Harmonic persistent homology

Saugata Basu[*] and Nathanael Cox[†]

Abstract. We introduce harmonic persistent homology spaces for filtrations of finite simplicial complexes. As a result we can associate concrete subspaces of cycles to each bar of the barcode of the filtration. We prove stability of the harmonic persistent homology subspaces, as well as the subspaces associated to the bars of the barcodes, under small perturbations of functions defining them. We relate the notion of "essential simplices" introduced in an earlier work to identify simplices which play a significant role in the birth of a bar, with that of harmonic persistent homology. We prove that the harmonic representatives of simple bars maximizes the "relative essential content" amongst all representatives of the bar, where the relative essential content is the weight a particular cycle puts on the set of essential simplices.

Key words. Harmonic homology, persistent homology, barcodes, essential simplices, stability

AMS subject classifications. 55N31

1. Introduction.

The main topic of this paper concerns the theory of *persistent homology* which is a central object of interest in the burgeoning field of *topological data analysis*.

1.1. Background on persistent homology.

We begin with some motivation behind the introduction of persistent homology. One way that simplicial complexes arise in topological data analysis is via the Čech (or its closely related cousin, the Vietoris-Rips) complex [14, pp. 60-61] associated to a point set. Let X be a (finite) subset of some metric space which for concreteness let us assume to be $\mathbb{R}^d$ (with its Euclidean metric). In practice, X may consist of a finite set of points (often called "point-cloud data") which approximates some subspace or sub-manifold M of $\mathbb{R}^d$. The topology (in particular, the homology groups) of the manifold M is not reflected in the set of points X (which is a discrete topological space under the subspace topology induced from that of $\mathbb{R}^d$). Now for $r \geq 0$, let X_r denote the union of closed Euclidean balls, $B(x, r)$, of radius r centered at the points $x \in X$. In particular, $X_0 = X$. Also, for $0 \leq r \leq r'$, we have that $X_r \subset X_{r'}$. Thus, $(X_r)_{r\in\mathbb{R}_{\geq 0}}$ is an increasing family of topological spaces indexed by $r \geq 0$. This is an example of a (continuous) filtration of topological spaces. For each $r \geq 0$, we can associate a finite simplicial complex K_r – the *nerve complex* of the tuple of balls $(B(x, r))_{x\in X}$. Informally, the simplicial complex K_r has vertices indexed by the set X, and for each subset X' of X of cardinality $p+1$, we include the p-dimensional simplex spanned by the vertex set corresponding to X' if and only if

$$\bigcap_{x\in X'} B(x, r) \neq \emptyset.$$

It is a basic result in algebraic topology (the "nerve lemma") that the simplicial homology groups, $\mathrm{H}_*(K_r)$, are isomorphic to the (say singular) homology groups, $\mathrm{H}_*(X_r)$, of X. More precisely, the nerve lemma states that the geometric realization $|K_r|$ is homotopy equivalent to (in fact, is a deformation retract of) X_r (homotopy equivalent spaces have isomorphic homology groups). Observe that for $r \leq r'$, K_r is a sub-simplicial complex of $K_{r'}$, and since there are only finitely many simplicial complexes on card(X)-many vertices, there are finitely many distinct simplicial complexes in the tuple $(K_r)_{r\geq 0}$. Thus, we obtain a finite nested sequence $\mathcal{F}$ of simplicial complexes, $K_0 \subset K_{r_1} \subset K_{r_2} \subset \cdots \subset K_{r_n}$ in which each complex is a subcomplex of the next. We will refer to $\mathcal{F}$ as a finite *filtration* of simplicial complexes.

[*]Department of Mathematics, Purdue University, 150 N. University St., West Lafayette IN, USA (sbasu@purdue.edu, https://www.math.purdue.edu/~sbasu/).

[†]Department of Mathematics, Purdue University, 150 N. University St., West Lafayette IN, USA (cox175@purdue.edu).

Let us return to the picture of the point-cloud X approximating an underlying manifold M. The homology of the manifold is captured (by virtue of the nerve lemma) by the simplicial homology groups of the various simplicial complexes occurring in the finite filtration $\mathcal{F}$. However, this correspondence is not bijective. As one can easily visualize, as r starts growing from 0, there are many spurious homology classes that are born and quickly die off (i.e. the corresponding holes are filled in) and these have nothing to do with the topology of M. Persistent homology is a tool that can be used to separate this "noise" from the bona fide homology classes of M. The persistent homology of the filtration $\mathcal{F}$ is encoded as a set of intervals (called bars) in the barcode of the filtration $\mathcal{F}$ (see Definition 3.7)). Intervals (bars) of short length corresponds to noise, while the ones which are long (persistent) reflect the homology of the underlying manifold M. The barcode of the filtration associated to X can be used as a feature of X for learning or comparison purposes. In particular, the barcodes of two finite sets X, X', which are "close" as finite metric spaces, are themselves close under an appropriately defined notion of distance between barcodes. Such results (called stability theorems) form the theoretical basis of practical applications of persistent homology, and we will state new stability theorems later on in the paper.

1.2. Associating cycles to bars. As mentioned previously, the output of a persistent homology computation is often displayed as a "barcode" (see Example 1.1 and Figure 1 for an illustration). The barcode is considered an important invariant of the given filtration (or of the underlying metric space giving rise to the filtration). We will define precisely barcodes of filtrations later in the paper (see Definition 3.7). For the moment it will suffice to note that the individual "bars" in the barcode of a filtration have some intuitive topological meaning (as explained above in the context of point-cloud data). They correspond loosely speaking to the lifetime of homology classes appearing in the homology of the simplicial complexes that appear in the filtration (here we are thinking of the ordered index set of the filtration as time). However, a new homology class that is "born" at a certain time is defined only modulo a certain subspace in the homology of the complex at that time – thus identifying a bar with a particular homology class is problematic.

Often in practice there is a demand to associate not just a homology class, but a specific *cycle from the chain group representing this class* or at least a *set of simplices* to each bar. This is because in applications the simplices of the simplicial complexes of a filtration themselves often have special significance. For instance, the vertices of a given simplicial complex could be labelled by genes and a p-simplex $\sigma = (g_0, \ldots, g_p)$ may signify positive correlation between the genes $g_0, \ldots, g_p$ (say in causing a certain disease). As an example, in [25], the authors associate bars with representative cycles to determine how topological features correlate with genes that are associated with cancer biogenesis.

There has been several approaches to the problem of associating specific cycle representatives to persistent homology classes. Most of these approaches involve minimization of some weight on the space of cycles representing a homology class. For instance, volume-optimal cycles were proposed in the non-persistent setting in [10] and in the setting of persistent homology in [28]. Volume-optimal cycles are cycles of a homology class with the fewest number of simplices, and they can be found as solutions to a linear programming optimization problem [10]. In Dey et. al. [11] the authors give a polynomial time algorithm for computing an optimal representative for a given finite bar (interval) in the p-th persistence diagram of a filtration of a simplicial complex which is a weak $(p + 1)$-pseudo-manifold. Algorithms for computing such optimal cycles have been implemented – see for instance [22]. A different approach for selecting a representative cycle can be found in [16]. The authors obtain a representative cycle by tracking when the addition or removal of a simplex causes a class to be born or die.

In this paper we describe a new approach based on the theory of *harmonic chains*. We consider homology groups with coefficients in $\mathbb{R}$ and impose an inner product on the chain group to make the chain groups an Euclidean space. As a result we are able to identify the various persistent homology groups, as well as the bars in the barcode of the given filtration, as subspaces of the simplicial chain groups themselves. Note that in contrast with ordinary persistent homology theory, we are able to associate canonically (only depending on the chosen inner product) a certain subspace of the chain space to each bar. When the bar is of multiplicity one (see Definition 3.7), this subspace is spanned by a single vector, and we have a uniquely defined (up to scalar multiplication) cycle representing the bar – we call such a cycle a *harmonic representative* of the bar. Intuitively, instead of selecting a representative cycle of the smallest possible weight (as in [10, 28]), which might in fact omit some simplices altogether, the harmonic representative of a bar will tend to produce an "average" representative.

There are several reasons to consider harmonic representatives. Instead of trying to optimize the length of the cycle the harmonic representatives put more relative weight on certain important simplices. Since as remarked earlier, the simplices themselves in the simplicial complex underlying the filtration often have domain dependent meaning – if a particular simplex shows up with non-zero coefficient in *every* cycle representing the homology class, then this fact may be considered significant from an application point of view (the lengths of representative cycles are not so significant in these applications). This idea was formalized in [3] where the notion of *essential* simplices corresponding to the bars of a barcode was introduced. Informally, a simplex is essential relative to a bar, if it occurs with a non-zero coefficient in *every* cycle representing the bar. We generalize this notion and associate a set of essential simplices to any simple bar (bars having multiplicity one – see Definition 4.4). The harmonic representative of a bar will maximize amongst all representative cycles the relative weight of the essential simplices (see Section 1.3).

A second and perhaps more important justification of considering harmonic persistent homology is that the harmonic persistent homology more accurately reflect the "geometry" of the filtrations on labelled simplicial complexes. We prove stability results (see Section 3.3) – harmonic persistent homology subspaces of simplicial filtrations which are close will be close (in a technical sense to be defined later) as elements of certain Grassmannians. In addition, filtrations whose harmonic persistent homology are close as subspaces will also have the harmonic representatives of their bars which are close (angle between corresponding subspaces will be small). Thus, bar diagrams augmented with the harmonic representatives of the bars is potentially a stronger signature of the data in some applications (see Section 1.7).

1.3. Harmonic and essential. We establish an important connection between the harmonicity of a representative cycle and the set of essential simplices of a bar. If a bar in the barcode of a filtration is of multiplicity one (this happens generically), then it is represented by a unique harmonic representative (unique up to multiplication by non-zero scalar). We define for each cycle,

$$z = \sum_{\sigma} c_\sigma \cdot \sigma,$$

(not necessarily harmonic) representing any given simple bar the relative essential content,

$$\mathrm{content}(z) = \left(\frac{\sum_{\sigma \text{ is essential}} c_\sigma^2}{\sum_\sigma c_\sigma^2} \right)^{1/2},$$

of the cycle (see Definition 4.5 for the precise definition) which measures the relative weight in the cycle of the essential as opposed to the non-essential simplices.

1.3.1. Our Result. We prove (Theorem 4.8) that the harmonic representatives of bars maximize (amongst all representative cycles) the relative essential content of the bar (see Definition 4.5) i.e.

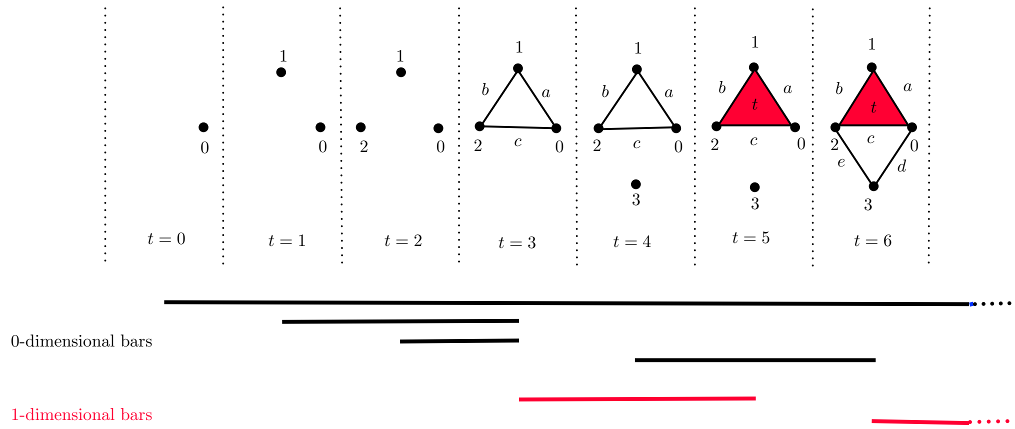

Figure 1: Barcode of a filtration

if z_0 is a harmonic representative of a simple bar b, then for any cycle z representing b,

$$\mathrm{content}(z) \leq \mathrm{content}(z_0).$$

This indicates that in applications where one would like to emphasize the role of essential simplices harmonic representatives of bars are preferable over (say) volume optimal ones mentioned above.

1.4. Example. Before proceeding further we discuss an example which illustrates the notion of harmonic represesentatives and essential simplices. A bar b in the barcode of a filtration is described by a triple (s, t, μ), where s denotes the birth time, t the death time and μ the multiplicity (see Definition 3.7).

We denote by $\mathcal{P}_p^{s,t}(\mathcal{F})$ the p-dimensional persistent harmonic homology subspace of the chain space $C_p(K)$ corresponding to the bar born at time s and which dies at time t (or never dies if $t = \infty$).

We denote by $\Sigma(b)$ the set of essential simplices associated to a simple bar b.

Example 1.1. *Let K be the simplicial complex defined by:*

$$\begin{aligned} K^{[0]} &= \{[0], [1], [2], [3]\}, \\ K^{[1]} &= \{a = [0,1], b = [1,2], c = [0,2], d = [0,3], e = [2,3]\}, \\ K^{[2]} &= \{t = [0,1,2]\}. \end{aligned}$$

For $p = 0, 1, 2$, we choose the standard inner product on $C_p(K)$ (see (2.1)).

Let $\mathcal{F}$ be the following filtration on the K:

$$\emptyset \subset \{0\} \subset \{0,1\} \subset \{0,1,2\} \subset \{0,1,2,a,b,c\} \subset \{0,1,2,3,a,b,c\}$$

$$\subset \{0,1,2,3,a,b,c,t\} \subset \{0,1,2,3,a,b,c,d,e,t\}.$$

For simplicity, we assume that vertex $\{0\}$ is added at time 0, and each complex in the filtration occurs at time 1 greater than the complex preceding it (see Figure 1). It is clear that all of the bars of the barcode of this filtration are simple. The corresponding harmonic persistent homology subspaces are listed in the following table. Note that since all the bars are simple, all these subspaces have dimension 1.

$p=0$	
$\mathcal{P}_0^{0,\infty}(\mathcal{F})$	$\mathrm{span}\{0\}$
$\mathcal{P}_0^{1,3}(\mathcal{F})$	$\mathrm{span}\{1\}$
$\mathcal{P}_0^{2,3}(\mathcal{F})$	$\mathrm{span}\{2\}$
$\mathcal{P}_0^{4,6}(\mathcal{F})$	$\mathrm{span}\{3\}$

$p=1$	
$\mathcal{P}_1^{3,5}(\mathcal{F})$	$\mathrm{span}\{a+b-c\}$
$\mathcal{P}_1^{6,\infty}(\mathcal{F})$	$\mathrm{span}\left\{a+b+2c-3d+3e\right\}$

For $p=1$, the set of simplices for each bar is listed below.

$$\Sigma((3,5,1)) = \{a,b,c\},$$
$$\Sigma((6,\infty,1)) = \{d,e\}.$$

The relative essential content of the harmonic representatives of these two bars are given by

$$\mathrm{content}((3,5,1)) = 1,$$
$$\mathrm{content}((6,\infty,1)) = \left(\frac{3}{4}\right)^{1/2}.$$

Consider the simple bar born at time 6. The homology class corresponding to this bar can be represented by many (infinitely) different cycles. The shortest or volume optimal cycle representing it is $c+d-e$. The harmonic representative is given by $a+b+2c-3d+3e$.

The relative essential content of the volume optimal cycle is $\left(\frac{2}{3}\right)^{1/2}$ which is strictly smaller than that of the harmonic representative which is equal to $\left(\frac{3}{4}\right)^{1/2}$.

Remark 1.1. Note that in [20, 19] the authors define the notion of critical simplices that is related but is not equivalent to the notion of being essential as defined in this paper. In fact it is easy to come up with examples of filtrations where the set of "critical simplices" in the sense of [20] is properly contained in the set of essential simplices.

1.5. Stability of harmonic persistent homology and harmonic barcodes. Stability theorems say that the persistent homology or its associated barcode is stable under perturbations of the input data. This makes persistent homology useful in applications, where the data often comes from physical measurements with their attendant sources of error. A basic body of results that makes persistent homology theory relevant in topological data analysis is about the stability of the persistence module [7, 9, 8]. There are many variations of stability results in the literature. We refer the reader to [14, Chapter VIII] and [5, §5.6] for a survey of these results.

The Euclidean space structure on the space of chains gives us the ability to talk about distances between harmonic homology subspaces corresponding to different sub-complexes of some fixed ambient simplicial complex. This is important because it allows us to prove stability theorems – *sub-complexes which are close should have harmonic homology spaces which are close under a natural metric* (see Theorem 2.15 and Example 2.1).

In order to discuss distance between harmonic homology subspaces, we need a notion of distance between subspaces of a finite dimensional real vector space V. The set of all d-dimensional subspaces of a vector space V is a well-studied topological space (in fact, has the structure of a projective algebraic variety) called the Grassmannian denoted $\mathrm{Gr}(d,V)$. Since, we will have subspaces of possibly different dimensions, we need a metric not just on $\mathrm{Gr}(d,V)$ but on the on the disjoint union

$$\mathrm{Gr}(V) = \coprod_{0\leq d\leq \dim V} \mathrm{Gr}(d,V)$$

where V is an Euclidean space V (V will be a chain space equipped with an inner product).

Metrics on Grassmannians were studied in detail by Lim and Ye [32] where the authors introduce several notions of distance between subspaces of varying dimensions. We use in this paper the distance called "Grassmann distance" in [32]. We prove a new class of stability theorems – bounding the distance between two harmonic filtration functions (see Definition 3.14) in terms of certain norms of the difference of the functions inducing the filtrations (see Theorems 3.20 and 3.18). These theorems should be compared with corresponding results for classical persistent homology groups (see [14, 5]).

1.5.1. Our results. We prove several different stability theorems.

Firstly, we leverage the fact that harmonic homology spaces (we denote the harmonic homology subspace of dimension p of a simplicial complex K by $\mathcal{H}_p(K)$) are elements of the Grassmannian $\mathrm{Gr}(C_p(K))$, where $C_p(K)$ is the p-th chain group of the simplicial complex K. The Grassmannian $\mathrm{Gr}(C_p(K))$ carries a natural metric induced by the Euclidean inner product on $C_p(K)$). Thus, it is meaningful to ask for a stability result for the subspaces of $\mathcal{H}_p(K)$ themselves. We prove (see Theorem 2.15 in Section 2) that for K a finite simplicial complex and K_1, K_2 sub-complexes of K, for each $p \geq 0$, $d_{K,p}(K_1, K_2) \leq \frac{\pi}{2} \cdot \Delta_p(K_1, K_2)^{1/2}$ where $d_{K,p}(K_1, K_2)$ denotes the Grassmannian distance between $\mathcal{H}_p(K_1)$ and $\mathcal{H}_p(K_2)$ in the Grassmannian $\mathrm{Gr}(C_p(K))$ (Definition 2.11), and $\Delta_p(K_1, K_2)$ is a natural measure of difference between K_1 and K_2 which is defined precisely in Theorem 2.15.

We next consider the harmonic persistent homology of filtrations of a finite simplicial complex K induced by certain admissible functions $f : K \to \mathbb{R}$ (see Definition 3.13). Such a function induces for each $p \geq 0$, two functions, $t \mapsto \mathcal{H}_p(K_{f \leq t})$, $(s,t) \mapsto \mathcal{H}_p^{s,t}(\mathcal{F}_f)$. Here $K_{f \leq t}$ denotes the sub-level complex of f (see Notation 3.3), and $\mathcal{H}_p^{s,t}(\mathcal{F}_f)$ is the (p,s,t)-th harmonic persistent homology subspace of the filtration $\mathcal{F}_f$ induced by f (see Definition 3.8).

A stability result on each of the above functions should state that for any pair of admissible functions $f, g : K \to \mathbb{R}$ which are close (under some metric) the corresponding pairs of functions $t \mapsto \mathcal{H}_p(K_{f \leq t}), t \mapsto \mathcal{H}_p(K_{g \leq t})$ as well as $(s,t) \mapsto \mathcal{H}_p^{s,t}(\mathcal{F}_f), (s,t) \mapsto \mathcal{H}_p^{s,t}(\mathcal{F}_g)$, should be close to each other. We prove that (Theorems 3.18 and 3.20), given two functions $f, g : K \to \mathbb{R}$ (satisfying a certain technical condition), the distance between the corresponding function pair $t \mapsto \mathcal{H}_p(K_{f \leq t}), t \mapsto \mathcal{H}_p(K_{g \leq t})$ as well as the pair $(s,t) \mapsto \mathcal{H}_p^{s,t}(\mathcal{F}_f), (s,t) \mapsto \mathcal{H}_p^{s,t}(\mathcal{F}_g)$, (defined via integrating over the appropriate domains the respective Grassmannian distances), is bounded from above by a constant times certain semi-norms of the functions of the function $f - g$ (see Definition 3.17).

We also study the stability of the harmonic barcodes. Let K be a finite simplicial complex and let $\mathcal{F}$ denote a finite filtration $K_0 \subset \cdots \subset K_N = K$. The harmonic barcode subspaces $\mathcal{P}_p^{s,t}(\mathcal{F})$ (see Definition 3.11) are subspaces of $\mathcal{H}_p(K_s)$ (corresponding to the birth of the homology classes corresponding to the bar $\mathcal{B} = (s, t, \mathcal{P}_p^{s,t}(\mathcal{F}))$ (assuming $\mathcal{P}_p^{s,t}(\mathcal{F}) \neq 0$). It also makes sense to consider the subspace of $\mathcal{H}_p(K_{t-1})$ representing the bar $\mathcal{B}$ just before its death – which we denote by $\widehat{\mathcal{P}}_p^{s,t}(\mathcal{F})$ (see Definition 3.23). We call $\widehat{\mathcal{P}}_p^{s,t}(\mathcal{F})$ the terminal harmonic representative of the corresponding bar. For technical reasons we prove stability of the terminal harmonic representatives of harmonic barcodes.

We first define an appropriate notion of distance between harmonic barcodes of two different filtrations using the terminal harmonic representatives. The distance measured introduced for proving stability of the harmonic homology subspaces and also the harmonic persistent homology subspaces (Theorems 3.18 and 3.20) are in the form of an integral (see Definitions 3.16 and 3.19). Since for a filtration $\mathcal{F}$ of a finite simplicial complex K, the subspaces $\mathcal{P}_p^{s,t}(\mathcal{F})$ will be non-zero only for a finitely many pairs (s,t), the integral form of the distance function is not suitable.

For two finite filtrations $\mathcal{F}$ and $\mathcal{G}$ indexed by the same ordered indexing set of cardinality $N+1$, we will use the sum (see Theorem 3.24 below) $\frac{1}{\binom{N+1}{2}} \cdot \sum_{s<t} d_{C_p(K)}(\widehat{\mathcal{P}}_p^{s,t}(\mathcal{F}), \widehat{\mathcal{P}}_p^{s,t}(\mathcal{G}))$ as a measure of distance between the harmonic barcodes of $\mathcal{F}, \mathcal{G}$, where $d_{C_p(K)}(\cdot,\cdot)$ denotes the Grassmann distance between subspaces in $C_p(K)$ (see Definition 2.11).

We prove that (Theorem 3.24) $\frac{1}{\binom{N+1}{2}} \cdot \sum_{0 \leq s<t \leq 1} d_{C_p(K)}(\widehat{\mathcal{P}}_p^{s,t}(\mathcal{F}), \widehat{\mathcal{P}}_p^{s,t}(\mathcal{G}))$ is bounded from above by a constant times the 1-norm of the function $f-g$, where $f, g : K \to [0,1]$ are admissible functions inducing the filtrations $\mathcal{F}, \mathcal{G}$.

The stability results described above give theoretical validity to the use of harmonicity in persistent homology theory.

1.6. Prior and related work. The definition of harmonic subspaces of chain spaces of a simplicial complex goes back to the work of Eckmann [12]. It has been discussed in the context of statistical inference and developing and studying graph Laplacians and Hodge theory on graphs by Lim in [24]. The theory of harmonic homology is closely related to generalized Hodge theory and L^2-cohomology. The study of Hodge theory on general metric spaces with motivation coming from data analysis and computer vision was initiated in Bartholdi et al in [1]. However, the emphasis in their work is different from that of the current paper and is concerned more about the connections and interplay between the generalized and the classical Hodge theory on Riemannian manifolds. The theory of L^2-cohomology of finite simplicial complexes was studied from an algorithmic perspective by Friedman in [15] who gave an efficient algorithm for computing them. This work is also not directly to the results of this paper. Persistent harmonic cohomology has also being mentioned before [23] (see also [6]), but the emphasis is more on manifold-learning rather than on general simplicial filtrations considered in this paper. Memoli et al. [27] also studies persistent homology groups, defining them in terms of the Laplace operator (see Remark 2.7) and gives efficient algorithms for computing them. They also establish interesting connections with spectral graph theory and prove certain stability results on the eigenvalues of the Laplace operator when applied to a simplicial filtration. These results and those in the current paper are somewhat orthogonal and it would be interesting to investigate if they are related. The definition of harmonic barcodes of filtrations defined in the current paper is new to our knowledge.

1.7. Applications. Persistent homology barcodes have found wide applications in many different areas. Persistent harmonic barcodes as defined in this paper carry more information but has not yet been applied in practice. In this section we discuss possible applications where the extra information in the persistent harmonic barcodes could potentially prove useful.

The data to which persistent homology methods are applied can be broadly categorized into two classes – labelled and unlabelled. The typical example of unlabelled data are point cloud data approximating some underlying manifold. Persistent homology is a tool to understand the global topology of the manifold. The individual points in the point cloud are not so significant by themselves – it is only their interaction with neighbors that is important. If the point cloud is contained in some Euclidean space, then the persistent homology barcode of the Vietoris-Rips filtrations is invariant under any isometry applied to the data.

The situation is quite different if the input data is labelled. In this case each labelled (weighted) simplex in the corresponding simplicial complex carry information about some relationship between its vertices. The situation is now more rigid and one needs to be able to distinguish between two isomorphic simplicial filtrations related (say) by a permutation of its vertices. The ordinary persis-

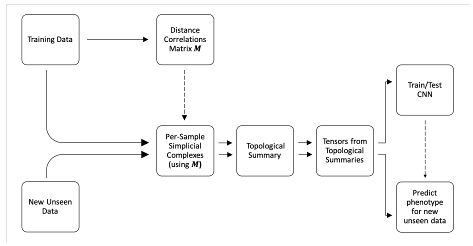

Figure 2: Pipeline reproduced from [26]

tent homology barcodes cannot distinguish between two such filtrations – however, the persistent harmonic barcodes introduced in this paper with the extra information will be able to distinguish between them. This is very important in applications (such as in genomics) where persistent homology methods are applied to "relationship" data as opposed to point-cloud data from some $\mathbb{R}^d$ (see [29] and also [26, 18, 21, 30] for recent representative examples of such applications)

In practical applications, often the barcodes obtained from the input data are fed into some standard machine learning algorithm such as a convolution neural net (CNN). The classification obtained by such a two-step pipeline is seen to be better in practice than that obtained using either TDA or CNN by itself. Since the harmonic barcodes carry more refined information and have desirable properties (stability and maximizing relative essential content) it is reasonable to postulate that in a two-step pipeline feeding extra information regarding the harmonic persistent barcodes to the machine learning programs would improve the quality of the output.

We now describe a specific scenario where harmonic persistent barcodes could prove useful. This is an example of an application of topological data analysis in genomics where the labels of vertices are significant.

1.7.1. Phenotype prediction. In [26] the authors investigate if gene expression measured from RNA sequencing contains enough signal to separate healthy from individuals afflicted with Parkinson's disease. Topological data analysis (persistence barcode) was used in conjunction with a certain standard machine learning tool (CNN) and the approach yielded improved results on Parkinson's disease phenotype prediction when measured against standard machine learning methods. The results in the paper thus confirm that gene expression can be a useful indicator of the presence or absence of a condition, and the subtle signal contained in this high dimensional data reveals itself when considering the intricate topological connections between expressed genes.

The input is sequencing-based gene expression values from blood samples. More precisely, the input is a matrix X of size whose rows correspond to subjects and whose columns correspond to genes. Each entry $X_{i,j}$ is the expression value of the j-th gene expression of the i-th subject.

Using a distance correlation matrix obtained from X (see Section 2.3 in [26]) for each individual and every pair of genes a value is computed which measures the pair-wise interaction of these two genes – and from these values one creates a Vietoris-Rips filtration of a simplicial comlpex (whose zero dimensional simplices are labelled by the genes). Using persistent homology computation one obtains for each individual their persistent homology barcodes. Actually, what is computed are persistence landscapes (this is a variant of barcode introduced in [4] and is particularly suitable to give at input to machine learning algorithm used in the second step of the pipeline). These landscapes are then used to train a convolution neural network to separate the healthy from the afflicted (see Figure 2 in [26] reproduced as Figure 2). This method of extracting the topological information first (i.e. computing the topological landscapes) and then using machine learning techniques was found to be more effective than just using machine learning methods on the raw data (i.e. on the matrix X).

The vertices of the Vietoris-Rips complex derived from the input data corresponds to genes and thus the filtration considered is a labelled filtration. We plan to augment the output of the persistent homology computation by adding to each bar $b = (s, t, 1)$ (assumed to be simple) unit vectors whose spans are the corresponding initial and terminal harmonic homology subspaces $\mathcal{P}_p^{s,t}(\mathcal{F}), \widehat{\mathcal{P}}_p^{s,t}(\mathcal{F})$ (Definition 3.23) associated to b. (Of course, we need to update the persistent homology software to compute these vectors and develop algorithms for computing them efficiently.) This should increase the amount of discerning information in the input to the ML algorithms in the second step of the pipeline. We postulate that doing so would improve the power of the method. The experimental results on whether such improvements are actually observed in practice is under investigation and will be reported in a subsequent paper.

The rest of the paper is organized as follows. In Section 2, we give the necessary mathematical background and define harmonic homology spaces and prove their basic properties, including a key theorem (Theorem 2.15) bounding the distance between harmonic homology subspaces of two simplicial complexes in terms of the difference between the two simplicial complexes. In Section 3, we define harmonic persistent homology, harmonic bar codes and prove three stability theorems (Theorems 3.18, 3.20, 3.24). Finally, in Section 4, we prove that harmonic representatives maximize relative essential content (Theorem 4.8).

2. Harmonic homology.

In this section we define harmonic homology spaces of simplicial complexes and prove their basic properties.

2.1. Simplicial complex.

We recall here some basic definitions and notation from simplicial homology theory.

Definition 2.1. *A* finite simplicial complex K *is a set of ordered subsets of* $[N] = \{0, \ldots N\}$ *for some* $N \geq 0$*, such that if* $\sigma \in K$ *and* τ *is a subset of* σ*, then* $\tau \in K$*.*

Notation 2.1. *If* $\sigma = \{i_0, \ldots, i_p\} \in K$*, with* K *a finite simplicial complex, and* $i_0 < \cdots < i_p$*, we will denote* $\sigma = [i_0, \ldots, i_p]$ *and call* σ *a* p-dimensional simplex of K*. We will denote by* $K^{(p)}$ *the subcomplex of* K *consisting of simplices of* K *of dimension* $\leq p$*. We will denote by* $K^{[p]} = K^{(p)} - K^{(p-1)}$ *the subset of* p*-dimensional simplices of* K*. Note that* $K^{[p]}$ *is not a subcomplex.*

Definition 2.2 (Chain groups). *Suppose* K *is a finite simplicial complex. For* $p \geq 0$*, we will denote by* $C_p(K) = C_p(K, \mathbb{R})$ *(the* p*-th chain group), the* $\mathbb{R}$*-vector space generated by the elements of* $K^{[p]}$*, i.e.*

$$C_p(K) = \bigoplus_{\sigma \in K^{[p]}} \mathbb{R} \cdot \sigma.$$

Definition 2.3 (The boundary map). *We denote by* $\partial_p(K) : C_p(K) \to C_{p-1}(K)$ *the linear map (called the* p*-th* boundary map*) defined as follows. Since* $(\sigma)_{\sigma \in K^{[p]}}$ *is a basis of* $C_p(K)$ *it is enough to define the image of each* $\sigma \in C_p(K)$*. We define for* $\sigma = [i_0, \ldots, i_p] \in K^{[p]}$*,* $\partial_p(K)(\sigma) = \sum_{0 \leq j \leq p} (-)^j [i_0, \ldots, \widehat{i_j}, \ldots, i_p] \in C_{p-1}(K)$*, where* $\widehat{\cdot}$ *denotes omission.*

Notation 2.2 (Cycles, boundaries, homology and the canonical surjection). *We denote* $Z_p(K) = \ker(\partial_p(K))$*, (the space of* p-dimensional cycles*),* $B_p(K) = \mathrm{Im}(\partial_{p+1}(K))$ *(the space of* p-dimensional boundaries*), and* $\mathrm{H}_p(K) = Z_p(K)/B_p(K)$ *(the* p-dimensional simplicial homology group *of* K*). We will denote by*

$$\phi_p(K) : Z_p(K) \to Z_p(K)/B_p(K) = \mathrm{H}_p(K)$$

the canonical surjection.

2.2. Representing homology classes by harmonic chains. Let K be a finite simplicial complex. We make the chain group $C_p(K)$ into an Euclidean space by fixing an inner product $\langle\cdot,\cdot\rangle_{C_p(K)}$. For the rest of the paper we fix the following inner product on $C_p(K)$ which we will refer to as the standard inner product on $C_p(K)$. We define:

$$\langle\sigma,\sigma'\rangle_{C_p(K)} = \delta_{\sigma,\sigma'}, \sigma,\sigma' \in K^{[p]} \tag{2.1}$$

(i.e. we declare the basis $(\sigma)_{\sigma\in K^{[p]}}$ to be an orthonormal basis). If the context is clear we will omit the subscript from the notation $\langle\cdot,\cdot\rangle_{C_p(K)}$.

We now come to a key definition – namely, that of harmonic homology (following [12]).

Definition 2.4 (Harmonic homology subspace). *For $p \geq 0$, we will denote*

$$\mathcal{H}_p(K) = Z_p(K) \cap B_p(K)^\perp$$

and call $\mathcal{H}_p(K) \subset C_p(K)$ the harmonic homology subspace *of K.*

2.2.1. Elementary properties. The following two propositions encapsulate the key properties of the harmonic homology subspaces.

Proposition 2.5. *The map $\mathfrak{f}_p(K)$ defined by*

$$z + B_p(K) \to \mathrm{proj}_{B_p(K)^\perp}(z), z \in Z_p(K) \tag{2.2}$$

gives an isomorphism $\mathfrak{f}_p(K) : \mathrm{H}_p(K) \to \mathcal{H}_p(K)$.

Proof. First observe that using the fact that $B_p(K) \subset Z_p(K)$, we have that for $z \in Z_p(K)$, $\mathrm{proj}_{B_p(K)^\perp}(z) \in Z_p(K)$, and so the map $\mathfrak{f}_p(K)$ is well defined. The injectivity and surjectivity of $\mathfrak{f}_p(K)$ are then obvious. ■

Proposition 2.6. $\mathcal{H}_p(K) = \ker(\partial_{p+1}(K)^*) \cap \ker(\partial_p(K))$ *(where L^* denotes the adjoint of a linear map L between two inner product spaces).*

Proof. From Definition 2.4 and the fact that $Z_p(K) = \ker(\partial_p)$, it suffices to prove that

$$\ker(\partial_{p+1}(K)^*) = B_p(K)^\perp.$$

For $z \in C_p(K)$,

$$\begin{aligned} z \in B_p(K)^\perp &\Leftrightarrow \langle z, z'\rangle_{C_p(K)} = 0 \text{ for all } z' \in B_p(K) \\ &\Leftrightarrow \langle z, \partial_{p+1}(w)\rangle_{C_p(K)} = 0 \text{ for all } w \in C_{p+1}(K) \\ &\Leftrightarrow \langle \partial^*_{p+1}(z), w\rangle_{C_{p+1}(K)} = 0 \text{ for all } w \in C_{p+1}(K) \\ &\Leftrightarrow z \in \ker(\partial^*_{p+1}). \end{aligned}$$

This completes the proof of the proposition. ■

Remark 2.7. The harmonic homology group $\mathcal{H}_p(K)$ as defined above is equal to the kernel of the linear map $\Delta_p = \partial_{p+1} \circ \partial^*_{p+1} + \partial^*_p \circ \partial_p$. The linear map $\Delta_p(K) : C_p(K) \to C_p(K)$ is a discrete analog of the Laplace operator and thus it makes sense to call its kernel the space of harmonic cycles.

Since the above description is often taken as a definition of harmonic homology groups we include the proof of the equivalence of the two definitions below.

Proposition 2.8. $\mathrm{Ker}(\Delta_p(K)) = \mathcal{H}_p(K)$.

Proof. It follows directly from Proposition 2.6 and the definition of $\Delta_p(K)$ that $\mathcal{H}_p(K) \subset \mathrm{Ker}(\Delta_p(K))$.

In order to prove the opposite inclusion. Suppose $z \in \mathrm{Ker}(\Delta_p(K))$. Then, $\partial_{p+1} \circ \partial^*_{p+1}(z) + \partial^*_p \circ \partial_p(z) = 0$.

Taking inner product with z we obtain

$$\langle z, \partial_{p+1} \circ \partial^*_{p+1}(z)\rangle_{C_p(K)} + \langle z, \partial^*_p \circ \partial_p(z)\rangle_{C_p(K)} = 0,$$

which using the defining property of adjoints gives

$$\langle \partial^*_{p+1}(z), \partial^*_{p+1}(z)\rangle_{C_{p+1}(K)} + \langle \partial_p(z), \partial_p(z)\rangle_{C_{p-1}(K)} = 0.$$

This implies that $z \in \mathrm{Ker}(\partial^*_{p+1}) \cap \mathrm{Ker}(\partial_p) = \mathcal{H}_p(K)$, the last equality being a consequence of Proposition 2.6. ■

2.2.2. Functoriality of the maps $\mathfrak{f}_p(K)$ under inclusion. Now suppose $K_1 \subset K_2$ are sub-complexes of the finite simplicial complex K. Then, $C_p(K_1)$ is a subspace of $C_p(K_2)$.

Proposition 2.9. *The restriction of* $\mathrm{proj}_{B_p(K_2)^\perp}$ *to* $\mathcal{H}_p(K_1)$ *gives a linear map*

$$\mathfrak{i}_p = \mathrm{proj}_{B_p(K_2)^\perp}|_{\mathcal{H}_p(K_1)} : \mathcal{H}_p(K_1) \to \mathcal{H}_p(K_2),$$

which makes the following diagram commute

$$\begin{tikzcd}
\mathrm{H}_p(K_1) \arrow[r, "i_p"] \arrow[d, "\mathfrak{f}_p(K_1)"] & \mathrm{H}_p(K_2) \arrow[d, "\mathfrak{f}_p(K_2)"] \\
\mathcal{H}_p(K_1) \arrow[r, "\mathfrak{i}_p"] & \mathcal{H}_p(K_2)
\end{tikzcd}$$

where $i_p : \mathrm{H}_p(K_1) \to \mathrm{H}_p(K_2)$ *is the map induced by the inclusion* $K_1 \hookrightarrow K_2$.

Before proving the proposition we first prove a very basic lemma.

Lemma 2.10. *Let* $B_1 \subset Z_1$, $B_2 \subset Z_2$ *be subspaces of an Euclidean vector space* V*, and suppose that* $B_1 \subset B_2$*, and* $Z_1 \subset Z_2$*. Then for* $z_1 \in Z_1$,

$$\mathrm{proj}_{B_2^\perp} \circ \mathrm{proj}_{B_1^\perp}(z_1) = \mathrm{proj}_{B_2^\perp}(z_1).$$

Proof. First observe that $B_1 \subset B_2$ implies that $B_2^\perp \subset B_1^\perp$.

Now let $w_1 = \mathrm{proj}_{B_1^\perp}(z_1)$. Then $w_1 = z_1 - \mathrm{proj}_{B_1}(z_1)$.

$$\begin{aligned}
\mathrm{proj}_{B_2^\perp} \circ \mathrm{proj}_{B_1^\perp}(z_1) &= \mathrm{proj}_{B_2^\perp}(w_1) \\
&= \mathrm{proj}_{B_2^\perp}(z_1) - \mathrm{proj}_{B_2^\perp}(\mathrm{proj}_{B_1}(z_1)) \\
&= \mathrm{proj}_{B_2^\perp}(z_1),
\end{aligned}$$

noting that $\mathrm{proj}_{B_2^\perp}(\mathrm{proj}_{B_1}(z_1)) = 0$, since (as noted earlier) $B_2^\perp \subset B_1^\perp$. ■

Proof of Proposition 2.9. Let $z_1 + B_p(K_1) \in \mathrm{H}_p(K_1)$. Using the definition of $\mathfrak{f}_p(K_1)$ (see Eqn. (2.2) in Proposition 2.5) we have

$$\mathfrak{f}_p(K_1)(z_1 + B_p(K_1)) = \mathrm{proj}_{B_p(K_1)^\perp}(z_1) \in \mathcal{H}_p(K_1) \subset B_p(K_1)^\perp.$$

Now observe that $B_p(K_1) \subset Z_p(K_1)$, $B_p(K_2) \subset Z_p(K_2)$, $B_p(K_1) \subset B_p(K_2)$, $Z_p(K_1) \subset Z_p(K_2)$. So

$$\begin{aligned}
\mathfrak{i}_p(K) \circ \mathfrak{f}_p(K_1)(z_1 + B_p(K_1)) &= \mathrm{proj}_{B_p(K_2)^\perp} \circ \mathrm{proj}_{B_p(K_1)^\perp}(z_1) \text{ (using (2.2))} \\
&= \mathrm{proj}_{B_p(K_2)^\perp}(z_1) \text{ (using Lemma 2.10)} \\
&= \mathfrak{f}_p(K_2) \circ i_p(z_1 + B_p(K_1)).
\end{aligned}$$

This proves the proposition. ■

2.3. Stability of harmonic homology subspaces. Suppose K is a finite simplicial complex and K_1, K_2 sub-complexes of K. In various applications one would want to compare the homology spaces of K_1 and K_2 quantitatively. In particular, one wants to say that if K_1 and K_2 are close under some natural metric then so are $\mathrm{H}_*(K_1)$ and $\mathrm{H}_*(K_2)$. Harmonic homology allows us to make rigorous this intuitive statement.

Since for each $p \geq 0$, $\mathcal{H}_p(K_1), \mathcal{H}_p(K_2)$ will be subspaces of $C_p(K)$, and thus correspond to points in $\mathrm{Gr}(b_p(K_1), C_p(K)), \mathrm{Gr}(b_p(K_2), C_p(K))$ respectively, (denoting by $b_p(K_i) = \dim \mathcal{H}_p(K_i)$) we first introduce a metric on the disjoint union of Grassmannians, $\coprod_{0 \leq d \leq \dim C_p(K)} \mathrm{Gr}(d, C_p(K))$.

2.3.1. Metric on Grassmannian. Let V be an Euclidean space, and for $0 \leq d \leq \dim V$ we will denote by $\mathrm{Gr}(d,V)$ the real Grassmannian variety of d-dimensional subspaces of V, and we will denote

$$\mathrm{Gr}(V) = \coprod_{0 \leq d \leq \dim V} \mathrm{Gr}(d,V).$$

Give two subspaces $A, B \subset V$, one way to measure how far away they are from each other is to take the square root of the sum of the squares of the principal angles between A and B. When the dimensions of A and B are equal this works well and produces a metric on the Grassmannian $\mathrm{Gr}(d,V)$ where $d = \dim A = \dim B$. However, if the dimensions of A and B are not necessarily equal, then this quantity can be zero even when $A \neq B$ (for example, if $A \subset B$). However, we would like to distinguish between such subspaces in our applications. In order to obtain a metric on the disjoint union $\mathrm{Gr}(V)$, one needs to tweak the above definition.

Following Lim and Ye [32], we now define a metric on $\mathrm{Gr}(V)$.

Definition 2.11. *For $A \in \mathrm{Gr}(k,V), B \in \mathrm{Gr}(\ell,V)$ with $0 \leq k, \ell \leq \dim V$, we define*

$$d_V(A,B) = \left(|k-\ell| \frac{\pi^2}{4} + \sum_{i=1}^{\min\{k,\ell\}} \theta_i^2 \right)^{1/2}$$

where θ_i is the i-th principal angle between A and B.

2.3.2. Principal angles in terms of singular values. The cosines of the principle angles between two subspaces of an Euclidean space V can be expressed in terms of the singular values of an associated matrix. Let P, Q be two subspaces of V. Fix an orthonormal basis $\mathbf{B}$ of V. Suppose $\dim P = p \geq q = \dim Q$. Let A, B be matrices whose columns are the coordinates with respect to $\mathbf{B}$ of some orthonormal bases of P and Q respectively. Let $U^T \Sigma V$ be the singular value decomposition of the $p \times q$ matrix $P^T Q$, with Σ a $p \times q$ matrix with diagonal entries the singular values $\sigma_1 \geq \cdots \geq \sigma_q \geq 0$.

Then for $1 \leq i \leq q$,

$$\sigma_i = \cos(\theta_i), \tag{2.3}$$

where θ_i is the i-th principle angle between P and Q.

Lemma 2.12. *Let V be an Euclidean space and $W_1, W_2 \subset V$ be subspaces. Let*

$$k = \max(\dim W_1, \dim W_2),$$

and

$$\ell = \dim(W_1 \cap W_2).$$

Then,

$$d_V(W_1, W_2) \leq \frac{\pi}{2} \cdot (k-\ell)^{1/2}.$$

Proof. Let $k_1 = \dim W_1, k_2 = \dim W_2$. It follows from Definition 2.11, and the definition of principle angles that

$$\begin{aligned} d_V(W_1, W_2) &\leq \frac{\pi}{2} \cdot (|k_1 - k_2| + \min(k_1, k_2) - \ell)^{1/2} \\ &= \frac{\pi}{2} \cdot (k-\ell)^{1/2}. \end{aligned}$$

We will also need the following lemma.

Lemma 2.13. *Let V be an Euclidean space and $W_1, W_2 \subset V$ be subspaces having dimensions p and q respectively, with $p \geq q$. Let $L \subset W_2$ be an 1-dimensional subspace. Let θ be the principal angle between L and W_1, and $\theta_1 \leq \cdots \leq \theta_q$ be the principle angles between W_1 and W_2. Then,*

$$\theta_1 \leq \theta \leq \theta_q.$$

Proof. The first inequality follows from the variational characterization of the principle angles. We prove the second inequality. Let $\{e_1, \ldots, e_p\}$ be an orthonormal basis of W_1. Let $f = f_1$ be a unit length vector such that $\mathrm{span}(f) = L$, and let $\{f_1, f_2, \ldots, f_q\}$ be an orthonormal basis of W_2. Let P and Q be matrices whose columns are the coordinates of the e_i's and the f_i's respectively. Let Q' be the submatrix of Q with consisting of the first column of Q (i.e. the column of coordinates of f).

Let $M = P^T Q, M' = P^T Q'$. Then, M' is a $p \times 1$ submatrix of the $p \times q$ matrix M. Let σ be the singular value of M', and $\sigma_1 \geq \cdots \geq \sigma_q$ the singular values of M. It follows from Eqn. (2.3) that

$$\sigma = \cos\theta,$$

$$\sigma_i = \cos\theta_i, 1 \leq i \leq q.$$

Finally it follows from the interlacing inequality [31, Theorem 10, Inequality (10)] that $\sigma = \cos\theta \geq \sigma_q = \cos\theta_q$, from which it follows that $\theta \leq \theta_q$, since $0 \leq \theta, \theta_q \leq \pi/2$. ∎

2.3.3. Stability theorem. Using harmonic homology spaces we now define a distance function between the homology of two sub-complexes of a fixed simplicial complex in any fixed dimension.

Let K be a finite simplicial complex and K_1, K_2 be two sub-complexes.

Definition 2.14. *We define*

$$d_{K,p}(K_1, K_2) = d_{C_p(K)}(\mathcal{H}_p(K_1), \mathcal{H}_p(K_2)).$$

We are now in a position to make quantitative the intuitive idea that two sub-complexes of a fixed simplicial complex which are close to each other should have homology spaces that are also close. We prove the following theorem.

Theorem 2.15 (Stability of harmonic homology). *Let K be a finite simplicial complex and K_1, K_2 sub-complexes of K. Then, for each $p \geq 0$,*

$$d_{K,p}(K_1, K_2) \leq \frac{\pi}{2} \cdot \Delta_p(K_1, K_2)^{1/2} \tag{2.4}$$

where $\Delta_p(K_1, K_2)$ is defined to be the maximum of

$$\mathrm{card}\left(K_1^{[p]} - K_2^{[p]}\right) + \mathrm{card}\left(K_2^{[p+1]} - K_1^{[p+1]}\right),$$

and

$$\mathrm{card}\left(K_2^{[p]} - K_1^{[p]}\right) + \mathrm{card}\left(K_1^{[p+1]} - K_2^{[p+1]}\right).$$

Before proving Theorem 2.15 we first prove a lemma that we will need in the proof of Theorem 2.15 and also later in the paper.

Lemma 2.16. *Let K be a finite simplicial complex and K_1, K_2 sub-complexes of K. Then, for each $p \geq 0$,*

$$\dim \mathcal{H}_p(K_1) - \dim(\mathcal{H}_p(K_1) \cap \mathcal{H}_p(K_2)) \leq \mathrm{card}\left(K_1^{[p]} - K_2^{[p]}\right) + \mathrm{card}\left(K_2^{[p+1]} - K_1^{[p+1]}\right), \tag{2.5}$$

$$\dim \mathcal{H}_p(K_2) - \dim(\mathcal{H}_p(K_1) \cap \mathcal{H}_p(K_2)) \leq \mathrm{card}\left(K_2^{[p]} - K_1^{[p]}\right) + \mathrm{card}\left(K_1^{[p+1]} - K_2^{[p+1]}\right). \tag{2.6}$$

Proof. We prove the inequality (2.5). The proof of inequality (2.6) is similar.

For each $p \geq 0$, and $i = 1, 2$, let $M_p(K_i)$ denote the matrix corresponding to the boundary map $\partial_p : C_p(K_i) \to C_{p-1}(K_i)$ with respect to the orthonormal bases $\mathcal{A}_p(K_i) = (\sigma)_{\sigma \in K_i^{[p]}}$ of $C_p(K_i)$ and $\mathcal{A}_{p-1}(K_i) = (\sigma)_{\sigma \in K_i^{[p-1]}}$ of $C_{p-1}(K_i)$.

Note that the rows of $M_p(K_i)$ are indexed by $K_i^{[p-1]}$ and the columns of $M_p(K_i)$ are indexed by $K_i^{[p]}$.

Observe that $\mathcal{H}_p(K_1) \subset C_p(K_1)$ is the intersection of the nullspaces of the matrices $M_p(K_1)$

and $M_{p+1}(K_1)^T$ i.e.

$$z \in \mathcal{H}_p(K_1) \Leftrightarrow [z]_{\mathcal{A}_p(K_1)} \in \mathrm{null}(M_p(K_1)) \cap \mathrm{null}(M_{p+1}(K_1)^T).$$

The subspace $\mathcal{H}_p(K_1) \cap \mathcal{H}_p(K_2)$ of $\mathcal{H}_p(K_1)$ is cut out of $\mathcal{H}_p(K_1)$ by additional equations. Let M'_p be the matrix whose columns are indexed by $K_1^{[p]}$ and whose rows are indexed by $K_1^{[p-1]} \cup (K_1^{[p]} - K_2^{[p]})$ defined by

$$\begin{aligned}(M'_p)_{\sigma,\sigma'} &= M_p(K_1)_{\sigma,\sigma'} \text{ if } \sigma \in K_1^{[p-1]}, \sigma' \in K_1^{[p]},\\ &= 1, \text{ if } \sigma = \sigma' \in K_1^{[p]} - K_2^{[p]},\\ &= 0, \text{ otherwise.}\end{aligned}$$

Similarly, Let M'_{p+1} be the matrix whose columns are indexed by $K_1^{[p+1]} \cup K_2^{[p+1]}$ and whose rows are indexed by $K_1^{[p]}$ be defined by

$$\begin{aligned}(M'_{p+1})_{\sigma,\sigma'} &= M_{p+1}(K_1)_{\sigma,\sigma'}, \text{ if } \sigma \in K_1^{[p]}, \sigma' \in K_1^{[p+1]},\\ &= M_{p+1}(K_2)_{\sigma,\sigma'}, \text{ if } \sigma \in K_1^{[p]}, \sigma' \in K_2^{[p+1]} - K_1^{[p+1]}.\end{aligned}$$

Then,

$$z \in \mathcal{H}_p(K_1) \cap \mathcal{H}_p(K_2) \Leftrightarrow [z]_{\mathcal{A}_p(K_1)} \in \mathrm{null}(M'_p) \cap \mathrm{null}(M'^T_{p+1}).$$

Observe that M'_p contains $M_p(K_1)$ as a submatrix and has $\mathrm{card}(K_1^{[p]} - K_2^{[p]})$ extra rows. Similarly, M'^T_{p+1} has $M_{p+1}(K_1)^T$ as a submatrix and has $\mathrm{card}(K_2^{[p+1]} - K_1^{[p+1]})$ extra rows.

It follows that the codimension of $\mathcal{H}_p(K_1) \cap \mathcal{H}_p(K_2)$ in $\mathcal{H}_p(K_1)$ is bounded by

$$\mathrm{card}\left(K_1^{[p]} - K_2^{[p]}\right) + \mathrm{card}\left(K_2^{[p+1]} - K_1^{[p+1]}\right),$$

which completes the proof of inequality (2.5). ∎

Proof of Theorem 2.15. The theorem now follows from Lemma 2.12, and inequalities (2.5), (2.6) in Lemma 2.16. ∎

Corollary 2.17. *With the same assumptions as in Theorem 2.15:*

$$d^2_{K,p}(K_1, K_2) \le \frac{\pi^2}{4} \cdot \left(\sum_{\sigma \in K^{[p]} \cup K^{[p+1]}} |\chi_{K_1}(\sigma) - \chi_{K_2}(\sigma)| \right), \tag{2.7}$$

where for any sub-complex K' of K we denote by $\chi_{K'}(\cdot)$ the characteristic function of K' (considered as a set of simplices).

Proof. It is easy to see that the quantity $\Delta_p(K_1, K_2)$ defined in Theorem 2.15 is bounded from above by

$$\sum_{\sigma \in K^{[p]} \cup K^{[p+1]}} |\chi_{K_1}(\sigma) - \chi_{K_2}(\sigma)|.$$

The corollary is now an immediate consequence of Theorem 2.15 after squaring both sides of the inequality (2.4). ∎

Remark 2.18. The bound in Theorem 2.15 is a bit crude in that it only depends on the dimension of the intersection of the two subspaces $\mathcal{H}_p(K_1)$ and $\mathcal{H}_p(K_2)$. However, even if the dimension of the intersection stays constant, the principal angles between the two subspaces give a measure of how "close" these two subspaces are. The following example is instructive.

Example 2.1. *Let $n > 0$ be a fixed integer, and for $m \le n$ let $K_{m,n}$ denote the one dimensional simplicial complex depicted in Figure 3. Then, $\mathcal{H}_1(K_{m,n})$ is a two-dimensional subspace of $C_1(K)$, where K denotes the union of all the $K_{m,n}$ as sub-complexes.*

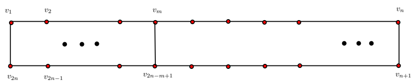

Figure 3: Convergence of harmonic homology subspaces

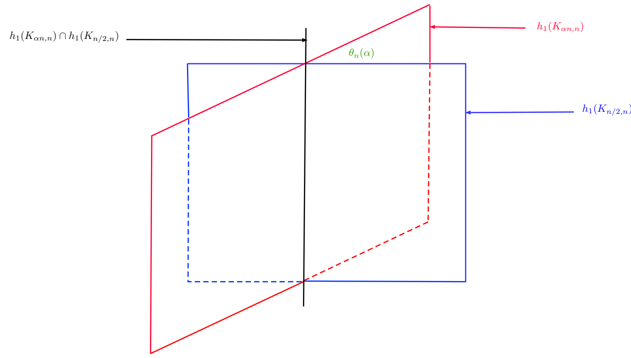

Figure 4: The subspaces $\mathcal{H}_1(K_{\alpha n,n})$ and $\mathcal{H}_1(K_{n/2,n})$.

Now, the harmonic homology spaces $\mathcal{H}_1(K_{m,n})$ have a one-dimensional subspace in common – namely, the subspace spanned by the vector

$$w = [v_1, v_2] + [v_2, v_3] + \cdots + [v_{2n-1}, v_{2n}] + [v_{2n}, v_1].$$

For $0 < \alpha < \frac{1}{2}$, we consider the subspaces $\mathcal{H}_1(K_{\lfloor \alpha n \rfloor,n})$ and $\mathcal{H}_1(K_{\lfloor n/2 \rfloor,n})$. Note there is at most one non-zero principal angle between these subspaces since they have a one-dimensional subspace in common. Let $\theta_n(\alpha)$ denote this principal angle. Intuitively, one should expect that as $\alpha \to 1/2$, the subspaces $\mathcal{H}_1(K_{\lfloor \alpha n \rfloor,n})$ and $\mathcal{H}_1(K_{\lfloor n/2 \rfloor,n})$ should come closer to each other and the non-zero principal angle between them should go to 0. Let $K_1 = K_{\lfloor \alpha n \rfloor,n}$ and $K_2 = K_{\lfloor n/2 \rfloor,n}$. Since we are going to take the limit as n goes to infinity, the floor function becomes irrelevant. With this caveat a standard calculation which we omit shows that

$$\cos(\theta_n(\alpha)) = \frac{\alpha n}{(2\alpha n(1-\alpha)^2 + (2n)(1-\alpha)\alpha^2 + 1)^{\frac{1}{2}} \left(\frac{n}{2} + 1\right)^{\frac{1}{2}}}.$$

This implies that $\lim_{n\to\infty} \cos\theta_n(\alpha) = \left(\frac{\alpha}{1-\alpha}\right)^{1/2}$, which agrees with the intuition expressed before, since clearly the angle $\theta_n(\alpha)$ goes to 0, as α approaches 1/2.

3. Harmonic persistent homology and harmonic barcodes of a filtration. In this section we define harmonic persistent homology and the associated harmonic persistent barcodes and prove stability results. We first recall the notion of persistent homology and also the definition of the associated barcodes. There is some subtlety in the definition of the latter that is explained in the following section.

3.1. Persistent homology and barcodes. Let T be an ordered set (with or without endpoints), and $\mathcal{F} = (K_t)_{t\in T}$, a tuple of sub-complexes of a finite simplicial complex K, such that $s \leq t \Rightarrow K_s \subset K_t$. We call $\mathcal{F}$ a filtration of the simplicial complex K.

Convention 3.1. *If the ordered set T has end points, $a = \min(T), b = \max(T)$ (for example if T is finite), then we will use formally adjoin two new elements $-\infty$ and ∞ to T with $-\infty < c, \infty > c$, for every $c \in T$, and adopt the convention that $K_{-\infty} = \emptyset$ and $K_\infty = K_b$ (or equivalently that $K_s = \emptyset$ for $s < a$ and $K_t = K_b$ for $t > b$).*

Notation 3.1. *For $s, t \in T, s \leq t$, and $p \geq 0$, we let $i_p^{s,t} : \mathrm{H}_p(K_s) \longrightarrow \mathrm{H}_p(K_t)$, denote the homomorphism induced by the inclusion $K_s \hookrightarrow K_t$.*

Definition 3.2 (Persistent homology groups). *[13] For each triple $(p, s, t) \in \mathbb{Z}_{\geq 0} \times T \times T$ with $s \leq t$ the* persistent homology group, *$\mathrm{H}_p^{s,t}(\mathcal{F})$ is defined by*

$$\mathrm{H}_p^{s,t}(\mathcal{F}) = \mathrm{Im}(i_p^{s,t}).$$

Note that $\mathrm{H}_p^{s,t}(\mathcal{F}) \subset \mathrm{H}_p(K_t)$, *and* $\mathrm{H}_p^{s,s}(\mathcal{F}) = \mathrm{H}_p(K_s)$.

Notation 3.2 (Persistent Betti numbers). *We denote by* $b_p^{s,t}(\mathcal{F}) = \dim_{\mathbb{R}}(\mathrm{H}_p^{s,t}(\mathcal{F}))$, *and call* $b_p^{s,t}(\mathcal{F})$ (p,s,t)*-th persistent Betti number.*

We now define barcode of a filtration.

3.1.1. Barcodes of filtrations.

Definition 3.3. *In the lingua franca of the theory of persistent homology, for* $s \leq t \in T$, *and* $p \geq 0$,

- *we say that a homology class* $\gamma \in \mathrm{H}_p(K_s)$ *is* born *at time* s, *if* $\gamma \notin \mathrm{H}_p^{s',s}(\mathcal{F})$, *for any* $s' < s$;
- *for a class* $\gamma \in \mathrm{H}_p(K_s)$ *born at time* s, *we say that* γ dies *at time* t, *if* $i_p^{s,t'}(\gamma) \notin \mathrm{H}_p^{s',t'}(\mathcal{F})$ *for all* s', t' *such that* $s' < s < t' < t$, *but* $i_p^{s,t}(\gamma) \in \mathrm{H}_p^{s'',t}(\mathcal{F})$, *for some* $s'' < s$.

First observe that it follows from Definition 3.2 that for all $s' \leq s \leq t$ and $p \geq 0$, $\mathrm{H}_p^{s',t}(\mathcal{F})$ is a subspace of $\mathrm{H}_p^{s,t}(\mathcal{F})$, and both are subspaces of $\mathrm{H}_p(K_t)$. This is because the homomorphism $i_p^{s',t} = i_p^{s,t} \circ i_p^{s',s}$, and so the image of $i_p^{s',t}$ is contained in the image of $i_p^{s,t}$. It follows that, for $s \leq t$, the union $\bigcup_{s'<s} \mathrm{H}_p^{s',t}(\mathcal{F})$ is an increasing union of subspaces, and is itself a subspace of $\mathrm{H}_p(K_t)$. In particular, setting $t = s$, $\bigcup_{s'<s} \mathrm{H}^{s',s}(\mathcal{F})$ is a subspace of $\mathrm{H}_p(K_s)$.

The following definitions are taken from [2] (see also the references therein and [17, Theorem 1]). We follow the same notation as above and first define certain subspaces of the homology groups $\mathrm{H}_p(K_s), s \in T, p \geq 0$.

Definition 3.4 (Subspaces of $\mathrm{H}_p(K_s)$). *For* $s \leq t$, *and* $p \geq 0$, *we define*

$$M_p^{s,t}(\mathcal{F}) = \bigcup_{s'<s} (i_p^{s,t})^{-1}(\mathrm{H}_p^{s',t}(\mathcal{F})),$$

$$M_p^{s,\infty}(\mathcal{F}) = \bigcup_{s \leq t'} M_p^{s,t'}(\mathcal{F}),$$

$$N_p^{s,t}(\mathcal{F}) = \bigcup_{s'<s\leq t'<t} (i_p^{s,t'})^{-1}(\mathrm{H}_p^{s',t'}(\mathcal{F})).$$

The "meaning" of these subspaces in terms of birth and death of homological cycles (as per Definition 3.3) is encapsulated in the following proposition.

Proposition 3.5. 1. *For every* $s, t \in T, s \leq t$, $M_p^{s,t}(\mathcal{F})$ *is a subspace of* $\mathrm{H}_p(K_s)$ *consisting of homology classes in* $\mathrm{H}_p(K_s)$ *which are either* 0 *or*

"are either born before time s, *or born at time* s *and die at time* t *or earlier".*

2. *For every fixed* $s \in T$, $M_p^{s,\infty}(\mathcal{F})$ *is a subspace of* $\mathrm{H}_p(K_s)$ *consisting of homology classes in* $\mathrm{H}_p(K_s)$ *which are either* 0

"are either born before time s, *or born at time* s *and die at some time* $t \geq s$".

3. *Similarly, for every* $s, t \in T, s \leq t$, $N_p^{s,t}(\mathcal{F})$ *is a subspace of* $\mathrm{H}_p(K_s)$ *consisting of homology classes in* $\mathrm{H}_p(K_s)$ *which are either* 0 *or*

"are either born before time s, *or born at time* s *and die strictly earlier than* t".

Proof. 1. Let $0 \neq \gamma \in M_p^{s,t}(\mathcal{F})$.
Since $M_p^{s,t}(\mathcal{F}) \subset \mathrm{H}_p(K_s)$, and every element of $\mathrm{H}_p(K_s)$ is born at time s or earlier, it is clear that γ is born at time s or earlier.
If $s = \min(T)$, then using Convention 3.1 $\mathrm{H}_p^{s',t}(\mathcal{F}) = 0$ for $s' < s$, and $\gamma \in \ker(i_p^{s,t})$. This implies that γ is born at time s and dies at or before time t.

If $s \neq \min(T)$, then it follows from Definition 3.4 that

$$\gamma \in M_p^{s,t}(\mathcal{F}) \Leftrightarrow \exists s' < s \text{ such that } i_p^{s,t}(\gamma) \in \mathrm{H}_p^{s',t}(\mathcal{F}) = \mathrm{Im}(i_p^{s',t}).$$

This implies as per Definition 3.3 that
Now $i_p^{s',t} = i_p^{s,t} \circ i_p^{s',s}$. So there exists $s' < s$ and $\gamma' \in \mathrm{H}_p(K_{s'})$ such that

$$i_p^{s',t}(\gamma') = i_p^{s,t} \circ i_p^{s',s}(\gamma') = i_p^{s,t}(\gamma). \tag{3.1}$$

If there exists γ' satisfying (3.1) such that $i_p^{s',s}(\gamma') = \gamma$ then γ is born before time s.
Else, s is born at time s. Moreover, the image of γ in $\mathrm{H}_p^{s,t}(\mathcal{F})$ equals the image of γ' in $\mathrm{H}_p^{s,t}(\mathcal{F})$ for some $\gamma' \in \mathrm{H}_p(K_{s'})$. It follows that γ is born at time s and dies at time t or earlier.
Conversely, suppose that $\gamma \in \mathrm{H}_p(K_s)$, and either γ is born before time s, or γ is born at time s and dies at time t or earlier.
There are two possibilities. Either γ is born before time s. In this case, there exists $s' < s$ and $\gamma' \in \mathrm{H}_p(K_{s'})$ such that $i_p^{s,t}(\gamma) = i_p^{s',t}(\gamma')$ which implies that $\gamma \in M_p^{s,t}(\mathcal{F})$.
Or, γ is born at time s and dies at time t or earlier. In this case, $i_p^{s,t}(\gamma) = 0$, which implies that $\gamma \in M_p^{s,t}(\mathcal{F})$.
This proves Part (1).

2. Part (2) is follows directly from Part (1).
3. It is clear from the definition that each $\gamma \in N_p^{s,t}(\mathcal{F})$ belongs to $M_p^{s,t'}(\mathcal{F})$ for some t' with $s \leq t' < t$. Part (3) now follows from Part (1). ■

We now define certain *subquotients* of the homology groups of $\mathrm{H}_p(K_s), s \in T, p \geq 0$, in terms of the subspaces defined above in Definition 3.4. These subquotients (which are vector spaces and therefore have well defined dimensions) are in bijection with equivalence classes of homology classes "born at time s and which die at time t".

Definition 3.6 (Subquotients associated to a filtration). *For $s \leq t$, and $p \geq 0$, we define*

$$P_p^{s,t}(\mathcal{F}) = M_p^{s,t}(\mathcal{F})/N_p^{s,t}(\mathcal{F}),$$
$$P_p^{s,\infty}(\mathcal{F}) = \mathrm{H}_p(K_s)/M_p^{s,\infty}(\mathcal{F}).$$

We will call

1. $P_p^{s,t}(\mathcal{F})$ *the* space of p-dimensional cycles born at time s and which dies at time t*; and*
2. $P_p^{s,\infty}(\mathcal{F})$ *the* space of p-dimensional cycles born at time s and which never die.

Definition 3.7 (Persistent multiplicity, barcode, simple bars). *We will denote for $s \in T, t \in T \cup \{\infty\}$,*

$$\mu_p^{s,t}(\mathcal{F}) = \dim P_p^{s,t}(\mathcal{F}), \tag{3.2}$$

and call $\mu_p^{s,t}(\mathcal{F})$ the persistent multiplicity of p-dimensional cycles born at time s and dying at time t if $t \neq \infty$, or never dying in case $t = \infty$.

Finally, we will call the set

$$\mathbf{B}_p(\mathcal{F}) = \{(s, t, \mu_p^{s,t}(\mathcal{F})) \mid \mu_p^{s,t}(\mathcal{F}) > 0\}$$

the p-dimensional barcode associated to the filtration $\mathcal{F}$.

We will call an element $b = (s, t, \mu_p^{s,t}(\mathcal{F})) \in \mathbf{B}_p(\mathcal{F})$ a bar of $\mathcal{F}$ of multiplicity $\mu_p^{s,t}(\mathcal{F})$*. If $\mu_p^{s,t}(\mathcal{F}) = 1$, we will call b a* simple *bar.*

3.2. Harmonic persistent homology and harmonic barcodes. We now define the harmonic versions of persistent homology groups. These will all be subspaces of $C_p(K), p \geq 0$.

Using Proposition 2.9 we have for each $s, t \in T, s \leq t$, a linear map

$$\mathrm{i}_p^{s,t} := \mathrm{proj}_{B_p(K_t)^\perp}|_{\mathcal{H}_p(K_s)} : \mathcal{H}_p(K_s) \to \mathcal{H}_p(K_t),$$

which makes the following diagram commute

$$\begin{tikzcd} \mathrm{H}_p(K_s) \arrow[r, "i_p^{s,t}"] \arrow[d, "\mathfrak{f}_p(K_s)"] & \mathrm{H}_p(K_t) \arrow[d, "\mathfrak{f}_p(K_t)"] \\ \mathcal{H}_p(K_s) \arrow[r, "\mathrm{i}_p^{s,t}"] & \mathcal{H}_p(K_t) \end{tikzcd} . \tag{3.3}$$

Definition 3.8 (Harmonic persistent homology subspaces). *For each triple* $(p, s, t) \in \mathbb{Z}_{\geq 0} \times T \times T$ *with* $s \leq t$ *the* harmonic persistent homology subspace, $\mathcal{H}_p^{s,t}(\mathcal{F})$ *is defined by*

$$\mathcal{H}_p^{s,t}(\mathcal{F}) = \mathrm{Im}(\mathrm{i}_p^{s,t}(\mathcal{F})) \subset C_p(K).$$

3.2.1. Harmonic barcodes of filtrations. We now give the harmonic analogs of the above spaces. They are all subspaces of $C_p(K)$ (in fact, of the various harmonic homology spaces $\mathcal{H}_p(K_s), s \in T$).

Definition 3.9 (Harmonic barcode of a filtration). *For* $s \leq t$*, and* $p \geq 0$*, we define*

$$\mathcal{M}_p^{s,t}(\mathcal{F}) = \bigcup_{s'<s} (\mathrm{i}_p^{s,t})^{-1}(\mathcal{H}_p^{s',t}(\mathcal{F})),$$

$$\mathcal{N}_p^{s,t}(\mathcal{F}) = \bigcup_{s'<s\leq t'<t} (\mathrm{i}_p^{s,t'})^{-1}(\mathcal{H}_p^{s',t'}(\mathcal{F})),$$

$$\mathcal{P}_p^{s,t}(\mathcal{F}) = \mathcal{M}_p^{s,t}(\mathcal{F}) \cap \mathcal{N}_p^{s,t}(\mathcal{F})^{\perp},$$

$$\mathcal{P}_p^{s,\infty}(\mathcal{F}) = \mathcal{H}_p(K_s) \cap \bigcap_{s\leq t} \mathcal{M}_p^{s,t}(\mathcal{F})^{\perp}.$$

The vector spaces defined in Definition 3.9 are isomorphic to the corresponding ones in Definitions 3.4 and 3.6. Following the same notation as in Definition 3.9 we have the following proposition.

Proposition 3.10.

$$\mathcal{M}_p^{s,t}(\mathcal{F}) \cong M_p^{s,t}(\mathcal{F}, \tag{3.4}$$

$$\mathcal{N}_p^{s,t}(\mathcal{F}) \cong N_p^{s,t}(\mathcal{F}), \tag{3.5}$$

$$\mathcal{P}_p^{s,t}(\mathcal{F}) \cong P_p^{s,t}(\mathcal{F}), \tag{3.6}$$

$$\mathcal{P}_p^{s,\infty}(\mathcal{F}) \cong P_p^{s,\infty}(\mathcal{F}). \tag{3.7}$$

Proof. The isomorphisms in (3.4) and (3.5) follow from the fact that the vertical arrows in the commutative diagram (3.3) are isomorphisms (Proposition 2.5).

In order to prove the isomorphism in (3.6), first observe that $\mathcal{N}_p^{s,t}(\mathcal{F})$ is a subspace of $\mathcal{M}_p^{s,t}(\mathcal{F})$. The isomorphism now follows from the following simple observation. If $W_1 \subset W_2$ are subspaces of an Euclidean space V, then $W_1^{\perp} \cap W_2$ equals the orthogonal complement of W_1 in W_2 (restricting the inner product of V to W_2).

Finally, the isomorphism in (3.7) follows from the fact that $\bigcup_{s\leq t} \mathcal{M}_p^{s,t}(\mathcal{F})$ is a subspace of $\mathcal{H}_p(K_s)$, and $\left(\bigcup_{s\leq t} \mathcal{M}_p^{s,t}(\mathcal{F})\right)^{\perp} = \bigcap_{s\leq t} \mathcal{M}_p^{s,t}(\mathcal{F})^{\perp}$, and the previous observation. ■

In analogy with Definition 3.7 we now define harmonic barcodes of filtrations.

Definition 3.11 (Harmonic barcodes). *We will call the set*

$$\mathcal{B}_p(\mathcal{F}) = \{(s, t, \mathcal{P}_p^{s,t}(\mathcal{F})) \mid \mathcal{P}_p^{s,t}(\mathcal{F})) \neq 0\}$$

the p-dimensional harmonic barcode associated to the filtration $\mathcal{F}$. *For* $b = (s, t, \mu_p^{s,t}(\mathcal{F})) \in B_p(\mathcal{F})$*, we will call the subspace* $\mathcal{P}_p^{s,t}(\mathcal{F}) \subset \mathcal{H}_p(K_s)$*, the harmonic homology subspace associated to* b.

Remark 3.12. Note that

$$\mathbf{B}_p(\mathcal{F}) \subset T \times (T \cup \{\infty\}) \times \mathbb{Z}_{>0},$$

and

$$\mathcal{B}_p(\mathcal{F}) \subset T \times (T \cup \{\infty\}) \times \coprod_{0 \le d \le \dim C_p(K)} \mathrm{Gr}(d, C_p(K)).$$

3.3. Harmonic persistent homology spaces of filtrations induced by functions. We now consider filtrations of finite simplicial complexes induced by functions. Our goal is to prove that the harmonic persistent homology spaces of such filtrations are stable with respect to perturbations of of the functions defining them.

3.3.1. Filtration induced by functions.

Definition 3.13 (Admissible functions). *Let K be a finite simplicial complex and $f : K \to \mathbb{R}$. We say that f is an* admissible function *if f satisfies for each $\sigma, \tau \in K$, with $\sigma \prec \tau$, $f(\sigma) < f(\tau)$.*

Notation 3.3 (Filtration induced by an admissible function). *If $f : K \to \mathbb{R}$ is an admissible function, then for each $t \in \mathbb{R}$, $K_{f \le t} = f^{-1}(-\infty, t]$ is a sub-complex of K, and $(K_{f \le t})_{t \in \mathbb{R}}$ is a filtration with respect to the usual order of $\mathbb{R}$. We will call this filtration* the filtration induced by f *and denote it by $\mathcal{F}_f$.*

Definition 3.14 (Harmonic filtration function). *Let K be a finite simplicial complex, and let $f : K \to \mathbb{R}$ be an admissible function.*

For each $p \ge 0$ we define:

$$\mathcal{H}_p(K, f) : \mathbb{R} \to \mathrm{Gr}(C_p(K)) = \coprod_{0 \le d \le \dim C_p(K)} \mathrm{Gr}(d, C_p(K))$$

$$t \mapsto \mathcal{H}_p(K_{f \le t}).$$

We will denote $\mathcal{H}_p(K, f)(t)$ by $\mathcal{H}_p^t(K, f)$. We call $\mathcal{H}_p(K, f)$ a harmonic filtration function of dimension p.

Remark 3.15. Notice that the harmonic filtration function $\mathcal{H}_p(K, f)$ contains all the information of persistent homology. Once $\mathcal{H}_p(K, f)$ is known, for $s \le t$ we have

$$\mathrm{proj}_{\mathcal{H}_p(K,f)(t)}(\mathcal{H}_p(K, f)(s)) = \mathcal{H}_p^{s,t}(\mathcal{F}_f) \cong \mathrm{H}_p^{s,t}(\mathcal{F}_f).$$

This was already observed by Lieutier (see [23, Lemma 2]).

3.3.2. Metric on the space of harmonic filtration functions. Observe that for any finite dimensional real vector space V, and for $d \ge 0$, the Grassmannian $\mathrm{Gr}(d, V)$ has the structure of a semi-algebraic set – and hence, $\mathrm{Gr}(V) = \coprod_{0 \le d \le \dim V} \mathrm{Gr}(d, V)$ is also a semi-algebraic set. For a finite simplicial complex K, and $p \ge 0$, we denote by $\mathcal{S}_p(K)$ the set of piece-wise constant maps with at most finitely many discontinuities $F : \mathbb{R} \to \mathrm{Gr}(C_p(K))$.

We now introduce a class of (pseudo)-metrics on $\mathcal{S}_p(K)$.

Definition 3.16. *Let $\ell > 0$. For $F, G \in \mathcal{S}_p(K)$ we define*

$$\mathrm{dist}_{K,p,\ell}(F, G) = \left(\int_{\mathbb{R}} d_{K,p}(F(t), G(t))^\ell \, dt \right)^{1/\ell}. \tag{3.8}$$

We also define a family of semi-norms on the space of functions $f : K \to \mathbb{R}$.

Definition 3.17. *Let $\ell > 0$. Given a function $f : K \to \mathbb{R}$ we define*

$$||f||_\ell^{(p)} = \left(\sum_{\sigma \in K^{[p]}} |f(\sigma)|^\ell \right)^{1/\ell}. \tag{3.9}$$

We are now in a position to state our first stability result.

Theorem 3.18. *Let K be a finite simplicial complex, $f, g : K \to \mathbb{R}$ be admissible functions and $p \geq 0$. Then,*

$$\mathrm{dist}_{K,p,2}(\mathcal{H}_p(K,f), \mathcal{H}_p(K,g)) \leq \frac{\pi}{2} \cdot \left(||f-g||_1^{(p)} + ||f-g||_1^{(p+1)} \right)^{1/2}.$$

Proof. Using Corollary 2.17 we have

$$\begin{aligned}
\mathrm{dist}_{K,p,2}(\mathcal{H}_p(K,f), \mathcal{H}_p(K,g))^2 &= \int_{\mathbb{R}} d_{K,p}(\mathcal{H}_p^t(K,f), \mathcal{H}_p^t(K,g))^2 dt \\
&\leq \frac{\pi^2}{4} \cdot \int_{\mathbb{R}} \left(\sum_{\sigma \in K^{[p]} \cup K^{[p+1]}} |\chi_{K_{f \leq t}}(\sigma) - \chi_{K_{g \leq t}}(\sigma)| dt \right) \\
&= \frac{\pi^2}{4} \cdot \left(\sum_{\sigma \in K^{[p]} \cup K^{[p+1]}} \int_{\mathbb{R}} |\chi_{K_{f \leq t}}(\sigma) - \chi_{K_{g \leq t}}(\sigma)| dt \right) \\
&= \frac{\pi^2}{4} \cdot \left(\sum_{\sigma \in K^{[p]}} |f(\sigma) - g(\sigma)| + \sum_{\sigma \in K^{[p]+1}} |f(\sigma) - g(\sigma)| \right) \\
&= \frac{\pi^2}{4} \cdot \left(||f-g||_1^{(p)} + ||f-g||_1^{(p+1)} \right).
\end{aligned}$$

The theorem follows. ■

We prove now a similar stability theorem for the harmonic persistent homology spaces

$$\mathcal{H}_p^{s,t}(K,f) := \mathcal{H}_p^{s,t}(\mathcal{F}_f), \mathcal{H}_p^{s,t}(K,g) := \mathcal{H}_p^{s,t}(\mathcal{F}_g).$$

For convenience (and in order to make sure that the distance does not blow up for trivial reasons) we will only consider admissible functions $f : K \to \mathbb{R}$ taking values in $[0,1]$. This is not a serious restriction, since K is a finite complex and one can always scale and translate functions without affecting the induced finite filtration of K other than the index.

We define:

Definition 3.19. *Let $\ell > 0$. For $F, G \in \mathcal{S}_p(K)$ we define*

$$\mathrm{dist}_{\mathcal{H},p,\ell}^{\mathrm{persistent}}(F,G) = \left(\int_0^1 \int_0^t d_{K,p}(F^{s,t}, G^{s,t})^\ell ds dt \right)^{1/\ell}, \tag{3.10}$$

where

$$F^{s,t} = \mathrm{proj}_{F(t)} F(s), G^{s,t} = \mathrm{proj}_{G(t)} G(s).$$

Theorem 3.20. *Let K, f, g be as above and $p \geq 0$. Then,*

$$\mathrm{dist}_{\mathcal{H},p,1}^{\mathrm{persistent}}(\mathcal{H}_p(K,f), \mathcal{H}_p(K,g)) \leq \pi \cdot \left(||f-g||_1^{(p)} + ||f-g||_1^{(p+1)} \right).$$

Before proving Theorem 3.20 we need a lemma.

Lemma 3.21. *Let V be a finite dimensional inner product space and P, V_1, V_2, subspaces of V. Let for $i = 1,2$, $Z_i = \mathrm{proj}_{V_i}(P)$ and $\Delta_i = \dim(V_i) - \dim(V_1 \cap V_2)$. Then for $i = 1,2$,*

$$\dim Z_i - \dim(Z_1 \cap Z_2) \leq \Delta_1 + \Delta_2.$$

Proof. Let $V' = V_1 + V_2$, and let

$$\pi_1 = \mathrm{proj}_{V_1}|_P, \pi_2 = \mathrm{proj}_{V_2}|_P, \pi = \mathrm{proj}_{V_1 \cap V_2}|_P : P \to V',$$

denote the orthogonal projections from P on to V_1, V_2, and $V_1 \cap V_2$ respectively. Denote for $i = 1,2$, $r_i = \mathrm{rank}(\pi_i)$, and $r = \mathrm{rank}(\pi)$.

Using the notation introduced above we have:

$$\begin{aligned}
\dim(Z_1 \cap Z_2) &= \dim(Z_1) + \dim(Z_2) - \dim(Z_1 + Z_2) \\
&= r_1 + r_2 - \dim(Z_1 + Z_2). \qquad (3.11)
\end{aligned}$$

Now let for $i = 1, 2$, W_i be the orthogonal complement of $V_1 \cap V_2$ in V_i. Thus, for $i = 1, 2$ we have an orthogonal decomposition $V_i = (V_1 \cap V_2) \oplus W_i$, so that $\Delta_i = \dim(W_i)$, and a decomposition $\pi_i = \pi \oplus \pi'_i$, where $\pi'_i = \mathrm{proj}_{W_i}|_P$.

Thus,

$$\begin{aligned} Z_1 + Z_2 &= \mathrm{Im}(\pi_1) + \mathrm{Im}(\pi_2) \\ &= \mathrm{Im}(\pi) + \mathrm{Im}(\pi'_1) + \mathrm{Im}(\pi'_2). \end{aligned}$$

Hence,

$$\begin{aligned} \dim(Z_1 + Z_2) &\leq \mathrm{rank}(\pi) + \mathrm{rank}(\pi'_1) + \mathrm{rank}(\pi'_2) \\ &\leq r + \Delta_1 + \Delta_2. \end{aligned} \tag{3.12}$$

Using inequalities (3.11), (3.12), and the fact that for $i, 1, 2$, $r \leq r_i$ (since π factors through each of the π_i's), we obtain

$$\begin{aligned} \dim(Z_1 \cap Z_2) &\geq r_1 + r_2 - r - \Delta_1 - \Delta_2 \\ &\geq r_i - \Delta_1 - \Delta_2 \\ &= \dim Z_i - \Delta_1 - \Delta_2, \end{aligned}$$

from which the lemma follows. ■

Proof of Theorem 3.20. First observe that the subspace $\mathcal{H}_p^{s,t}(K, f)$ (resp. $\mathcal{H}_p^{s,t}(K, g)$) is equal to the orthogonal projection of $\mathcal{H}_p^s(K, f)$ into $\mathcal{H}_p^t(K, f)$ (respectively, $\mathcal{H}_p^t(K, g)$).

Denote

$$F_p = \mathcal{H}_p(K, f), G_p = \mathcal{H}_p(K, g). \tag{3.13}$$

Let

$$\begin{aligned} U_f &= \mathcal{H}_p^s(K, f), & U_g &= \mathcal{H}_p^s(K, g), \\ V_f &= \mathcal{H}_p^t(K, f), & V_g &= \mathcal{H}_p^t(K, g), \\ W_f &= \mathrm{proj}_{V_f}(U_f), & W_g &= \mathrm{proj}_{V_g}(U_g), \\ Z_f &= \mathrm{proj}_{V_f}(U_f \cap U_g), & Z_g &= \mathrm{proj}_{V_g}(U_f \cap U_g). \end{aligned}$$

Using triangle inequality we get

$$\begin{aligned} d_{K,p}(F_p^{s,t}, G_p^{s,t}) &= d_{K,p}(W_f, W_g) \\ &\leq d_{K,p}(W_f, Z_f) + d_{K,p}(Z_f, Z_g) + d_{K,p}(W_g, Z_g) \end{aligned}$$

(see Eqn. (3.13) and Definition 3.19 for the definition of $F_p^{s,t}, G_p^{s,t}$). It is easy to verify from the definition of the metric $d_{K,p}$ that orthogonal projection does not increase the distance between subspaces if one is a subspace of the other. Hence,

$$\begin{aligned} d_{K,p}(W_f, Z_f) &\leq d_{K,p}(U_f, U_f \cap U_g) \\ &= \frac{\pi}{2} \cdot (\dim U_f - \dim(U_f \cap U_g))^{1/2}, \\ d_{K,p}(W_g, Z_g) &\leq d_{K,p}(U_g, U_f \cap U_g) \\ &= \frac{\pi}{2} \cdot (\dim U_g - \dim(U_f \cap U_g))^{1/2}. \end{aligned}$$

Moreover, using Lemma 3.21, (with $P = U_f \cap U_g$) we have that

$$\begin{aligned} d_{K,p}(Z_f, Z_g) &\leq \frac{\pi}{2} \left(\max(\dim Z_f, \dim Z_g) - \dim(Z_f \cap Z_g)\right)^{1/2} \\ &\leq \frac{\pi}{2} \left((\dim V_f - \dim(V_f \cap V_g)) + (\dim V_g - \dim(V_f \cap V_g))\right)^{1/2}. \end{aligned}$$

Now it follows from Lemma 2.16 that,

$$\begin{aligned}
\dim U_f - \dim(U_f \cap U_g) &= \dim \mathcal{H}_p^s(K,f) - \dim(\mathcal{H}_p^s(K,f) \cap \mathcal{H}_p^s(K,g)) \\
&\le \operatorname{card}\left(K^{[p]}_{f\le s} - K^{[p]}_{g \le s}\right) + \operatorname{card}\left(K^{[p+1]}_{g\le s} - K^{[p+1]}_{f\le s}\right), \\
\dim U_g - \dim(U_f \cap U_g) &= \dim \mathcal{H}_p^s(K,g) - \dim(\mathcal{H}_p^s(K,f) \cap \mathcal{H}_p^s(K,g)) \\
&\le \operatorname{card}\left(K^{[p]}_{g\le s} - K^{[p]}_{f \le s}\right) + \operatorname{card}\left(K^{[p+1]}_{f\le s} - K^{[p+1]}_{g\le s}\right), \\
\dim V_f - \dim(V_f \cap V_g) &= \dim \mathcal{H}_p^t(K,f) - \dim(\mathcal{H}_p^t(K,f) \cap \mathcal{H}_p^t(K,g)) \\
&\le \operatorname{card}\left(K^{[p]}_{f\le t} - K^{[p]}_{g \le t}\right) + \operatorname{card}\left(K^{[p+1]}_{g\le t} - K^{[p+1]}_{f\le t}\right), \\
\dim V_g - \dim(V_f \cap V_g) &= \dim \mathcal{H}_p^t(K,g) - \dim(\mathcal{H}_p^t(K,f) \cap \mathcal{H}_p^t(K,g)) \\
&\le \operatorname{card}\left(K^{[p]}_{g\le t} - K^{[p]}_{f \le t}\right) + \operatorname{card}\left(K^{[p+1]}_{f\le t} - K^{[p+1]}_{g\le t}\right).
\end{aligned}$$

It follows from the above inequalities that

$$\begin{aligned}
d_{K,p}(F_p^{s,t}, G_p^{s,t}) &\le \frac{\pi}{2} \cdot \left(\Delta^{(p)}_{f,g}(s)^{1/2} + \Delta^{(p)}_{g,f}(s)^{1/2} + (\Delta^{(p)}_{f,g}(t) + \Delta^{(p)}_{g,f}(t))^{1/2}\right) \\
&\le \frac{\pi}{2} \cdot \left(\Delta^{(p)}_{f,g}(s) + \Delta^{(p)}_{g,f}(s) + \Delta^{(p)}_{f,g}(t) + \Delta^{(p)}_{g,f}(t)\right), \qquad (3.14)
\end{aligned}$$

where for any two functions $h_1, h_2 : K \to \mathbb{R}$ and $r \in \mathbb{R}$,

$$\Delta^{(p)}_{h_1,h_2}(r) = \operatorname{card}\left(K^{[p]}_{h_1\le r} - K^{[p]}_{h_2 \le r}\right) + \operatorname{card}\left(K^{[p+1]}_{h_2\le r} - K^{[p+1]}_{h_1\le r}\right).$$

Using Definition 3.19 we have

$$\operatorname{dist}^{\text{persistent}}_{\mathcal{H},p,1}(\mathcal{H}_p(K,f), \mathcal{H}_p(K,g)) = \int_0^1 \int_0^t d_{K,p}(\mathcal{H}_p^{s,t}(K,f), \mathcal{H}_p^{s,t}(K,g)) ds dt. \qquad (3.15)$$

Using equality (3.15) and inequality (3.14) we get

$$\begin{aligned}
\operatorname{dist}^{\text{persistent}}_{\mathcal{H},p,1}(\mathcal{H}_p(K,f), \mathcal{H}_p(K,g)) &\le \frac{\pi}{2} \cdot \int_0^1 \int_0^t \left(\Delta^{(p)}_{f,g}(s) + \Delta^{(p)}_{g,f}(s) + \Delta^{(p)}_{f,g}(t) + \Delta^{(p)}_{g,f}(t)\right) ds dt \\
&\le \frac{\pi}{2} \cdot \left(\int_0^1 \left(\Delta^{(p)}_{f,g}(t) + \Delta^{(p)}_{g,f}(t)\right) dt + \int_0^1 \left(\Delta^{(p)}_{f,g}(s) + \Delta^{(p)}_{g,f}(s)\right) ds\right) \\
&= \pi \cdot \int_0^1 \left(\Delta^{(p)}_{f,g}(s) + \Delta^{(p)}_{g,f}(s)\right) ds \\
&= \pi \cdot \left(||f-g||_1^{(p)} + ||f-g||_1^{(p+1)}\right).
\end{aligned}$$

This completes the proof. ∎

3.4. Stability of harmonic persistent barcodes.

We have proved the stability of the harmonic homology, as well as the persistent homology subspaces of filtrations of finite simplicial complexes induced by admissible functions (Theorems 3.18, and 3.15). We now study the stability of the harmonic barcodes.

Let K be a finite simplicial complex and let $\mathcal{F}$ denote a finite filtration $K_0 \subset \cdots \subset K_N = K$. By convention we will assume that $K_s = \emptyset$ for $s < 0$ and $K_t = K$ for $t \ge N$.

The harmonic barcode subspaces $\mathcal{P}_p^{s,t}(\mathcal{F})$ (see Definition 3.11) are subspaces of $\mathcal{H}_p(K_s)$ (corresponding to the birth of the homology classes corresponding to the bar $b = (s, t, \mu_p^{s,t}(\mathcal{F}))$ (assuming $\mu_p^{s,t}(\mathcal{F}) \neq 0$).

We show below that in the case the bar b is simple (i.e. $\mu_p^{s,t}(\mathcal{F}) = 1$), it also makes sense to associate a subspace of $\mathcal{H}_p(K_{t-1})$ representing the bar b just before its death. With the notation introduced above:

Proposition 3.22. *Let $0 \le s < t \le N$, and suppose that $\mu_p^{s,t}(\mathcal{F}) = 1$. Then, the map $i_p^{s,t-1}$ induces an isomorphism $P_p^{s,t}(\mathcal{F}) \to \mathrm{H}_p^{s,t-1}(\mathcal{F})/\mathrm{H}_p^{s-1,t-1}(\mathcal{F})$.*

Proof. Recall that

$$\begin{aligned} M_p^{s,t}(\mathcal{F}) &= (i_p^{s,t})^{-1}(\mathrm{H}_p^{s-1,t}(\mathcal{F}) \\ &= (i_p^{s,t-1})^{-1}(\mathrm{H}_p^{s,t-1}(\mathcal{F})), \\ N_p^{s,t}(\mathcal{F})) &= (i_p^{s,t-1})^{-1}(\mathrm{H}_p^{s-1,t-1}(\mathcal{F})), \text{ and} \\ P_p^{s,t}(\mathcal{F})) &= M_p^{s,t}(\mathcal{F})/N_p^{s,t}(\mathcal{F}). \end{aligned}$$

It is clear that $i_p^{s,t-1}$ induces a surjective map $P_p^{s,t}(\mathcal{F}) \to \mathrm{H}_p^{s,t-1}(\mathcal{F})/\mathrm{H}_p^{s-1,t-1}(\mathcal{F})$. It thus suffices to prove that $\dim \mathrm{H}_p^{s,t-1}(\mathcal{F}) - \dim \mathrm{H}_p^{s-1,t-1}(\mathcal{F}) = 1$.

Since,

$$\begin{aligned} P_p^{s,t}(\mathcal{F})) &= M_p^{s,t}(\mathcal{F})/N_p^{s,t}(\mathcal{F}) \\ &\cong (i_p^{s,t-1})^{-1}(\mathrm{H}_p^{s,t-1}(\mathcal{F}))/(i_p^{s,t-1})^{-1}(\mathrm{H}_p^{s-1,t-1}(\mathcal{F})), \end{aligned}$$

it follows that

$$\begin{aligned} \dim P_p^{s,t}(\mathcal{F})) &= \dim(i_p^{s,t-1})^{-1}(\mathrm{H}_p^{s,t-1}(\mathcal{F})) - \dim(i_p^{s,t-1})^{-1}(\mathrm{H}_p^{s-1,t-1}(\mathcal{F})) \\ &\geq \dim \mathrm{H}_p^{s,t-1}(\mathcal{F})/\mathrm{H}_p^{s-1,t-1}(\mathcal{F}) \\ &= \dim \mathrm{H}_p^{s,t-1}(\mathcal{F}) - \dim \mathrm{H}_p^{s-1,t-1}(\mathcal{F}) \end{aligned}$$

Since $\dim P_p^{s,t}(\mathcal{F})) = 1$, it is immediate that

$$\dim \mathrm{H}_p^{s,t-1}(\mathcal{F}) - \dim \mathrm{H}_p^{s-1,t-1}(\mathcal{F}) \leq 1.$$

If

$$\dim \mathrm{H}_p^{s,t-1}(\mathcal{F}) - \dim \mathrm{H}_p^{s-1,t-1}(\mathcal{F}) = 0,$$

it would imply that $\dim P_p^{s,t}(\mathcal{F})) = 0$.

So,

$$\dim \mathrm{H}_p^{s,t-1}(\mathcal{F}) - \dim \mathrm{H}_p^{s-1,t-1}(\mathcal{F}) = 1.$$

■

Proposition 3.22 motivates the following definition.

Definition 3.23 (Terminal harmonic homology subspace associated to a simple bar). *Let $0 \leq s \leq N$, and let $b = (s, t, 1) \in \mathbf{B}_p(\mathcal{F})$ be a simple bar.*

We denote

$$\widehat{\mathcal{P}}_p^{s,t}(\mathcal{F}) = \mathcal{H}_p^{s,t-1}(\mathcal{F}) \cap \mathcal{H}_p^{s-1,t-1}(\mathcal{F})^{\perp}.$$

We call $\widehat{\mathcal{P}}_p^{s,t}(\mathcal{F})$ the terminal harmonic homology subspace associated to *b. We will refer to $\mathcal{P}_p^{s,t}(\mathcal{F})$ itself as the* initial *harmonic homology subspace associated to b.*

For technical reasons we will prove stability of the terminal rather than the (initial) harmonic subspaces associated to simple bars.

We first define an appropriate notion of distance between the harmonic barcodes of two different filtrations using the terminal harmonic subspaces. The distance measured introduced for proving stability of the harmonic homology subspaces and also the harmonic persistent homology subspaces (Theorems 3.18 and 3.20) are in the form of an integral (see Definitions 3.16 and 3.19). Since for a filtration $\mathcal{F}$ of a finite simplicial complex K, the subspaces $\mathcal{P}_p^{s,t}(\mathcal{F})$ will be non-zero only for a finitely many pairs (s,t), the integral form of the distance function is not suitable.

For two finite filtrations $\mathcal{F}, \mathcal{G}$, indexed by $[N]$, we will use the averaged sum (see Theorem 3.24 below) $\frac{1}{\binom{N+1}{2}} \cdot \sum_{0 \leq i < j \leq N} d_{C_p(K)}(\widehat{\mathcal{P}}_p^{i,j}(\mathcal{F}), \widehat{\mathcal{P}}_p^{i,j}(\mathcal{G}))$ as a measure of distance between the harmonic barcodes of $\mathcal{F}, \mathcal{G}$ in dimension p.

We prove the following theorem.

Theorem 3.24 (Stability of harmonic barcodes). *Let $\mathcal{F}, \mathcal{G}$ denote filtrations $K_0 \subset \cdots \subset K_N = K$, $K'_0 \subset \cdots \subset K'_N = K$ of a finite simplicial complex K.*

Let $f, g : K \to [0,1]$ be the admissible maps defined by

$$f(\sigma) = \frac{1}{N} \cdot \min\{s \in [N] \mid \sigma \in K_s\},$$
$$g(\sigma) = \frac{1}{N} \cdot \min\{s \in [N] \mid \sigma \in K'_s\},$$

each $\sigma \in K$.

Moreover suppose that for all $i, j \in [N]^2, 0 \le i < j \le N$

$$\dim \mathcal{P}_p^{i,j}(\mathcal{F})), \dim \mathcal{P}_p^{i,j}(\mathcal{G})) \le 1.$$

Then,

$$\frac{1}{\binom{N+1}{2}} \cdot \sum_{0 \le i < j \le N} d_{C_p(K)}(\widehat{\mathcal{P}}_p^{i,j}(\mathcal{F}), \widehat{\mathcal{P}}_p^{i,j}(\mathcal{G})) \le 2 \cdot \mathrm{dist}^{\text{persistent}}_{\mathcal{H},p,1}(\mathcal{H}_p(K,f), \mathcal{H}_p(K,g))$$
$$\le \frac{\pi^3}{2}\left(||f-g||_1^{(p)} + ||f-g||_1^{(p+1)}\right).$$

Before proving Theorem 3.24 we first prove a proposition that is key to the proof of the theorem. This proposition might be of independent interest since it shows that the distance (angle) between terminal harmonic subspaces associated to simple bars is bounded from above by the distances between certain harmonic persistent homology subspaces of the two filtrations.

Proposition 3.25. *Let $\mathcal{F}, \mathcal{G}$ denote filtrations $K_0 \subset \cdots \subset K_N = K$, $K'_0 \subset \cdots \subset K'_N = K$ of a finite simplicial complex K, and let $p \ge 0$. Then, for each $i, j, 0 \le i < j \le N$ such that $\dim \mathcal{P}_p^{i,j}(\mathcal{F})) = \dim \mathcal{P}_p^{i,j}(\mathcal{G})) = 1$,*

$$d_{C_p(K)}(\widehat{\mathcal{P}}_p^{i,j}(\mathcal{F})), \widehat{\mathcal{P}}_p^{i,j}(\mathcal{G}))) \le \frac{\pi^2}{4} \cdot (d_{C_p(K)}(\mathcal{H}_p^{i,j-1}(\mathcal{F}), \mathcal{H}_p^{i,j-1}(\mathcal{G}))$$
$$+ d_{C_p(K)}(\mathcal{H}_p^{i-1,j-1}(\mathcal{F}), \mathcal{H}_p^{i-1,j-1}(\mathcal{G}))).$$

Before proving Proposition 3.25 we need a lemma.

Lemma 3.26. *Let $W, W' \subset V$ be two subspaces of an Euclidean space V, with $\dim W = \dim W'$. Suppose $U \subset W, U' \subset W'$ be codimension one subspaces. Let $P = W \cap U^\perp, P' = W' \cap U'^\perp$, α the principle angle between P, P', θ_0 denote the largest principle angle between W, W', and $\theta_1, \dots, \theta_N$ denote the principle angles between U, U'. Then, $\alpha^2 \le \frac{\pi^2}{4} \cdot \left(\sum_{i=0}^N \theta_i^2\right)$.*

Proof. Let $e_1, \dots, e_N$ (resp. $e'_1, \dots, e'_N$) be an orthonormal basis of U (resp. U') such the angle between e_i, e'_i equals θ_i. Let e_0, e'_0 be a unit vectors spanning P_1, P'_1 respectively, such that $\langle e_0, e'_0\rangle = \cos\alpha$.

Now, $e'_0 = \cos\alpha e_0 + \sum_{i=1}^N a_i e_i + e$, where $a_i = \langle e'_0, e_i\rangle, 1 \le i \le N$ and $e = \mathrm{proj}_{W^\perp} e'_0$.

Let f be a unit length vector in W such that the angle θ between the $P' = \mathrm{span}(e'_0)$ and $\mathrm{span}(f)$ is the smallest possible. Then, using Lemma 2.13, $\theta \le \theta_0$, since θ_0 is the largest principle angle between W', W.

Since e is orthogonal to f, it follows that the angle between $\mathrm{span}(e)$ and $P' = \mathrm{span}(e'_0)$ is bounded between $\pi/2 - \theta$ and $\pi/2 + \theta$, and using the inequality from the previous paragraph, we obtain that this angle is between $\pi/2 - \theta_0$ and $\pi/2 + \theta_0$. It follows that $||e||^2 \le \sin^2\theta_0$. Now, $\langle e'_0, e'_i\rangle = 0, \langle e'_i, e_i\rangle = \cos\theta_i$, which implies that the angle between e'_0, e_i is between $\pi/2 - \theta_i, \pi/2 + \theta_i$. Hence, $|a_i| \le \sin\theta_i, 1 \le i \le N$ implying $1 = \cos^2\alpha + \sum_{i=0}^N a_i^2 \le \cos^2\alpha + \sum_{i=0}^N \sin^2\theta_i$. It follows that $\sin^2\alpha \le \sum_{i=0}^N \sin^2\theta_i$. Since, $\frac{2}{\pi}x \le \sin x \le x$, for $0 \le x \le \frac{\pi}{2}$, we get that $\frac{4}{\pi^2}\alpha^2 \le \sum_{i=0}^N \theta_i^2$ from which the stated inequality follows immediately. ∎

Proof of Proposition 3.25. Since $d_{C_p(K)}(\widehat{\mathcal{P}}_p^{i,j}(\mathcal{F})), \widehat{\mathcal{P}}_p^{i,j}(\mathcal{G})))$ is at most $\frac{\pi^2}{4}$ we can assume that $\dim \mathcal{H}_p^{i,j-1}(\mathcal{F}) = \dim \mathcal{H}_p^{i,j-1}(\mathcal{G})$ and $\dim \mathcal{H}_p^{i-1,j-1}(\mathcal{F}) = \dim \mathcal{H}_p^{i-1,j-1}(\mathcal{G})$. Otherwise, the claimed inequality is true.

Now apply Lemma 3.26 with
$$W = \mathcal{H}_p^{i,j-1}(\mathcal{F}), W' = \mathcal{H}_p^{i,j-1}(\mathcal{G}), U = \mathcal{H}_p^{i-1,j-1}(\mathcal{F}), U' = \mathcal{H}_p^{i-1,j-1}(\mathcal{G}).$$ ■

Proof of Theorem 3.24. Follows from Proposition 3.25, Definition 3.19, and Theorem 3.20. ■

4. Essential simplices and harmonic representative. We now make precise the notion of essential simplices referred to previously. Let K be a finite simplicial complex and $\mathcal{F} = (K_t)_{t\in T}$ denote a filtration of K.

Let $\phi_p^s = \phi_p(K_s) : Z_p(K_s) \to \mathrm{H}_p(K_s)$ be the canonical surjection. For every $s \in T$ we denote,
$$\widetilde{M}_p^{s,\infty}(\mathcal{F}) = \bigcup_{s\leq t} (\phi_p^s)^{-1}(M_p^{s,t}(\mathcal{F})),$$
and for $s \leq t$, we denote
$$\widetilde{M}_p^{s,t}(\mathcal{F}) = (\phi_p^s)^{-1}(M_p^{s,t}(\mathcal{F})),$$
$$\widetilde{N}_p^{s,t}(\mathcal{F}) = (\phi_p^s)^{-1}(N_p^{s,t}(\mathcal{F})).$$

Proposition 4.1. *Let K be a finite simplicial complex and $\mathcal{F} = (K_t)_{t\in T}$ a filtration. Let $s \in T, t \in T \cup \{\infty\}$, $s \leq t$.*

1. *If $t = \infty$ the following diagram is commutative and all maps are isomorphisms.*

$$\begin{tikzcd} Z_p(K_s)/\widetilde{M}_p^{s,\infty}(\mathcal{F}) \arrow[rr, "\phi_p^s"] \arrow[dr, "f_p"'] & & P_p^{s,\infty}(\mathcal{F}) , \arrow[dl, "g_p"] \\ & \mathcal{P}_p^{s,\infty}(\mathcal{F}) & \end{tikzcd}$$

where
$$\phi_p^s(z + \widetilde{M}_p^{s,\infty}(\mathcal{F})) = [z] + M_p^{s,\infty}(\mathcal{F}),$$
$$f_p(z + \widetilde{M}_p^{s,\infty}(\mathcal{F})) = \mathrm{proj}_{\widetilde{M}_p^{s,t}(\mathcal{F})^\perp}(z),$$
$$g_p([z] + M_p^{s,\infty}(\mathcal{F})) = \mathrm{proj}_{\widetilde{M}_p^{s,\infty}(\mathcal{F})^\perp}(z).$$
(here we denote for a cycle $z \in Z_p(K_s)$, by $[z] = z + B_p(K_s)$ the homology class of z).

2. *If $t \neq \infty$ the following diagram is commutative and all maps are isomorphisms.*

$$\begin{tikzcd} \widetilde{M}_p^{s,t}(\mathcal{F})/\widetilde{N}_p^{s,t}(\mathcal{F}) \arrow[rr, "\phi_p"] \arrow[dr, "f_p"'] & & P_p^{s,t}(\mathcal{F}) , \arrow[dl, "g_p"] \\ & \mathcal{P}_p^{s,t}(\mathcal{F}) & \end{tikzcd}$$

where
$$\phi_p(z + \widetilde{N}_p^{s,t}(\mathcal{F})) = [z] + N_p^{s,t}(\mathcal{F}),$$
$$f_p(z + \widetilde{N}_p^{s,t}(\mathcal{F})) = \mathrm{proj}_{\widetilde{N}_p^{s,t}(\mathcal{F})^\perp}(z),$$
$$g_p([z] + N_p^{s,t}(\mathcal{F})) = \mathrm{proj}_{\widetilde{N}_p^{s,t}(\mathcal{F})^\perp}(z).$$

Proof. The fact that the map ϕ_p is an isomorphism follows from a standard isomorphism theorem for vector spaces. The fact that the remaining maps are well defined, and are isomorphisms, and also the commutativity of the diagrams are easily checked using Lemma 2.10, Proposition 2.9, and the definitions of the various subspaces involved. ■

Definition 4.2 (Support of a chain). *For $z = \sum_{\sigma\in K^{[p]}} c_\sigma \cdot \sigma \in C_p(K)$, we denote* $\mathrm{supp}(z) = \{\sigma \in K^{[p]} \mid c_\sigma \neq 0\}$*, and call* $\mathrm{supp}(z)$ *the* the support of z.

The following definition subsumes the one in [3].

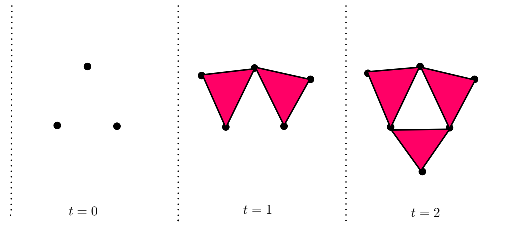

Figure 5: Empty set of essential edges

Definition 4.3. *Let $b = (s, t, 1) \in \mathbf{B}_p(\mathcal{F})$ be a simple bar of $\mathcal{F}$ (see Definition 3.7). We define*

$$\begin{aligned}\mathrm{Rep}(b) &= Z_p(K_s) \setminus \widetilde{M}_p^{s,\infty}(\mathcal{F}) \text{ if } t = \infty \\ &= \widetilde{M}_p^{s,t}(\mathcal{F}) - \widetilde{N}_p^{s,t}(\mathcal{F}) \text{ else.}\end{aligned}$$

We call $\mathrm{Rep}(b)$ the set of cycles representing the bar b. *More precisely, for $z \in Z_p(K_s)$, $z \in \mathrm{Rep}(b)$, if and only if z represents a non-zero element in $P_p^{s,t}(\mathcal{F})$.*

Definition 4.4 (The set of essential simplices associated to a simple bar). *Let $b = (s, t, 1) \in \mathbf{B}_p(\mathcal{F})$ be a simple bar of $\mathcal{F}$. We define $\Sigma(b) = \bigcap_{z \in \mathrm{Rep}(b)} \mathrm{supp}(z)$. We will call* $\Sigma(b)$ the set of essential simplices of b.

Definition 4.5 (Relative essential content). *Let $b = (s, t, 1) \in \mathbf{B}_p(\mathcal{F})$ be a simple bar of $\mathcal{F}$. For $z = \sum_{\sigma \in K^{[p]})} c_\sigma \cdot \sigma \in \mathrm{Rep}(b)$, we denote* $\mathrm{content}(z) = \left(\frac{\sum_{\sigma \in \Sigma(b)} c_\sigma^2}{\sum_{\sigma \in K^{[p]}} c_\sigma^2}\right)^{1/2}$. *We will call* $\mathrm{content}(z)$ the relative essential content of z.

Definition 4.6 (Harmonic representative). *Given a simple bar $b = (s, t, 1)$ we will call any non-zero cycle in $\mathcal{P}_p^{s,t}(\mathcal{F})$* a harmonic representative of b.

Remark 4.7. Note that the set of essential simplices of a bar can be empty. For example, the unique bar, $(2, \infty, 1) \in \mathbf{B}_1(\mathcal{F})$ of the filtration $\mathcal{F}$ shown in Figure 5 has no essential simplices (edges). None of the edges are indispensable for obtaining a representative of the unique non-zero homology class in dimension one that is born at $t = 2$. However, this situation cannot occur if the filtration is simplex-wise (see [3]). In that case the last simplex added at the time a bar is created is always in the set of essential simplices of the bar. The filtration in Figure 5 is not a simplex-wise filtration.

4.1. Relative essential content.

Theorem 4.8 (Harmonic representatives maximize relative essential content). *Let K be a finite simplicial complex and $\mathcal{F} = (K_t)_{t \in T}$ denote a filtration of K. Suppose $p \geq 0$, and let $b = (s, t, 1) \in \mathbf{B}_p(\mathcal{F})$ be a simple bar. Let z_0 be a harmonic representative of b. Then for any $z \in \mathrm{Rep}(b)$,*

$$\mathrm{content}(z) \leq \mathrm{content}(z_0).$$

Remark 4.9. Note that Theorem 4.8 implies in particular that the relative essential contents of any two harmonic representatives of a simple bar are equal. But this is clear also from the definition of the relative essential content and the fact that any two harmonic representatives of a simple bar are proportional.

Before proving Theorem 4.8 we first prove a technical lemma.

Lemma 4.10. *Let $V, W \subset Z_p(K)$ be subspaces with $W \subset V$ with $\dim V - \dim W = 1$. Let $\Gamma = \bigcap_{z \in V \setminus W} \mathrm{supp}(z)$. Then, for each $z \in V \setminus W$*

1. *$\Gamma \subset \mathrm{supp}(z)$;*

2. *for each* $\sigma \in \operatorname{supp}(z)$, $\sigma \in \Gamma$ *if and only if* $\sigma \in W^{\perp}$.

Proof. Part (1) is clear.

We now prove Part (2). Let $z \in V \setminus W$, and $\sigma_0 \in \operatorname{supp}(z)$.

1. $\sigma_0 \in \Gamma \Rightarrow \sigma_0 \in W^{\perp}$: Suppose that $\sigma_0 \in \Gamma$ and that $\sigma_0 \notin W^{\perp}$. Then, there exists $z' \in W$, with $\langle \sigma_0, z' \rangle \neq 0$. Otherwise, σ_0 is orthogonal to every cycle in W, and and hence $\sigma_0 \in W^{\perp}$. Let $z'' = z - \frac{\langle \sigma_0, z \rangle}{\langle \sigma_0, z' \rangle} z'$. Then, $z'' \in V \setminus W$ (otherwise $z \in W$) but $\sigma_0 \notin \operatorname{supp}(z'')$. This contradicts the fact that $\sigma_0 \in \Gamma$. Hence, $\sigma_0 \in W^{\perp}$.

2. $\sigma_0 \in W^{\perp} \Rightarrow \sigma_0 \in \Gamma$: Let $\sigma_0 \in W^{\perp}$, and let $z'' \in V \setminus W$. Suppose that $\sigma_0 \notin \operatorname{supp}(z'')$. Since $\dim V - \dim W = 1$, there exists $c \in \mathbb{R}, c \neq 0$, such that $z + cz'' \in W$.
Then, $\sigma_0 \in \operatorname{supp}(z + cz'')$, but $z + cz'' \in W$. This contradicts the fact that $\sigma_0 \in W^{\perp}$.
So, $\sigma_0 \in \operatorname{supp}(z'')$ for every $z'' \in V \setminus W$. Hence, $\sigma_0 \in \Gamma$. ■

Proof of Theorem 4.8. There are two cases.

If $t = \infty$, set $V = Z_p(K_s), W = \bigcup_{s \leq t} \widetilde{M}_p^{s,t}(\mathcal{F}), \Gamma = \Sigma(b)$.

If $t \neq \infty$, set $V = \widetilde{M}_p^{s,t}(\mathcal{F}), W = \widetilde{N}_p^{s,t}(\mathcal{F}), \Gamma = \Sigma(b)$.

Let $z = \sum_{\sigma \in K^{[p]}} c_\sigma \cdot \sigma \in \operatorname{Rep}(b),$, $z_1 = \sum_{\sigma \in \Sigma(b)} c_\sigma \cdot \sigma$, and $z_2 = z - z_1$. Using Lemma 4.10, $z_1 \in W^{\perp}$. Clearly $z_1 \perp z_2$. Also using Lemma 4.10 we have that for each $\sigma \in \Sigma(b) = \operatorname{supp}(z_1)$, $\sigma \in W^{\perp}$, from which it follows that $\sigma \notin \operatorname{supp}(\operatorname{proj}_W(z_2))$. Since it is clear from the definition of z_2, that for each $\sigma \in \Sigma(b) = \operatorname{supp}(z_1)$, $\sigma \notin \operatorname{supp}(z_2)$, we obtain that for each $\sigma \in \Sigma(b) = \operatorname{supp}(z_1)$,

$$\sigma \notin \operatorname{supp}(z_2 - \operatorname{proj}_W(z_2)) = \operatorname{supp}(\operatorname{proj}_{W^{\perp}}(z_2)). \tag{4.1}$$

In particular, this implies that

$$z_1 \perp \operatorname{proj}_{W^{\perp}}(z_2). \tag{4.2}$$

Let $z_0 = \operatorname{proj}_{W^{\perp}}(z)$. Then, using Proposition 4.1, z_0 is a harmonic representative of b.

Moreover,

$$\begin{aligned} z_0 &= \operatorname{proj}_{W^{\perp}}(z_1 + z_2) \\ &= \operatorname{proj}_{W^{\perp}}(z_1) + \operatorname{proj}_{W^{\perp}}(z_2) \\ &= z_1 + \operatorname{proj}_{W^{\perp}}(z_2), \end{aligned}$$

and (4.1) and (4.2) imply that

$$||z_0||^2 = ||z_1||^2 + ||\operatorname{proj}_{W^{\perp}}(z_2)||^2, \tag{4.3}$$

$$\operatorname{content}(z_0) = \frac{||z_1||}{||z_0||}. \tag{4.4}$$

Hence,

$$\begin{aligned} \operatorname{content}(z)^2 &= \frac{||z_1||^2}{||z_1||^2 + ||z_2||^2} \\ &\leq \frac{||z_1||^2}{||z_1||^2 + ||\operatorname{proj}_{W^{\perp}}(z_2)||^2} \text{ (projection does not increase length)} \\ &= \frac{||z_1||^2}{||z_1 + \operatorname{proj}_{W^{\perp}}(z_2)||^2} \text{ (using (4.2))} \\ &= \frac{||z_1||^2}{||z_0||^2} \text{ (using (4.3))} \\ &= \operatorname{content}(z_0)^2 \text{ (using (4.4))}. \end{aligned}$$

This completes the proof of the theorem. ■

REFERENCES

[1] Laurent Bartholdi, Thomas Schick, Nat Smale, and Steve Smale, *Hodge theory on metric spaces*, Found. Comput. Math. **12** (2012), no. 1, 1–48, Appendix B by Anthony W. Baker. MR 2886155

[2] Saugata Basu and Negin Karisani, *Efficient simplicial replacement of semi-algebraic sets*, 2020, URL:https://arxiv.org/pdf/2009.13365.pdf.

[3] Saugata Basu, Filippo Utro, and Laxmi Parida, *Essential simplices in persistent homology and subtle admixture detection*, 18th International Workshop on Algorithms in Bioinformatics (Laxmi Parida and Esko Ukkonen, eds.), Leibniz International Proceedings in Informatics, vol. 113, Schloss Dagstuhl, 2018, pp. 14:1–14:10.

[4] Peter Bubenik, *Statistical topological data analysis using persistence landscapes*, J. Mach. Learn. Res. **16** (2015), 77–102. MR 3317230

[5] Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot, *The structure and stability of persistence modules*, SpringerBriefs in Mathematics, Springer, [Cham], 2016. MR 3524869

[6] Jiahui Chen, Rundong Zhao, Yiying Tong, and Guo-Wei Wei, *Evolutionary de Rham-Hodge method*, Discrete & Continuous Dynamical Systems - B **26** (2021), no. 7, 3785–3821.

[7] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer, *Stability of persistence diagrams*, Discrete Comput. Geom. **37** (2007), no. 1, 103–120. MR 2279866

[8] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko, *Lipschitz functions have L_p-stable persistence*, Found. Comput. Math. **10** (2010), no. 2, 127–139. MR 2594441

[9] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Dmitriy Morozov, *Persistent homology for kernels, images, and cokernels*, Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, PA, 2009, pp. 1011–1020. MR 2807543

[10] Tamal K. Dey, Anil N. Hirani, and Bala Krishnamoorthy, *Optimal homologous cycles, total unimodularity, and linear programming*, Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010 (Leonard J. Schulman, ed.), ACM, 2010, pp. 221–230.

[11] Tamal K. Dey, Tao Hou, and Sayan Mandal, *Computing minimal persistent cycles: Polynomial and hard cases*, Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020 (Shuchi Chawla, ed.), SIAM, 2020, pp. 2587–2606.

[12] Beno Eckmann, *Harmonische funktionen und randwertaufgaben in einem komplex*, Commentarii Mathematici Helvetici **17** (1944), no. 1, 240–255.

[13] Herbert Edelsbrunner and John Harer, *Persistent homology—a survey*, Surveys on discrete and computational geometry, Contemp. Math., vol. 453, Amer. Math. Soc., Providence, RI, 2008, pp. 257–282. MR 2405684 (2009h:55003)

[14] Herbert Edelsbrunner and John L. Harer, *Computational topology*, American Mathematical Society, Providence, RI, 2010, An introduction. MR 2572029 (2011e:00001)

[15] Joel Friedman, *Computing betti numbers via combinatorial laplacians*, Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, Philadelphia, Pennsylvania, USA, May 22-24, 1996 (Gary L. Miller, ed.), ACM, 1996, pp. 386–391.

[16] Jennifer Gamble, Harish Chintakunta, and Hamid Krim, *Adaptive tracking of representative cycles in regular and zigzag persistent homology*, CoRR (2014).

[17] Robert Ghrist and Gregory Henselman-Petrusek, *Saecular persistence*, 2021.

[18] Aldo Guzmán-Sáenz, Niina Haiminen, Saugata Basu, and Laxmi Parida, *Signal enrichment with strain-level resolution in metagenomes using topological data analysis*, BMC Genom. **20** (2019), no. S2.

[19] Sara Kališnik, Vitaliy Kurlin, and Davorin Lešnik, *A higher-dimensional homologically persistent skeleton*, Adv. in Appl. Math. **102** (2019), 113–142. MR 3866989

[20] V. Kurlin, *A one-dimensional homologically persistent skeleton of an unstructured point cloud in any metric space*, Computer Graphics Forum **34** (2015), no. 5, 253–262.

[21] Michael Lesnick, Raúl Rabadán, and Daniel I. S. Rosenbloom, *Quantifying genetic innovation: Mathematical foundations for the topological study of reticulate evolution*, SIAM J. Appl. Algebra Geom. **4** (2020), no. 1, 141–184.

[22] Lu Li, Connor Thompson, Gregory Henselman-Petrusek, Chad Giusti, and Lori Ziegelmeier, *Minimal cycle representatives in persistent homology using linear programming: An empirical study with user's guide*, Front. of Artif. Intell. (2021).

[23] A. Lieutier, *Talk: Persistent harmonic forms*, URL: https://project.inria.fr/gudhi/files/2014/10/Persistent-Harmonic-Forms.pdf. Last visited on 2021/05/28.

[24] Lek-Heng Lim, *Hodge Laplacians on graphs*, SIAM Rev. **62** (2020), no. 3, 685–715. MR 4131346

[25] S. Lockwood and B. Krishnamoorthy, *Topological features in cancer gene expression data*, Biocomputing 2015 (2014).

[26] Sayan Mandal, Aldo Guzmán-Sáenz, Niina Haiminen, Saugata Basu, and Laxmi Parida, *A topological data analysis approach on predicting phenotypes from gene expression data*, Algorithms for Computational Biology - 7th International Conference, AlCoB 2020, Missoula, MT, USA, April 13-15, 2020, Proceedings

(Carlos Martín-Vide, Miguel A. Vega-Rodríguez, and Travis J. Wheeler, eds.), Lecture Notes in Computer Science, vol. 12099, Springer, 2020, pp. 178–187.

[27] Facundo Memoli, Zhengchao Wan, and Yusu Wang, *Persistent laplacians: Properties, algorithms and implications*, SIAM Journal on Mathematics of Data Science **4** (2022), no. 2, 858–884.

[28] Ippei Obayashi, *Volume-optimal cycle: Tight representative cycle of a generator in persistent homology*, SIAM Journal on Applied Algebra and Geometry **2** (2018), no. 4, 508–534.

[29] Laxmi Parida, *Talk: Combinatorics, statistics and topology enabling genomics*, Distinguished Keynote at ISMB, Montreal, Canada (virtual) July, 2020, URL: https://www.youtube.com/watch?v=ZtHYXou_yR4. Last visited on 2021/05/28.

[30] Raul Rabadan and Andrew J. Blumberg, *Topological data analysis for genomics and evolution: Topology in biology*, Cambridge University Press, 2019.

[31] R. C. Thompson, *Principal submatrices. IX. Interlacing inequalities for singular values of submatrices*, Linear Algebra Appl. **5** (1972), 1–12. MR 297791

[32] Ke Ye and Lek-Heng Lim, *Schubert varieties and distances between subspaces of different dimensions*, SIAM J. Matrix Anal. Appl. **37** (2016), no. 3, 1176–1197. MR 3543155