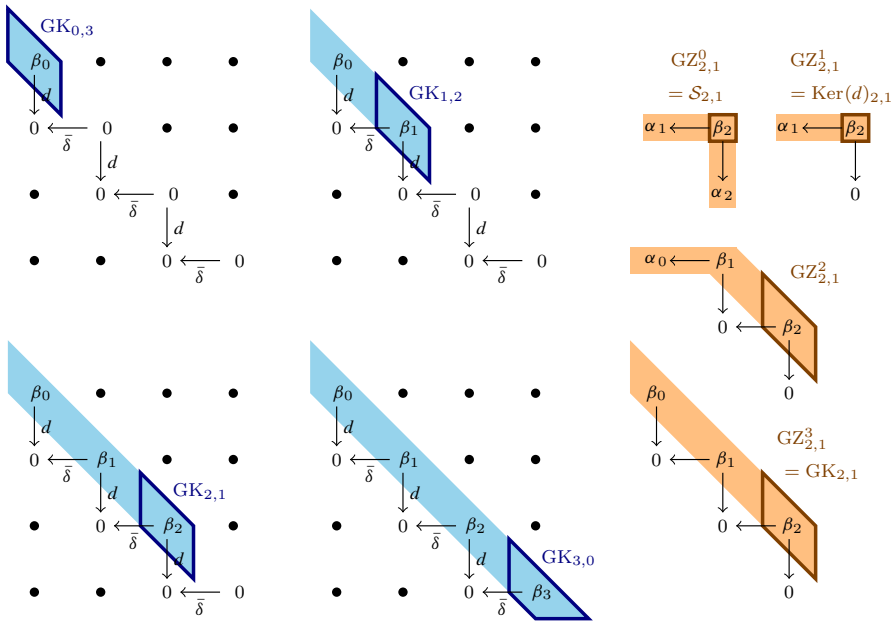


Fig. 4 On the left, in cyan the four direct summands of $\text{Ker}(d^{\text{Tot}})_4$. The corresponding $\text{GK}_{r,3-r}$ are framed to indicate that these are subspaces of $\mathcal{S}_{r,3-r}$ for all $0 \leq r \leq 3$. On the right, in orange the subspaces $\text{GZ}_{2,1}^r$, eventually shrinking to $\text{GK}_{2,1}$. For convenience, we denote $\alpha_2 = d(\beta_2)$, $\alpha_1 = \delta(\beta_2)$ and $\alpha_0 = \delta(\beta_1)$

for all $p, q \geq 0$. For intuition see Fig. 4, and also Fig. 6 for a depiction of $\text{GZ}_{3,1}^2$ on a lattice. We also write $\text{GZ}_{p,q}^r = \text{Ker}(d) \cap (\delta^{-1} \circ d)^{r-1}(\mathcal{S}_{p-r+1,q+r-1})$ for all $r \geq 1$, where by $(\delta^{-1} \circ d)^r$ we denote composing r times $\delta^{-1} \circ d$. In particular, since $\text{GZ}_{p,q}^r = \text{GZ}_{p,q}^{r+1}$ for all $r \geq p+1$, we use the convention $\text{GZ}_{p,q}^\infty := \text{GZ}_{p,q}^{p+1} = \text{GK}_{p,q}$.

Now, we explain the notation $\text{GK}_{p,q}$ and the isomorphism (6). We start defining a vertical filtration F^* on $\mathcal{S}_{*,*}$ by the following subcomplexes for all $r \geq 0$:

$$F^r(\mathcal{S}_{*,*})_{p,q} := \begin{cases} \mathcal{S}_{p,q} & \text{whenever } p \leq r, \\ 0 & \text{otherwise.} \end{cases}$$

This is an increasing filtration, i.e., $F^r(\mathcal{S}_{*,*}) \subseteq F^{r+1}(\mathcal{S}_{*,*})$ for all $r \geq 0$. Additionally, there are isomorphisms $F^r(\mathcal{S}_{*,*})/F^{r-1}(\mathcal{S}_{*,*}) \cong \mathcal{S}_{r,*}$ for all $r \geq 0$. See Fig. 5 for a depiction of F^* . Also, F^* respects the differentials from $\mathcal{S}_{*,*}$ in the sense that $d(F^r(\mathcal{S}_{*,*})) \subseteq F^r(\mathcal{S}_{*,*})$ and $\delta(F^r(\mathcal{S}_{*,*})) \subseteq F^r(\mathcal{S}_{*,*})$ for all $r \geq 0$. Consequently, F^* filters the total complex $\mathcal{S}_*^{\text{Tot}}$, respecting its differential d^{Tot} . That is, $\mathcal{S}_n^{\text{Tot}}$ is filtered by subcomplexes, $F^r \mathcal{S}_n^{\text{Tot}} := \bigoplus_{p+q=n}^{p \leq r} \mathcal{S}_{p,q}$, for all $r \geq 0$.

In particular, $\text{Ker}(d^{\text{Tot}})$ also inherits the filtration F^* , where $F^r \text{Ker}(d^{\text{Tot}})_n = F^r \mathcal{S}_n^{\text{Tot}} \cap \text{Ker}(d^{\text{Tot}})_n$ for all $r \geq 0$. We define the associated modules of $\text{Ker}(d^{\text{Tot}})_n$ to be the quotients $G^p \text{Ker}(d^{\text{Tot}})_n = F^p \text{Ker}(d^{\text{Tot}})_n / F^{p-1} \text{Ker}(d^{\text{Tot}})_n$, which can be checked to be isomorphic with $\text{GK}_{p,q}$ for all $p+q=n$. This follows by considering

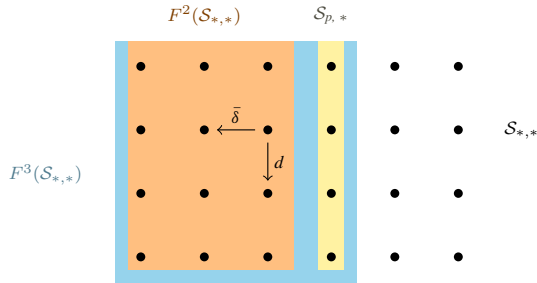


Fig. 5 Note that $F^3(S_{*,*})/F^2(S_{*,*}) \cong S_{3,*}$. Also notice that $\bar{\delta}$ and d respect the vertical filtration F^*

morphisms

$$\begin{aligned} G^p \text{Ker}(d^{\text{Tot}})_n &\longrightarrow \text{GK}_{p,q}, \\ [(s_{0,n}, s_{1,n-1}, \dots, s_{p,q}, 0, \dots, 0)] &\longrightarrow s_{p,q}, \end{aligned} \tag{7}$$

which are well defined since $s_{p,q}$ does not change for representatives of the same class. In fact, this morphism is injective since two classes with the same image will be equal by definition of $G^p \text{Ker}(d^{\text{Tot}})_n$. On the other hand, the definition of $\text{GK}_{p,q}$ ensures surjectivity. In particular, as we are working with vector spaces, we have that

$$\text{Ker}(d_n^{\text{Tot}}) \cong \bigoplus_{p+q=n} G^p \text{Ker}(d^{\text{Tot}})_n \cong \bigoplus_{p+q=n} \text{GK}_{p,q},$$

which justifies isomorphism (6).

Next, we explain the notation $\text{GZ}^r_{p,q}$. We introduce the objects

$$Z^r_{p,q} := \{z \in F^p \mathcal{S}^{\text{Tot}}_{p+q} : d^{\text{Tot}}(z) \in F^{p-r} \mathcal{S}^{\text{Tot}}_{p+q-1}\}$$

for all $r \geq 0$. We can think of these as kernels of d^{Tot} up to some previous filtration. Then, by definition, we have that $Z^0_{p,q} = F^p \mathcal{S}^{\text{Tot}}_{p+q}$ and $Z^{p+1}_{p,q} = Z^\infty_{p,q} = F^p \text{Ker}(d^{\text{Tot}}_{p+q})$. Using a morphism analogous to (7), one can check that the quotients $Z^{r+1}_{p,q}/Z^r_{p-1,q+1}$ are isomorphic to $\text{GZ}^{r+1}_{p,q}$ for all $p+q = n$. This is depicted in Fig. 6. Thus, computing these quotients increasing $r \geq 0$ leads to $\text{Ker}(d^{\text{Tot}})$. With a little more work, we can do the same for computing the homology.

The *Mayer–Vietoris spectral sequence* leads to $H_n(\mathcal{S}_*^{\text{Tot}})$ after a series of small, computable steps. This is done similarly as we did before for computing $\text{Ker}(d^{\text{Tot}})$. However, in this case we need to take quotients by the images of d^{Tot} . First, notice that the vertical filtration F^* transfers to homology $H_n(\mathcal{S}_*^{\text{Tot}})$ by the inclusions $F^p \mathcal{S}_*^{\text{Tot}} \subseteq \mathcal{S}_*^{\text{Tot}}$ for all $p \geq 0$. That is, we have filtered sets

$$F^p H_n(\mathcal{S}_*^{\text{Tot}}) := \text{Im}\left(H_n(F^p \mathcal{S}_*^{\text{Tot}}) \longrightarrow H_n(\mathcal{S}_*^{\text{Tot}})\right),$$

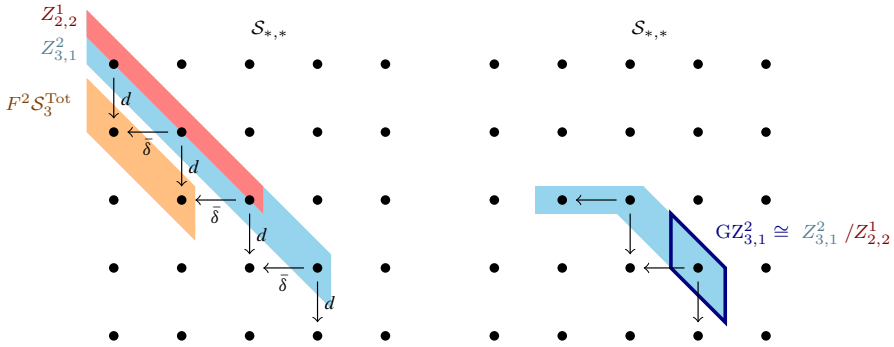


Fig. 6 On the left the sets $Z_{3,1}^2$ and $Z_{2,2}^1$. On the right their respective quotient $GZ_{3,1}^2$

which induce a filtration on $H_n(S_*^{\text{Tot}})$. For this filtration, the associated modules are defined by the quotients $G^r H_n(S_*^{\text{Tot}}) = F^r H_n(S_*^{\text{Tot}}) / F^{r-1} H_n(S_*^{\text{Tot}})$ for all $r \geq 0$. As we are working over a field, we recover the homology by taking direct sums:

$$H_n(S_*^{\text{Tot}}) \cong \bigoplus_{r=0}^n G^r H_n(S_*^{\text{Tot}}). \tag{8}$$

Previously, we defined the sets $Z_{p,q}^r$ which are kernels “up to filtration”. In an analogous way, we define boundaries “up to filtration” by setting³

$$B_{p,q}^r := \{d^{\text{Tot}}(c) \in F^p S_{p+q}^{\text{Tot}} : c \in F^{p+r} S_{p+q+1}^{\text{Tot}}\}$$

for all $r \geq 0$ and $p, q \geq 0$. These are images of d^{Tot} coming from a previous filtration. In particular, the equalities $d^{\text{Tot}}(Z_{p,q}^r) = B_{p-r,q+r-1}^r$ and $d^{\text{Tot}}(B_{p,q}^r) = 0$ hold. Additionally, for all $p, q \geq 0$, there is a sequence of inclusions,

$$B_{p,q}^0 \subset B_{p,q}^1 \subset \dots \subset B_{p,q}^{q+1} = B_{p,q}^\infty \subset Z_{p,q}^\infty = Z_{p,q}^p \subset \dots \subset Z_{p,q}^1 \subset Z_{p,q}^0.$$

Hence, we define the *first page* of the spectral sequence as the quotient

$$E_{p,q}^1 := \frac{Z_{p,q}^1}{Z_{p-1,q+1}^0 + B_{p,q}^0} \cong \frac{GZ_{p,q}^1}{\text{Im}(B_{p,q}^0 \rightarrow GZ_{p,q}^1)},$$

for all $p, q \geq 0$. Recall that $\text{Ker}(d)_{p,q} = GZ_{p,q}^1 = Z_{p,q}^1 / Z_{p-1,q+1}^0$ and also one can see that $\text{Im}(B_{p,q}^0 \rightarrow GZ_{p,q}^1)$ is isomorphic to $\text{Im}(d)_{p,q}$. Then we deduce that $E_{p,q}^1 \cong H_q(S_{p,*}, d)$. On this page d^{Tot} induces differentials $d^1 : E_{p,q}^1 \rightarrow E_{p-1,q}^1$.

³ Here we have adopted the definition of $Z_{p,q}^r$ and $B_{p,q}^r$ that one can find in [20]. Other sources such as [3] and [19] use the same notation for other terms.

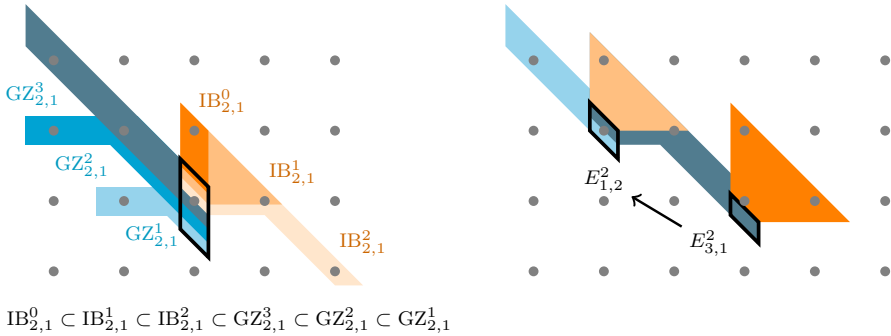


Fig. 7 On the left, the different subspaces on $S_{2,1}$. Here $IB_{2,1}^r = \text{Im}(B_{2,1}^r \rightarrow GZ_{2,1}^{r+1})$, for all $0 \leq r \leq 2$. The framed region represents $S_{2,1}$. Brighter colours represent bigger regions than darker colours. Note that blue and orange colours have been assigned to $GZ_{2,1}^*$ and $IB_{2,1}^*$ respectively. On the right, the morphism $d^2: E_{3,1}^2 \rightarrow E_{1,2}^2$ on the second page. The two framed regions represent the codomain and domain of d^2 , these have been assigned brighter and darker colours, respectively

I.e., noticing that $d^{\text{Tot}}(Z_{p,q}^1) = B_{p-1,q}^1 \subset Z_{p-1,q}^1$ and also $d^{\text{Tot}}(Z_{p-1,q+1}^0 + B_{p,q}^0) = d^{\text{Tot}}(Z_{p-1,q+1}^0) + 0 = B_{p-1,q}^0$ we have that $d^1: E_{p,q}^1 \rightarrow E_{p-1,q}^1$ is well defined.

Next, we compute the second page. First, notice that

$$\text{Ker}(d^1) = \frac{Z_{p,q}^2}{Z_{p,q}^2 \cap (Z_{p-1,q+1}^0 + B_{p,q}^0)} = \frac{Z_{p,q}^2}{Z_{p-1,q+1}^1 + B_{p,q}^0}, \quad \text{Im}(d^1) = \frac{B_{p,q}^1}{B_{p,q}^0}.$$

Since $d^{\text{Tot}} \circ d^{\text{Tot}} = 0$ we also have $d^1 \circ d^1 = 0$ and, in particular, $\text{Im}(d^1) \subseteq \text{Ker}(d^1)$. Then, the second page terms are given by

$$E_{p,q}^2 := H_{p,q}(E_{*,*}^1, d^1) = \frac{\text{Ker}(d^1)}{\text{Im}(d^1)} = \frac{Z_{p,q}^2}{Z_{p-1,q+1}^1 + B_{p,q}^1}.$$

The second page has differential d^2 induced by the total complex differential d^{Tot} . Figure 7 illustrates this principle. Doing the same for all pages we obtain the definition of the r -page terms

$$E_{p,q}^r := H_{p,q}(E_{*,*}^{r-1}, d^{r-1}) = \frac{Z_{p,q}^r}{Z_{p-1,q+1}^{r-1} + B_{p,q}^{r-1}},$$

for all $r \geq 2$. Of course, we can express alternatively the r -page terms as

$$E_{p,q}^r := \frac{GZ_{p,q}^r}{\text{Im}(B_{p,q}^{r-1} \rightarrow GZ_{p,q}^r)}.$$

Thus, the ∞ -page is

$$E_{p,q}^\infty = \frac{Z_{p,q}^\infty}{Z_{p-1,q+1}^\infty + B_{p,q}^\infty} \cong \frac{\text{GK}_{p,q}}{\text{Im}(B_{p,q}^\infty \rightarrow \text{GK}_{p,q})}.$$

Then, for $n = p + q$, one has that $G^p H_n(\mathcal{S}_*^{\text{Tot}})$ is isomorphic to

$$\begin{aligned} \frac{F^p H_n(\mathcal{S}_*^{\text{Tot}})}{F^{p-1} H_n(\mathcal{S}_*^{\text{Tot}})} &= \frac{\text{Im}(H_n(F^p \mathcal{S}_*^{\text{Tot}}) \rightarrow H_n(\mathcal{S}_*^{\text{Tot}}))}{\text{Im}(H_n(F^{p-1} \mathcal{S}_*^{\text{Tot}}) \rightarrow H_n(\mathcal{S}_*^{\text{Tot}}))} \\ &\cong \frac{Z_{p,q}^\infty / B_{p,q}^\infty}{Z_{p-1,q+1}^\infty / B_{p-1,q+1}^\infty} \cong E_{p,q}^\infty, \end{aligned}$$

since $B_{p-1,q+1}^\infty \subseteq B_{p,q}^\infty$. Therefore, computing the spectral sequence is a way of approximating the associated module $G^p H_n(\mathcal{S}_*^{\text{Tot}})$. By (8), adding up these leads to $H_n(\mathcal{S}_*^{\text{Tot}})$. Also, since $E_{p,q}^\infty \cong G^p H_n(\mathcal{S}_*^{\text{Tot}})$, we say that $E_{p,q}^*$ converges to $H_n(\mathcal{S}^{\text{Tot}})$ and we denote this as $E_{p,q}^* \Rightarrow H_n(\mathcal{S}^{\text{Tot}})$. As the rows from $\mathcal{S}_{p,q}$ are exact, the following result is standard; see for example [3, Prop. 8.8] for a similar proof.

Proposition 4.1 $H_n(\mathcal{S}_*^{\text{Tot}}) \cong H_n(K)$.

Proof Consider the horizontal filtration F_H by $F_H^r \mathcal{S}_n^{\text{Tot}} := \bigoplus_{p+q=n}^{q \leq r} \mathcal{S}_{p,q}$. This whole section can be adapted to this filtration and one obtains a corresponding spectral sequence ${}_H E_{p,q}^* \Rightarrow H_n(\mathcal{S}^{\text{Tot}})$. As the Čech sequence (3) is exact, ${}_H E_{p,q}^*$ collapses on the first page, where the only nontrivial terms are ${}_H E_{0,n}^1 \simeq H_n(K)$ for all $n \geq 0$. This implies the claim. See [20, Thm. 2.15] for a general result on the spectral sequence of a double complex. \square

Therefore, using Proposition 4.1, we have that $E_{p,q}^* \Rightarrow H_n(K)$. In particular, since we are working with vector spaces, $H_n(K) \cong \bigoplus_{p+q=n} E_{p,q}^\infty$ for all $n \geq 0$. Throughout the following section, we adapt these results to the category of persistence modules.

5 Persistent Mayer–Vietoris

One can translate the method from Sect. 4 to **PMod**. The reason for this is that **PMod** is an abelian category, since **Vect** is an abelian category and **R** is a small category. The theory of spectral sequences can be developed for arbitrary abelian categories. For an introduction to this, see [30, Chap. 5].

Consider a filtered simplicial complex, K , together with a cover of K by filtered subcomplexes, $\mathcal{U} = \{U_i\}_{i \in I}$, so that $K = \bigcup_{i \in I} U_i$. Then, the persistence Mayer–Vietoris spectral sequence is given by

$$E_{p,q}^1 = \bigoplus_{\sigma \in \Delta_p^m} \text{PH}_q(U_\sigma) \Rightarrow \text{PH}_n(K),$$

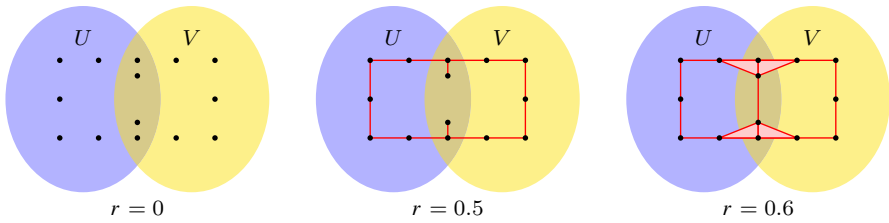


Fig. 8 As the radius increases, more edges are added. At radius $r = 0.5$ a circle will be across the two covers U and V . Later on, at radius $r = 0.6$ this circle will be split into two

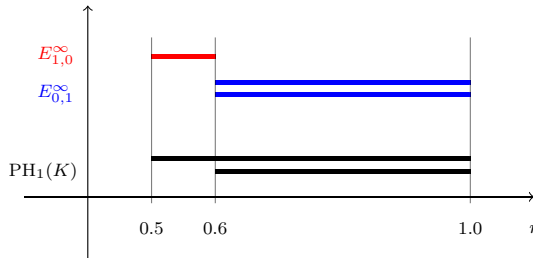


Fig. 9 Barcode on associated module

where $p + q = n$. However, unlike the case of vector spaces, we might have that $\bigoplus_{p+q=n} E_{p,q}^{\infty} \not\cong \text{PH}_n(K)$. All that we know is that $E_{p,q}^{\infty} \cong G^p \text{PH}_{p+q}(K)$ for all $p, q \geq 0$. In the literature, recovering $\text{PH}_n(K)$ from the terms $G^p \text{PH}_{p+q}(K)$ is known as the *extension problem*, which we solve in Sect. 5.1. Furthermore, we even obtain more information; as pointed out in [31], the knowledge of which subset $J \subset I$ detects a feature from $\text{PH}_n(K)$ can potentially add insight into the information given by ordinary persistent homology. The following example illustrates this.

Example 5.1 Consider the case of a point cloud X covered by two open sets as in Fig. 8. From Sects. 3 and 4, we know how to compute the ∞ -page $(E_{*,*}^{\infty})^r$ associated to any value $r \in \mathbf{R}$. In particular, when we take $r = 0.5$, then the combination of U and V detects a 1-cycle. On the other hand, when $r = 0.6$ this cycle splits into two smaller cycles which are detected by U and V individually. Notice that if we want to come up with a persistent Mayer–Vietoris method then we need to be able to track this behaviour. That is, we need to know how cycles develop as r increases. In particular, the barcode from $\text{PH}_1(X)$ is broken down into $E_{1,0}^{\infty} \cong \mathbb{F}_{[0.5,0.6)}$ and also $E_{0,1}^{\infty} \cong \mathbb{F}_{[0.6,1.0)} \oplus \mathbb{F}_{[0.6,1.0)}$, see Fig. 9.

Figure 10 illustrates a filtered complex covered by three regions where, as in Example 5.1, there is a nontrivial extension problem.

5.1 The Extension Problem

Recall the definition of the total complex, vertical filtrations and associated modules from Sect. 4. Through this section we study the extension problem, that is, we recover $H_n(\mathcal{S}_*^{\text{Tot}})$ from the associated modules $G^p(H_n(\mathcal{S}_*^{\text{Tot}}))$. Also, we assume that

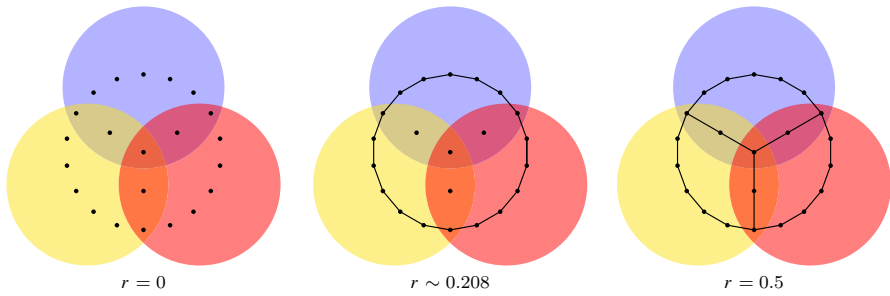


Fig. 10 A one loop is detected at value $r \sim 0.208$ which goes through three covers. Later, at radius $r = 0.5$, this loop splits into three loops, each included in one of the three covers

the spectral sequence collapses after a finite number of pages. Consider the persistence module

$$\mathbb{V} = \mathbb{V}(n) := H_n(S_*^{\text{Tot}}),$$

together with the corresponding filtration

$$0 = F^{-1}\mathbb{V} \subset F^0\mathbb{V} \subset \dots \subset F^n\mathbb{V} = \mathbb{V}.$$

We define the *associated modules* of (\mathbb{V}, F^*) as the quotients $\mathbb{G}^k = F^k\mathbb{V}/F^{k-1}\mathbb{V}$ for all $0 \leq k \leq n$. This gives rise to short exact sequences,

$$0 \longrightarrow F^{k-1}\mathbb{V} \xrightarrow{\iota} F^k\mathbb{V} \xrightarrow{p^k} \mathbb{G}^k \longrightarrow 0, \tag{9}$$

for all $0 \leq k \leq n$. On the sequences (9) we consider successive extension problems where we know the first and last term and want to know the middle term for $0 \leq k \leq n$. This leads to \mathbb{V} ; however, in this work we obtain directly a persistence module isomorphic to \mathbb{V} (see Proposition 5.2). Adding up all associated modules we obtain a persistence module $\mathbb{G} := \bigoplus_{i=0}^n \mathbb{G}^i$ with an additional filtration given by $F^k\mathbb{G} = \bigoplus_{i=0}^k \mathbb{G}^i$ for all $0 \leq k \leq n$. Since $\mathbb{G}^k \cong E_{k,n-k}^\infty$ for all $0 \leq k \leq n$, a spectral sequence algorithm leads to a barcode basis for \mathbb{G} . We formulate the extension problem as computing a basis for \mathbb{V} from the obtained basis \mathcal{G} of \mathbb{G} .

To start, notice that for each $r \in \mathbf{R}$ the sequence (9) splits, leading to morphisms

$$\mathcal{F}^k(r) : \mathbb{G}^k(r) \rightarrow F^k\mathbb{V}(r), \tag{10}$$

such that $p^k(r) \circ \mathcal{F}^k(r) = \text{Id}_{\mathbb{G}^k(r)}$ for all $0 \leq k \leq n$. In particular, $\mathcal{F}^k(r)$ is injective for all $0 \leq k \leq n$. On the other hand, for any class $[\beta_k]_{k,n-k}^\infty$ of $E_{k,n-k}^\infty(r)$ with representative $\beta_k \in E_{k,n-k}^0(r)$, since $\beta_k \in \text{GK}_{k,n-k}(r)$, we have that $d(\beta_k) = 0$ and there exists a sequence of $\beta_i \in S_{i,n-i}(r)$ such that $d(\beta_i) = -\bar{\delta}(\beta_{i+1})$ for all

$0 \leq i < k$. The choice of this sequence determines $\mathcal{F}^k(r)$, so that

$$\mathcal{F}^k(r)([\beta_k(r)]_{k,n-k}^\infty) = [(\beta_0(r), \beta_1(r), \dots, \beta_k(r), 0, \dots, 0)]_n^{\text{Tot}},$$

where $[\cdot]_n^{\text{Tot}}$ denotes the n -homology class of the total complex. Notice that if we already computed \mathcal{G} from the Mayer–Vietoris spectral sequence, then there is no need to do any extra computations to obtain these morphisms $\mathcal{F}^k(r)$. All we need to do is to store our previous results. Adding over all $0 \leq k \leq n$ we obtain the isomorphism $\mathcal{F}(r) = \bigoplus_{k=0}^n \mathcal{F}^k(r): \bigoplus_{k=0}^n \mathbb{G}^k(r) \rightarrow \mathbb{V}(r)$. This last morphism is an isomorphism since all its summands are injective, their images have mutual trivial intersection, and the dimensions of the domain and codomain coincide.

Recall that \mathbb{G} has induced morphisms $\mathbb{G}(r \leq s)$ from $\mathbb{V}(r \leq s)$ for all values $r \leq s$ in \mathbf{R} . Given a basis \mathcal{G} for \mathbb{G} , we would like to compute a basis \mathcal{B} for \mathbb{V} from this information. Notice that this is not a straightforward problem since (10) does not imply that one has an isomorphism $\mathcal{F}: \mathbb{G} \rightarrow \mathbb{V}$. A point to start is to define the image along each generator in \mathcal{G} . That is, for each barcode generator $g_i \sim [a_i, b_i]$ in \mathcal{G} , we choose an image at the start $\mathcal{F}(a_i)(g_i(a_i))$. Then, we set $\mathcal{F}(r)(g_i(r)) := \mathbb{V}(a_i < r) \circ \mathcal{F}(a_i)(g_i(a_i))$ for all $a_i < r < b_i$. This leads to commutativity of \mathcal{F} along each generator g_i . Nevertheless this is still far from even defining a morphism $\mathcal{F}: \mathbb{G} \rightarrow \mathbb{V}$.

The solution to the problem above is to define a new persistence module $\tilde{\mathbb{G}}$. We define $\tilde{\mathbb{G}}(s) := \mathbb{G}(s)$ for all $s \in \mathbf{R}$. Then, if $\mathcal{G} = \{g_i\}_{i=1}^{|\mathcal{G}|}$ is a barcode basis for \mathbb{G} , by Proposition 3.2, $\mathcal{G}^s(1_{\mathbb{F}})$ will be a basis of $\tilde{\mathbb{G}}(s)$ for all $s \in \mathbf{R}$. Now, given $g_i \sim [a_i, b_i]$ a generator in \mathcal{G} , we define the morphism $\tilde{\mathbb{G}}(r \leq s)$ by the recursive formula

$$\tilde{\mathbb{G}}(r \leq s)(g_i(r)(1_{\mathbb{F}})) := \begin{cases} \sum_{j=1}^{|\mathcal{G}|} c_{i,j} \tilde{\mathbb{G}}(b_i \leq s)(g_j(b_i)(1_{\mathbb{F}})) & \text{if } r \in [a_i, b_i], b_i \leq s, \\ g_i(s)(1_{\mathbb{F}}) & \text{if } r, s \in [a_i, b_i], \\ 0 & \text{otherwise,} \end{cases}$$

where $c_{i,j} \in \mathbb{F}$ for all $1 \leq i, j \leq |\mathcal{G}|$. We want to define $c_{i,j}$ in such a way that $\tilde{\mathbb{G}}$ is isomorphic to \mathbb{V} . For this we impose the commutativity condition

$$\tilde{\mathbb{G}}(a_i \leq b_i)(g_i(a_i)(1_{\mathbb{F}})) = \mathcal{F}(b_i)^{-1} \circ \mathbb{V}(a_i \leq b_i) \circ \mathcal{F}(a_i)(g_i(a_i)(1_{\mathbb{F}})),$$

which leads to the equation

$$\sum_{j=1}^{|\mathcal{G}|} c_{i,j} g_j(b_i)(1_{\mathbb{F}}) = \mathcal{F}(b_i)^{-1} \circ \mathbb{V}(a_i \leq b_i) \circ \mathcal{F}(a_i)(g_i(a_i)(1_{\mathbb{F}})). \tag{11}$$

This determines uniquely the coefficients $c_{i,j}$ for all $1 \leq i, j \leq |\mathcal{G}|$. Notice that $\tilde{\mathbb{G}}$ respects the filtration on \mathbb{V} , since the right hand side in (11) is a composition of filtration preserving morphisms. In particular, if $g_i \in \text{PVect}(F^k \tilde{\mathbb{G}})$, then $c_{i,j} = 0$ for all $1 \leq j \leq |\mathcal{G}|$ such that $g_j \notin \text{PVect}(F^k \tilde{\mathbb{G}})$.

Fix a generator $g_i \in \mathcal{G}$ such that $g_i \in \text{PVect}(\mathbb{G}^k)$ and such that $g_i \sim [a_i, b_i]$. Let us calculate the coefficients $c_{i,j}$. Suppose that we have a representative $\tilde{g}_j = (\beta_0^j, \beta_1^j, \dots, \beta_k^j, 0, \dots, 0) \in \mathcal{S}_n^{\text{Tot}}$ for each generator $g_j \in \mathcal{G}$, with $g_j = [\beta_k^j]_{k,n-k}^\infty$. Also, for all $0 \leq q \leq n$ we define the subset $I^q \subseteq \{1, \dots, |\mathcal{G}|\}$ of indices $1 \leq j \leq |\mathcal{G}|$ such that $g_j \in \text{PVect}(\mathbb{G}^q)$. Then, the coefficients $c_{i,j}$ for $j \in I^k \setminus \{i\}$ are determined by the following equality in $\mathbb{G}^k(b_i)$ (where we use p^k from the sequence (9))

$$p^k(b_i)([\tilde{g}_i(b_i)(1_{\mathbb{F}})]_n^{\text{Tot}}) = \sum_{j \in I^k \setminus \{i\}} c_{i,j} g_j(b_i)(1_{\mathbb{F}}).$$

Thus, we have

$$p^k(b_i) \left(\left[\tilde{g}_i(b_i)(1_{\mathbb{F}}) - \sum_{j \in I^k \setminus \{i\}} c_{i,j} \tilde{g}_j(b_i)(1_{\mathbb{F}}) \right]_n^{\text{Tot}} \right) = 0.$$

Hence, by (9), there must exist some chain $\Gamma \in \mathcal{S}_{n+1}^{\text{Tot}}(b_i)$ such that

$$\tilde{g}_i(b_i)(1_{\mathbb{F}}) - \sum_{j \in I^k \setminus \{i\}} c_{i,j} \tilde{g}_j(b_i)(1_{\mathbb{F}}) - d^{\text{Tot}} \Gamma \tag{12}$$

is contained in $F^{k-1} \mathcal{S}_n^{\text{Tot}}(b_i)$. How do we compute Γ ? We start by searching for the first page $r \geq 2$ such that

$$\left[\beta_k^i(b_i)(1_{\mathbb{F}}) - \sum_{j \in I^k \setminus \{i\}} c_{i,j} \beta_k^j(b_i)(1_{\mathbb{F}}) \right]_{k,n-k}^r = 0, \tag{13}$$

where $[\cdot]_{k,n-k}^r$ denotes the class in the r -page in position $(k, n - k)$. Notice that this r must exist since we assumed that (13) vanishes at the ∞ -page. Consequently, there exists $\Gamma_{k+r-1} \in E_{k+r-1,n-k-r+2}^{r-1}(b_i)$ such that

$$\left[\beta_k^i(b_i)(1_{\mathbb{F}}) - \sum_{j \in I^k \setminus \{i\}} c_{i,j} \beta_k^j(b_i)(1_{\mathbb{F}}) \right]_{k,n-k}^{r-1} - d^{r-1}(\Gamma_{k+r-1}) = 0$$

on $E_{k,n-k}^{r-1}(b_i)$. Repeating for all pages leads to $\Gamma_{k+t} \in E_{k+t,n-k-t+1}^t(b_i)$ for all $0 \leq t \leq r - 1$, such that

$$\beta_k^i(b_i)(1_{\mathbb{F}}) - \sum_{j \in I^k \setminus \{i\}} c_{i,j} \beta_k^j(b_i)(1_{\mathbb{F}}) - \sum_{t=0}^{r-1} d^t(\widetilde{\Gamma_{k+t}}) = 0, \tag{14}$$

where $d^l(\widetilde{\Gamma}_{k+t}) \in \mathcal{S}_{k,n-k}(b_i)$ is a representative for the class $d^l(\Gamma_{k+t}) \in E_{k,n-k}^l(b_i)$. Notice that (14) holds independently of the representatives, since if we changed some term, then the other representatives would adjust to the change. In particular, we have that the k component of (12) vanishes, whereas the $k - 1$ component will be equal to

$$\beta_{k-1}^i(b_i)(1_{\mathbb{F}}) - \sum_{j \in I^k \setminus \{i\}} c_{i,j} \beta_{k-1}^j(b_i)(1_{\mathbb{F}}) - \bar{\delta}(\Gamma_k).$$

Next we proceed to find coefficients $c_{i,j} \in \mathbb{F}$ so that in $\mathbb{G}^{k-1}(b_i)$ we get the equality

$$\left[\beta_{k-1}^i(b_i)(1_{\mathbb{F}}) - \sum_{j \in I^k \setminus \{i\}} c_{i,j} \beta_{k-1}^j(b_i)(1_{\mathbb{F}}) - \bar{\delta}(\Gamma_k) \right]_{k-1,n-k+1}^{\infty} = \sum_{j \in I^{k-1}} c_{i,j} g_j(b_i)(1_{\mathbb{F}}).$$

Then we proceed as we did on \mathbb{G}^k . Doing this for all parameters $0 \leq r \leq k$, there are coefficients $c_{i,j} \in \mathbb{F}$, and an element $\tilde{\Gamma} \in \mathcal{S}_{n+1}^{\text{Tot}}(b_i)$ so that

$$\tilde{g}_i(b_i)(1_{\mathbb{F}}) = \sum_{0 \leq r \leq k} \sum_{j \in I^r} c_{i,j} \tilde{g}_j(b_i)(1_{\mathbb{F}}) + d^{\text{Tot}} \tilde{\Gamma}.$$

Thus, recalling that $c_{i,j} = 0$ for all $g_j \notin \text{PVect}(F^k \mathbb{G})$, we have

$$\tilde{\mathbb{G}}(a_i \leq b_i)(g_i(a_i)(1_{\mathbb{F}})) = \sum_{0 \leq j \leq |\mathcal{G}|} c_{i,j} g_j(b_i)(1_{\mathbb{F}}).$$

Proposition 5.2 $\tilde{\mathbb{G}} \cong \mathbb{V}$.

Proof Since each $\mathcal{F}(s)$ is an isomorphism, and also we have commutative squares:

$$\begin{array}{ccc} \tilde{\mathbb{G}}(r) & \xrightarrow{\tilde{\mathbb{G}}(r \leq s)} & \tilde{\mathbb{G}}(s) \\ \mathcal{F}(r) \downarrow & & \downarrow \mathcal{F}(s) \\ \mathbb{V}(r) & \xrightarrow{\mathbb{V}(r \leq s)} & \mathbb{V}(s) \end{array}$$

for all $r \leq s$, then \mathcal{F} must be an isomorphism of persistence modules. □

This gives $\tilde{\mathbb{G}} \cong \mathbb{V}$, but we still need to compute a barcode basis. In fact, this can be done by considering a quotient. Define $\mathbb{A} \simeq \bigoplus_{g_i \in \mathcal{G}} \mathbb{F}_{[a_i, \infty)}$ where $g_i \sim [a_i, b_i)$ for all $g_i \in \mathcal{G}$; here the $\mathcal{A} = \{\alpha_i\}_{1 \leq i \leq |\mathcal{G}|}$ denotes the canonical base for \mathbb{A} . Consider the coefficients $c_{i,j}$ for $1 \leq i, j \leq |\mathcal{G}|$ from the construction of $\tilde{\mathbb{G}}$ and define the sets of

indices $\mathcal{S}_i = \{1 \leq j \leq |\mathcal{G}| : c_{i,j} \neq 0\}$ for all $1 \leq i \leq |\mathcal{G}|$. We consider a submodule $\mathbb{B} \subseteq \mathbb{A}$ such that $\text{PVect}(\mathbb{B})$ is generated by

$$\left\{ \mathbf{1}_{b_i} \left((-\alpha_i) \boxplus \bigoplus_{j \in \mathcal{S}_i} c_{i,j} \alpha_j \right) \right\}_{1 \leq i \leq |\mathcal{G}|} .$$

Also, notice that $\mathbb{B} \simeq \bigoplus_{1 \leq i \leq |\mathcal{G}|} \mathbb{F}_{[b_i, \infty)}$ and that, by construction, $\widetilde{\mathbb{G}} \simeq \mathbb{A}/\mathbb{B}$. Now, we pick up the canonical base for \mathbb{B} and consider the inclusion $\iota : \mathbb{B} \hookrightarrow \mathbb{A}$; this will lead to an associated matrix $(\iota(\mathcal{B}))_{\mathcal{A}}$. Thus, we send $((\mathcal{B} | \mathcal{A}), \mathcal{A}, ((\iota(\mathcal{B}))_{\mathcal{A}} | \text{Id}_{|\mathcal{A}|}))$ to `box_gauss_reduce` and obtain a basis for the quotient \mathbb{A}/\mathbb{B} ; i.e., a basis for $\widetilde{\mathbb{G}}$.

5.2 PERMAVISS

Here we outline a procedure for implementing the persistence Mayer–Vietoris spectral sequence. In Sect. 5.1 we worked with $\text{GZ}_{p,q}^r$ and $\text{IB}_{p,q}^r$, which is very intuitive from a mathematical perspective. However, it is more efficient to work directly with the sets $Z_{p,q}^r$ and $B_{p,q}^r$. By storing representatives in $Z_{p,q}^r$, we avoid repeating computations on each page and in the extension problem. Furthermore, this approach allows to easily track the complexity of the algorithm. For compactness, we work with barcode bases through this section. Before we study the algorithm, we make some notational remarks. Given a spectral sequence term $E_{p,q}^r$, we denote by $\mathcal{E}_{p,q}^r$ its barcode basis. Additionally, for a generator $\alpha \in \mathcal{E}_{p,q}^r$, we denote by $\widetilde{\alpha}$ its representative in $\mathcal{S}_{p+q}^{\text{Tot}}$ and, if $r > 0$, by $\bar{\alpha}$ its representative in $E_{p,q}^{r-1}$. Also, we denote by $[\cdot]_{p,q}^r$ a class in $E_{p,q}^r$. Thus, given $\alpha \in \mathcal{E}_{p,q}^r$ we have that $\alpha = [\bar{\alpha}]_{p,q}^r = [\widetilde{\alpha}]_{p,q}^r$.

0-Page We start by defining the 0-page as the quotient

$$E_{p,q}^0 = \frac{F^p \mathcal{S}_{p+q}^{\text{Tot}}}{F^{p-1} \mathcal{S}_{p+q}^{\text{Tot}}} \cong \mathcal{S}_{p,q} = \bigoplus_{\sigma \in N_p^{\mathcal{U}}} S_q(U_\sigma)$$

for all pairs of integers $p, q \geq 0$. The 0 differential, d^0 , is isomorphic to the standard chain differential

$$d_{p,q}^0 \cong d_q : \mathcal{S}_{p,q} \rightarrow \mathcal{S}_{p,q-1} .$$

In particular, for each simplex $\sigma \in N_q^{\mathcal{U}}$, the morphism $d_{p,q}^0$ restricts to a local differential

$$d_q^\sigma : S_q(U_\sigma) \rightarrow S_{q-1}(U_\sigma) .$$

Thus, we can compute a basis, $\mathcal{E}_{\sigma,q}^1$, for the persistent homology $\text{PH}_q(U_\sigma)$ as well as a basis for the image $\text{Im}(d_{q+1}^\sigma)$. Putting all of these together, we get a basis for $E_{p,q}^1$ as the union $\mathcal{E}_{p,q}^1 = \bigcup_{\sigma \in N_p^{\mathcal{U}}} \mathcal{E}_{\sigma,q}^1$. Further, for each generator $\alpha \in \mathcal{E}_{p,q}^1 \subseteq \text{PVect}(E_{p,q}^1)$,

we store a chain $\alpha_p \in \mathcal{S}_{p,q}$ so that $\alpha = [(0, \dots, 0, \alpha_p, 0, \dots, 0)]_{p,q}^1$. We define $\tilde{\mathcal{E}}_{p,q}^1$ to be the set of representatives in $\mathcal{S}_{p+q}^{\text{Tot}}$ given by $(0, \dots, 0, \alpha_0, 0, \dots, 0)$.

1-Page Recall that the first page elements are given as classes in the quotient

$$E_{p,q}^1 = \frac{Z_{p,q}^1}{Z_{p-1,q+1}^0 + B_{p,q}^0}.$$

Therefore, for each generator $\alpha \in \mathcal{E}_{p,q}^1$, with $\alpha \sim [a_\alpha, b_\alpha)$, there is a chain $\alpha_p \in \mathcal{S}_{p,q}$, so that $\alpha = [(0, \dots, 0, \alpha_p, 0, \dots, 0)]_{p,q}^0$. Then we compute

$$d_{p,q}^1(\alpha) = [d^{\text{Tot}}(0, \dots, 0, \alpha_p, 0, \dots, 0)]_{p-1,q}^1 = [(0, \dots, 0, \bar{\delta}_p(\alpha_p), 0, \dots, 0)]_{p-1,q}^1.$$

Now, for each simplex $\tau \in N_{p-1}^{\mathcal{U}}$, we consider the local chain $(\bar{\delta}_p(\alpha_p))_\tau \in S_q(U_\tau)$ and proceed to compute a subset $\mathcal{J}_{\tau,q}^1 \subseteq (\mathcal{E}_{\tau,q}^1)^{a_\alpha}$ together with $c_\beta^1 \in \mathbb{F}$ for all $\beta \in \mathcal{J}_{\tau,q}^1$ and an element $\Gamma_\tau \in \text{PVect}(S_{q+1}(U_\tau))$ such that

$$(\bar{\delta}_p(\alpha_p))_\tau = d_{q+1}(\Gamma_\tau) \boxplus \mathbf{1}_{a_\alpha} \left(\boxplus_{\beta \in \mathcal{J}_{\tau,q}^1} c_\beta^1 \bar{\beta} \right),$$

where recall that $\bar{\beta}$ denotes a representative in $S_q(U_\tau) \subseteq E_{p-1,q}^0$. This computation can be performed by using Algorithm 1. That is, we consider $(\mathcal{E}_{\tau,q}^1)^{a_\alpha}$ and a basis for $\text{Im}((d_{q+1})_\tau)$ and write $(\bar{\delta}_p(\alpha_p))_\tau$ in terms of these; the preimages lead to Γ_τ . Repeating this for all $\tau \in N_{p-1}^{\mathcal{U}}$, we get a subset $\mathcal{J}_{p-1,q}^1 \subseteq (\mathcal{E}_{p-1,q}^1)^{a_\alpha}$ together with coefficients $c_\beta^1 \in \mathbb{F} \setminus \{0\}$ for all $\beta \in \mathcal{J}_{p-1,q}^1$ and an element $\Gamma_{p-1} \in \text{PVect}(E_{p-1,q+1}^0)$ so that

$$\bar{\delta}_p(\alpha_p) = d_{p-1,q+1}^0(\Gamma_{p-1}) \boxplus \mathbf{1}_{a_\alpha} \left(\boxplus_{\beta \in \mathcal{J}_{p-1,q}^1} c_\beta^1 \bar{\beta} \right).$$

This leads to $d_{p,q}^1(\alpha) = \mathbf{1}_{a_\alpha}(\boxplus_{\beta \in \mathcal{J}_{p-1,q}^1} c_\beta^1 \bar{\beta})$. Repeating this procedure for all generators $\alpha \in \mathcal{E}_{p,q}^1$ leads to an associated matrix $D_{p,q}^1$ for $d_{p,q}^1$. Using `image_kernel`, we compute bases for the kernel and image, together with the corresponding preimages. Next, we compute a base $\mathcal{E}_{p,q}^2$ for the second page term $E_{p,q}^2$ by applying `box_gauss_reduce` to compute the quotient $\text{Ker}(d_{p,q}^1)/\text{Im}(d_{p+1,q}^1)$. This also leads to first page representatives $\bar{\alpha} \in \text{PVect}(E_{p,q}^1)$ for all $\alpha \in \mathcal{E}_{p,q}^2$. Finally, for each generator $\alpha \in \mathcal{E}_{p,q}^2$, we proceed to find a good representative by using $\tilde{\mathcal{E}}_{p,q}^1$ together with the computed coordinates of α in terms of $\mathcal{E}_{p,q}^1$. This leads to $\tilde{\alpha} = (0, \dots, 0, \alpha_p, 0, \dots, 0) \in \mathcal{S}_{p+q}^{\text{Tot}}$, so that $\alpha = [\tilde{\alpha}]_{p,q}^2$. Since by hypotheses $[\tilde{\alpha}]_{p,q}^1 \in \text{Ker}(d_{p,q}^1)$, we might find $\alpha_{p-1} \in E_{p-1,q+1}^0$ such that $d_{p-1,q}^0(\alpha_{p-1}) =$

$-\bar{\delta}(\alpha_p)$. Altogether, we set $\tilde{\alpha} \leftarrow (0, \dots, 0, \alpha_{p-1}, \alpha_p, 0, \dots, 0)$ and store it in $\tilde{\mathcal{E}}_{p,q}^2$; notice that $d^{\text{Tot}}(\tilde{\alpha}) \in F^{p-2}\mathcal{S}^{\text{Tot}}$.

k-Page Suppose that we have computed generators $\mathcal{E}_{p,q}^k \subseteq \text{PVect}(E_{p,q}^{k-1})$, together with total complex representatives $\tilde{\mathcal{E}}_{p,q}^k$ for some $k \geq 3$. Assume also that $d^{\text{Tot}}(\tilde{\alpha}) \in F^{p-k}(\mathcal{S}^{\text{Tot}})$ for all $\tilde{\alpha} \in \tilde{\mathcal{E}}_{p,q}^k$. Notice that if $k \geq p + 1$, then $d_{p,q}^k = 0$. Thus, we focus on the case that $k < p + 1$. Let $\alpha \in \mathcal{E}_{p,q}^k$ with $\alpha \sim [a_\alpha, b_\alpha)$ together with a representative $\tilde{\alpha} \in \tilde{\mathcal{E}}_{p,q}^k$ with $\tilde{\alpha} = (0, \dots, 0, \alpha_{p-k+1}, \dots, \alpha_p, 0, \dots, 0)$ so that

$$d^k(\alpha) = [d^{\text{Tot}}(\tilde{\alpha})]_{p-k,q+k-1}^k = [(0, \dots, 0, \bar{\delta}_{p-k+1}(\alpha_{p-k+1}), 0, \dots, 0)]_{p-k,q+k-1}^k.$$

We proceed by ‘lifting’ $d^{\text{Tot}}(\tilde{\alpha})$ to the k -page. As before, using Algorithm 1 in parallel, we obtain a subset $\mathcal{J}_{p-k,q+k-1}^1 \subseteq \mathcal{E}_{p-k,q+k-1}^1$ together with coefficients $c_\beta^1 \in \mathbb{F} \setminus \{0\}$ for all $\beta \in \mathcal{T}_{p-k,q+k-1}^1 \subseteq \mathcal{S}_{p+q-1}^{\text{Tot}}$ giving us an expression in terms of the first page basis $[d^{\text{Tot}}(\tilde{\alpha})]_{p-k,q+k-1}^1 = \mathbf{1}_{a_\alpha}(\boxplus_{\beta \in \mathcal{T}_{p-k,q+k-1}^1} c_\beta^1 \beta)$. Next, using Algorithm 1, we compute a subset $\mathcal{J}_{p-k,q+k-1}^r \subseteq (\mathcal{E}_{p-k,q+k-1}^r)^{a_\alpha}$ together with coefficients $c_\beta^r \in \mathbb{F} \setminus \{0\}$ for all $\beta \in \mathcal{J}_{p-k,q+k-1}^r$, and $\Gamma_{p-k+r-1} \in \text{PVect}(E_{p-k+r-1,q+k-r+1}^{r-1})$ such that

$$\mathbf{1}_{a_\alpha} \left(\boxplus_{\beta \in \mathcal{J}_{p-k,q+k-1}^{r-1}} c_\beta^{r-1} \beta \right) = d_{p-k+r-1,q+k-r+1}^{r-1}(\Gamma_{p-k+r-1}) \boxplus \mathbf{1}_{a_\alpha} \left(\boxplus_{\beta \in \mathcal{J}_{p-k,q+k-1}^r} c_\beta^r \bar{\beta} \right).$$

Thus, we deduce an expression of $[d^{\text{Tot}}\tilde{\alpha}]_{p-k,q+k-1}^r$ in terms of $\mathcal{E}_{p-k,q+k-1}^r$. In particular, this holds for $r = k$, which leads to the associated matrix of $d_{p,q}^k$, and then we can compute `image_kernel` to obtain bases for images and kernels of $d_{p,q}^k$. Also, using `box_gauss_reduce` we obtain basis, $\mathcal{E}_{p,q}^{k+1}$, for the terms $E_{p,q}^{k+1}$. As a byproduct of this computation, for each $\alpha \in \mathcal{E}_{p,q}^{k+1}$ we obtain a previous page representative $\bar{\alpha} \in \text{PVect}(E_{p,q}^k)$, so that $[\bar{\alpha}]_{p,q}^{k+1} = \alpha$. Now, consider $\alpha \in \mathcal{E}_{p,q}^{k+1}$ with $\alpha \sim [a_\alpha, b_\alpha)$ and proceed to obtain a ‘good’ total complex representative. There exists $\mathcal{T}(\alpha) \subseteq \mathcal{E}_{p,q}^k$ together with coefficients c_β^α for all $\beta \in \mathcal{T}(\alpha)$ such that $\alpha = [\mathbf{1}_{a_\alpha}(\boxplus_{\beta \in \mathcal{T}(\alpha)} c_\beta^\alpha \beta)]_{p,q}^{k+1}$. Then we define $\tilde{\alpha} = \mathbf{1}_{a_\alpha}(\boxplus_{\beta \in \mathcal{T}(\alpha)} c_\beta^\alpha \tilde{\beta})$, and notice that $\alpha = [\tilde{\alpha}]_{p,q}^{k+1}$ as well as $d^{\text{Tot}}(\tilde{\alpha}) \in F^{p-k}(\mathcal{S}^{\text{Tot}})$. We denote $\tilde{\alpha} = (0, \dots, 0, \alpha_{p-k+1}, \dots, \alpha_p, 0, \dots, 0)$. Now, by hypotheses

$$\begin{aligned} d_{p,q}^k(\alpha) &= [d^{\text{Tot}}(\tilde{\alpha})]_{p-k,q+k-1}^k \\ &= [(0, \dots, 0, \bar{\delta}_{p-k+1}(\alpha_{p-k+1}), 0, \dots, 0)]_{p-k,q+k-1}^k = 0. \end{aligned}$$

Then, there exist $\gamma_{p-1} \in \text{PVect}(E_{p-1,q+1}^{k-1})$ with

$$d_{p-1,q+1}^{p-1}(\gamma_{p-1}) = [d^{\text{Tot}}(\tilde{\alpha})]_{p-k,q+k-1}^{k-1}.$$

By writing γ_{p-1} in terms of $\mathcal{E}_{p-1,q+1}^{k-1}$ and using their stored representatives, we may get $\tilde{\gamma}_{p-1} \in \mathcal{S}_{p+q}^{\text{Tot}}$ such that $\gamma_{p-1} = [\tilde{\gamma}_{p-1}]^{k-1}$ and also $d^{\text{Tot}}(\tilde{\gamma}_{p-1}) \in F^{p-k} \mathcal{S}_{p+q-1}^{\text{Tot}}$. In particular,

$$[d^{\text{Tot}}(\tilde{\alpha} \boxplus (-\tilde{\gamma}_{k-1}))]^{k-1} = 0$$

and we set $\tilde{\alpha} \leftarrow \tilde{\alpha} \boxplus (-\tilde{\gamma}_{k-1})$. Hence, by induction, we can repeat this procedure for all $1 \leq r \leq k$. Eventually, we should obtain a representative $\tilde{\alpha} = (0, \dots, 0, \alpha_{p-k}, \dots, \alpha_p, 0, \dots, 0)$ such that $d^{\text{Tot}}(\tilde{\alpha}) \in F^{p-k-1} \mathcal{S}_{p+q-1}^{\text{Tot}}$. We denote the new set of representatives as $\tilde{\mathcal{E}}_{p,q}^{k+1}$.

5.3 Extension Problem

After computing all pages of the spectral sequence, we still have to solve the extension problem. Recall that a solution was given in Sect. 5.1; here we only give some algorithmic guidelines. We start from a basis $\mathcal{E}_{p,q}^\infty$, with total complex representatives $\tilde{\mathcal{E}}_{p,q}^\infty$. Since we assume that the spectral sequence is bounded, it collapses at an $L > 0$ page. Then, for each generator $\alpha \in \mathcal{E}_{p,q}^L$, with $\alpha \sim [a_\alpha, b_\alpha]$, we have a corresponding representative $\tilde{\alpha} \in \tilde{\mathcal{E}}_{p,q}^L$. Consider $\tilde{\gamma} \leftarrow \mathbf{1}_{b_\alpha}(\tilde{\alpha})$; we perform changes to $\tilde{\gamma}$ similarly as in Sect. 5.1. We start by computing the classes $[\tilde{\gamma}]_{p,q}^r$ for all $1 \leq r \leq L$. We do this by using Algorithm 1 in parallel, as done on the 1-page. This leads to a subset $\mathcal{J}_{p,q}^1 \subseteq (\mathcal{E}_{p,q}^1)^{b_\alpha}$ together with coefficients $c_\beta^1 \in \mathbb{F} \setminus \{0\}$ for all $\beta \in \mathcal{J}_{p,q}^1$ and $\Gamma_p \in \mathcal{S}_{p,q+1}$, so that $[\tilde{\gamma}]_{p,q}^0 = d_{q+1}(\Gamma_p) \boxplus \mathbf{1}_{b_\alpha}(\boxplus_{\beta \in \mathcal{J}_{p,q}^1} c_\beta^1 \tilde{\beta})$. The same happens for all pages $2 \leq r \leq L$; we find $\mathcal{J}_{p,q}^r \subseteq (\mathcal{E}_{p,q}^r)^{b_\alpha}$ together with coefficients $c_\beta^r \in \mathbb{F} \setminus \{0\}$ for all $\beta \in \mathcal{J}_{p,q}^r$ and $\Gamma_{p+r-1} \in \text{PVect}(E_{p+r-1,q-r+2}^{r-1})$, so that

$$\mathbf{1}_{b_\alpha}([\tilde{\gamma}]^{r-1}) = d_{p+r-1,q-r+2}^{r-1}(\Gamma_{p+r-1}) \boxplus \mathbf{1}_{b_\alpha} \left(\boxplus_{\beta \in \mathcal{J}_{p,q}^r} c_\beta^r \tilde{\beta} \right).$$

Now, we change the total complex representative by an assignment

$$\tilde{\gamma} \leftarrow \tilde{\gamma} \boxplus \left(- \boxplus_{\beta \in \mathcal{E}_{p,q}^L} c_\beta^L \tilde{\beta} \right).$$

In particular, notice that $[\tilde{\gamma}]_{p,q}^L = 0$. We might repeat this procedure for all integers $L - 1 \geq r \geq 0$ so that one has that $[\tilde{\gamma}]_{p,q}^r = 0$. As a consequence the p -component of $\tilde{\gamma}$ vanishes, so $\tilde{\gamma} \in F^{p-1} \mathcal{S}_{p+q}^{\text{Tot}}$. Then, as in Sect. 5.1, one can repeat this process for all $p - 1 \geq r \geq 0$. This leads to all coefficients $(c_\beta^L)_{\beta \in \mathcal{E}_{p-r,q+r}^L}$ for all $0 \leq r \leq p$. These can be used to define a basis for the submodule \mathbb{B} from end of Sect. 5.1; this solves the extension problem by computing the corresponding quotient.

5.4 Complexity Analysis

Let D_s be the maximum simplex dimension in K , and let L be the number of pages. Denote by $\dim(N^{\mathcal{U}})$ the dimension of the nerve. Let

$$X = \max_{\substack{q \geq 0 \\ \sigma \in N^{\mathcal{U}}}} \{\#q - \text{simplices in } \mathcal{U}_\sigma\}.$$

On the other hand, we define

$$H = \max_{p,q \geq 0} \{|\mathcal{E}_{p,q}^1| : \mathcal{E}_{p,q}^1 \text{ is a base for } E_{p,q}^1\}.$$

Assume P is the number of processors.

0-Page When computing the first page, all we need to do is calculate persistent homology in parallel. Then, the complexity is

$$\left\lceil \frac{|N^{\mathcal{U}}|}{P} \right\rceil \mathcal{O}(X^3).$$

This leads to generators for the first page.

1-Page For the first page, recall that we start from a generator $\alpha \in \mathcal{E}_{p,q}^1$ with $\alpha \sim [a_\alpha, b_\alpha)$ and proceed to solve $|N_{p-1}^{\mathcal{U}}|$ linear equations. Notice that this can be done for all generators from $\mathcal{E}_{p,q}^1$ simultaneously. This is because as the value a_α changes, we might select the columns associated to $(\mathcal{E}_{p,q}^1)^{a_\alpha}$ from the columns of $\mathcal{E}_{p,q}^1$ which does not affect the complexity of Algorithm 1. On the other hand, we need to execute `image_kernel` on at most $\dim(N^{\mathcal{U}}) \cdot D_s$ elements on the first page. Notice that for each of these, we first compute a basis for the images and kernels. Afterwards, we perform the quotients using `box_gauss_reduce` which takes a complexity of at most $\mathcal{O}(H^3)$. Also, we need to add the complexity of the Čech differential. An option for computing this, is to compare simplices in different covers by their vertices; two simplices are the same iff they share the same vertex set. This would take less than $\mathcal{O}(|N^{\mathcal{U}}| D_s X^2 H)$ operations. Thus the overall complexity becomes

$$\left\lceil \frac{|N^{\mathcal{U}}|}{P} \right\rceil \mathcal{O}(X^2 H) + \left\lceil \frac{\dim(N^{\mathcal{U}}) \cdot D_s}{P} \right\rceil (\mathcal{O}(|N^{\mathcal{U}}| D_s X^2 H) + \mathcal{O}(H^3)).$$

k-Page Now, we proceed for the complexity of the page $k \geq 2$. This is the same as for the 1 page, with the addition of Gaussian eliminations of higher pages. These take at most $\mathcal{O}(H^3)$ time. Denoting by L the infinity page, we have the new term

$$\left\lceil \frac{\dim(N^{\mathcal{U}}) \cdot D_s}{P} \right\rceil \mathcal{O}(LH^3)$$

which added to the complexity of the 1-page gives

$$\begin{aligned} \left\lceil \frac{|N^{\mathcal{U}}|}{P} \right\rceil \mathcal{O}(X^2H) + \left\lceil \frac{\dim(N^{\mathcal{U}}) \cdot D_s}{P} \right\rceil (\mathcal{O}(|N^{\mathcal{U}}|_{D_s} X^2H) + \mathcal{O}(H^3) + \mathcal{O}(LH^3)) \\ = \left\lceil \frac{\dim(N^{\mathcal{U}}) \cdot D_s}{P} \right\rceil (\mathcal{O}(|N^{\mathcal{U}}|_{D_s} X^2H) + \mathcal{O}(LH^3)). \end{aligned}$$

Extension problem If the spectral sequence collapses at $L > 0$, then the complexity of extending all generators in $\mathcal{E}_{p,q}^L$ is bounded by that of computing the L page about D_s times.

Overall complexity Altogether, we have a complexity bounded by that of computing the first page plus that of computing the L page $L + D_s$ times. Here the L comes from computing the L page L times and D_s from the extension problem. Thus, the overall complexity is bounded by

$$\left\lceil \frac{|N^{\mathcal{U}}|}{P} \right\rceil \mathcal{O}(X^3) + (L + D_s) \left\lceil \frac{\dim(N^{\mathcal{U}}) \cdot D_s}{P} \right\rceil (\mathcal{O}(|N^{\mathcal{U}}|_{D_s} X^2H) + \mathcal{O}(H^3)).$$

Notice that in general D_s , L and $\dim(N^{\mathcal{U}})$ are much smaller than H and X . Thus, for covers such that $|N^{\mathcal{U}}| \ll X$, and assuming we have enough processors, the complexity can be simplified to the two dominating terms $\mathcal{O}(X^3) + \mathcal{O}(H^3)$. Notice that this last case is satisfied for those covers whose mutual intersections are generally smaller than each cover. Also, in this case H is approximately of the order of nontrivial barcodes over all the input complex. This shows that PERMAVISS isolates simplicial data, while only merging homological information. It is worth to notice that in general H , being the number of nontrivial bars, is much smaller than the size of the whole simplicial complex.

6 Conclusion

We started by developing linear algebra for persistence modules. In doing so, we introduced bases of persistence modules, as well as associated matrices to morphisms. Also, we presented Algorithm 2, which computes bases for the image and the kernel of a persistence morphism between any pair of tame persistence modules. Then a generalization of traditional persistent homology was introduced in Sect. 3.5. This theory has helped us to define and understand the Persistent Mayer–Vietoris spectral sequence. Furthermore, we have provided specific guidelines for a distributed algorithm, with a solution to the extension problem presented in Sect. 5.1. The PERMAVISS method presented in Sect. 5.2 isolates simplicial information to local matrices, while merging only homological information between different covers. Thus, the complexity of this method is dominated by the size of a local complex plus the order of barcodes over all the data. A first implementation of these results can be found in [29]. Coding an efficient implementation from the pseudo-code given in this paper, and benchmarking its performance compared to other methods, will be a matter of future research.

Another interesting direction of research is how to merge this method with existing algorithms, such as those from [7, 8, 21, 26]. Especially it would be interesting to explore the possible interactions of discrete Morse theory and this approach, see [11]. Additionally, it will be worth exploring, both theoretically and practically, which are the most suitable covers for different applications. Finally, we would also like to study the additional information given by the covering. This will add locality information from persistent homology.

Acknowledgements The author would like to thank his supervisor Dr. Ulrich Pennig who suggested this topic and has been very helpful and supportive in the development of these ideas. Also, the author would like to express his gratitude to EPSRC for the Grant EP/N509449/1 support with Project Number 1941653, without which the author would not have been able to write this work.

Data availability There is no data associated to this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., Ziegelmeier, L.: Persistence images: a stable vector representation of persistent homology. *J. Mach. Learn. Res.* **18**, # 8 (2017)
2. Bauer, U., Kerber, M., Reininghaus, J.: Clear and compress: computing persistent homology in chunks. In: *Topological Methods in Data Analysis and Visualization III* (Davis 2013), pp. 103–117. Springer, Berlin (2014)
3. Bott, R., Tu, L.W.: *Differential Forms in Algebraic Topology*. Graduate Texts in Mathematics, vol. 82. Springer, New York (1982)
4. Bredon, G.E.: *Sheaf Theory*. Graduate Texts in Mathematics, vol. 170. Springer, New York (1997)
5. Carlsson, G.: Topology and data. *Bull. Am. Math. Soc.* **46**(2), 255–308 (2009)
6. Chazal, F., de Silva, V., Glisse, M., Oudot, S.: *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics. Springer, Cham (2016)
7. Chen, C., Kerber, M.: Persistent homology computation with a twist. In: *27th European Workshop on Computational Geometry* (Morschach 2011), pp. 197–200 (2011). <https://eurocg11.inf.ethz.ch/docs/Booklet.pdf>
8. Chen, C., Kerber, M.: An output-sensitive algorithm for persistent homology. *Comput. Geom.* **46**(4), 435–447 (2013)
9. Chow, T.Y.: You could have invented spectral sequences. *Not. Am. Math. Soc.* **53**(1), 15–19 (2006)
10. Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Morozov, D.: Persistent homology for kernels, images, and cokernels. In: *20th Annual ACM-SIAM Symposium on Discrete Algorithms* (New York 2009), pp. 1011–1020. SIAM, Philadelphia (2009)
11. Curry, J., Ghrist, R., Nanda, V.: Discrete Morse theory for computing cellular sheaf cohomology. *Found. Comput. Math.* **16**(4), 875–897 (2016)
12. Delfinado, C.J.A., Edelsbrunner, H.: An incremental algorithm for Betti numbers of simplicial complexes on the 3-sphere. *Comput. Aided Geom. Des.* **12**(7), 771–784 (1995)
13. Di Fabio, B., Landi, C.: Persistent homology and partial similarity of shapes. *Pattern Recognit. Lett.* **33**(11), 1445–1450 (2012)

14. Edelsbrunner, H., Harer, J.L.: *Computational Topology: An Introduction*. American Mathematical Society, Providence (2010)
15. Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. *Discrete Comput. Geom.* **28**(4), 511–533 (2002)
16. Ghrist, R.: *Elementary Applied Topology*. Createspace (2014)
17. Govc, D., Skraba, P.: An approximate nerve theorem. *Found. Comput. Math.* **18**(5), 1245–1297 (2018)
18. Lewis, R., Morozov, D.: Parallel computation of persistent homology using the blowup complex. In: *27th ACM Symposium on Parallelism in Algorithms and Architectures (Portland 2015)*, pp. 323–331. ACM, New York (2015)
19. Lipsky, D., Skraba, P., Vejdemo-Johansson, M.: A spectral sequence for parallelized persistence (2011). [arXiv:1112.1245](https://arxiv.org/abs/1112.1245)
20. McCleary, J.: *A User's Guide to Spectral Sequences*. Cambridge Studies in Advanced Mathematics, vol. 58. Cambridge University Press, Cambridge (2001)
21. Milosavljević, N., Morozov, D., Škraba, P.: Zigzag persistent homology in matrix multiplication time. In: *27th Annual Symposium on Computational Geometry (Paris 2011)*, pp. 216–225. ACM, New York (2011)
22. Munkres, J.R.: *Elements of Algebraic Topology*. Addison-Wesley, San Francisco (2018)
23. Robins, V., Turner, K.: Principal component analysis of persistent homology rank functions with case studies of spatial point patterns, sphere packing and colloids. *Physica D* **334**, 99–117 (2016)
24. Robinson, M.: *Topological Signal Processing*. Mathematical Engineering. Springer, Heidelberg (2014)
25. de Silva, V., Ghrist, R.: Coverage in sensor networks via persistent homology. *Algebr. Geom. Topol.* **7**(1), 339–358 (2007)
26. de Silva, V., Morozov, D., Vejdemo-Johansson, M.: Dualities in persistent (co)homology. *Inverse Probl.* **27**(12), # 124003 (2011)
27. Singh, G., Mémoli, F., Carlsson, G.: Topological methods for the analysis of high dimensional data sets and 3D object recognition. In: *Eurographics Symposium on Point-Based Graphics (Prague 2007)*, pp. 91–100. Eurographics Association (2007)
28. Skraba, P., Vejdemo-Johansson, M.: Persistence modules: algebra and algorithms (2013). [arXiv:1302.2015](https://arxiv.org/abs/1302.2015)
29. Torras Casas, Á.: PerMaViss: persistence Mayer Vietoris spectral sequence (2020). <https://doi.org/10.5281/zenodo.3613870>
30. Weibel, C.A.: *An Introduction to Homological Algebra*. Cambridge Studies in Advanced Mathematics, vol. 38. Cambridge University Press, Cambridge (1994)
31. Yoon, H.R.: Cellular Sheaves and Cosheaves for Distributed Topological Data Analysis. PhD thesis, University of Pennsylvania (2018). <https://repository.upenn.edu/edissertations/2936>
32. Yoon, H.R., Ghrist, R.: Persistence by parts: multiscale feature detection via distributed persistent homology (2020). [arXiv:2001.01623](https://arxiv.org/abs/2001.01623)
33. Zomorodian, A., Carlsson, G.: Localized homology. *Comput. Geom.* **41**(3), 126–148 (2008)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.