PERSISTENT MAGNITUDE

DEJAN GOVC AND RICHARD HEPWORTH

ABSTRACT. In this paper we introduce the *persistent magnitude*, a new numerical invariant of (sufficiently nice) graded persistence modules. It is a weighted and signed count of the bars of the persistence module, in which a bar of the form $[a, b)$ in degree d is counted with weight $(e^{-a} - e^{-b})$ and sign $(-1)^d$. Persistent magnitude has good formal properties, such as additivity with respect to exact sequences and compatibility with tensor products, and has interpretations in terms of both the associated graded functor, and the Laplace transform.

Our definition is inspired by Otter's notion of blurred magnitude homology: we show that the magnitude of a finite metric space is precisely the persistent magnitude of its blurred magnitude homology. Turning this result on its head, we obtain a strategy for turning existing persistent homology theories into new numerical invariants by applying the persistent magnitude. We explore this strategy in detail in the case of persistent homology of Morse functions, and in the case of Rips homology.

1. INTRODUCTION

Magnitude is a numerical invariant of metric spaces arising from category theory and with nontrivial geometric content. In this paper we apply the theory of magnitude and its categorifications to the study of persistence modules and persistent homology theories.

1.1. **Background.** Persistent homology, a fundamental notion in topological data analysis (or TDA), is a tool for measuring the shape of data sets and other objects. The general idea is to take a data set and produce an increasing sequence of topological spaces X_s, one for each value of a parameter s, where X_s describes the shape of the data set 'at scale s'. Taking the homology of the X_s produces the homology groups $H_*(X_s)$ together with structure maps $H_*(X_s) \to H_*(X_{s'})$ whenever $s \leqslant s'$. This structure is called the *persistent homology* of the data set, and it is an example of an algebraic structure called a graded *persistence module*. Any (sufficiently nice) persistence module has a *barcode decomposition* describing its isomorphism class in terms of a collection of intervals called *bars*. Each bar is interpreted as a feature of the data set: the start point of the interval is the scale at which the feature first comes into being, and the end point is the scale at which the feature evaporates. In one common interpretation, longer bars are interpreted as significant features, while shorter bars are interpreted as noise.

Magnitude is a numerical invariant of metric spaces introduced by Leinster [34] (see also the survey [35]), as an instance of a general category theoretical construction. Despite its abstract origins, magnitude is a rich geometric invariant: Meckes [40] showed that magnitude can detect the Minkowski dimension of compact subsets of Euclidean space, Barceló-Carbery [7] showed that it can detect the volume of compact subsets of Euclidean space, and Gimperlein-Goffeng [24] showed that it can in addition detect surface area and the second intrinsic volume V_2 of appropriate subsets of odd-dimensional Euclidean space.

Magnitude of metric spaces has a categorification, called *magnitude homology*, which was introduced by Hepworth-Willerton [31] and Leinster-Shulman [36]. The magnitude homology of a metric space is a bigraded abelian group, whose graded Euler characteristic recovers the magnitude of the metric space, at least when the space is finite. Thus the relationship between magnitude and magnitude homology is analogous to the relationship between Euler characteristic and singular homology. More recently, Otter [42] has introduced a *blurred* or persistent version of magnitude homology, which relates magnitude homology to the Rips complex and, importantly, to ordinary homology.

1.2. **Results.** Blurred magnitude homology assigns to each metric space X a graded persistence module $\mathrm{BMH}_*(X)$. When X is finite, we show that there is an attractive relationship between the barcode decomposition of $\mathrm{BMH}_*(X)$ and the magnitude $|X|$ of X:

$$|X| = \sum_{k=0}^{\infty} \sum_{i=1}^{m_k} (-1)^k (e^{-a_{k,i}} - e^{-b_{k,i}})$$

where $\mathrm{BMH}_*(X)$ has bars $[a_{k,1}, b_{k,1}), \ldots, [a_{k,m_k}, b_{k,m_k})$ in degree $k \geq 0$.

Observe that the right hand side of the equation above makes sense for any graded persistence module, so long as it is subject to a finiteness condition such as being finitely presented. We turn this observation into a definition: The *persistent magnitude* or simply *magnitude* $|M_*|$ of a finitely presented graded persistence module M_* is defined by

$$|M_*| = \sum_{k=0}^{\infty} \sum_{i=1}^{m_k} (-1)^k (e^{-a_{k,i}} - e^{-b_{k,i}})$$

where M_* has bars $[a_{k,1}, b_{k,1}), \ldots, [a_{k,m_k}, b_{k,m_k})$ in degree $k \geq 0$. Note that a bar $[a, b)$ makes a contribution of $\pm(e^{-a} - e^{-b})$ to the magnitude, so that longer bars make a greater contribution, in line with one of the general philosophies of persistent homology.

Persistent magnitude has good formal properties: we show that it is additive with respect to exact sequences, and that the magnitude of a tensor product of persistence modules is the product of the magnitudes of the factors, so long as the tensor product is understood in an appropriate derived sense.

Now suppose that we have a persistent homology theory defined for some class of mathematical objects, for example the Rips homology of metric spaces. By applying persistent magnitude to the persistent homology, we obtain a new numerical invariant of the mathematical objects in question. Our first example of this process is the case of the sublevel set persistent homology of Morse

functions, where the resulting magnitude invariant is a (signed and weighted) count of the critical points of the original function.

Our most detailed example of persistent magnitude in action is the *Rips magnitude*. This is the numerical invariant of finite metric spaces obtained by taking the persistent magnitude of the Rips homology, and is given by the weighted simplex-count

$$|X|_{\mathrm{Rips}} = \sum_{\emptyset \neq A \subseteq X} (-1)^{\#A-1} e^{-\operatorname{diam}(A)}.$$

We compute the Rips magnitude of cycle graphs with their path, Euclidean and geodesic metrics. In each case they are determined by a number-theoretical formula reminiscent of the sum of divisors function.

In the original setting of magnitude, defining the magnitude of infinite metric spaces is not straightforward: the simplest method is to take the supremum of the magnitude of all finite subspaces of the given infinite metric space, but there are alternatives, and currently the theory only works well in the case of *positive definite* spaces. We conclude the paper by investigating the question of whether Rips magnitude can be extended to infinite metric spaces. In the case of closed intervals in $\mathbb{R}$ the approach via a supremum works well and we find that $|[a,b]|_{\mathrm{Rips}} = 1 + (b-a)$. In the case of the circle with its Euclidean and geodesic metrics, which we study in detail, the results are attractive but inconclusive.

1.3. **Organisation.** We begin the paper with a series of generous background sections: persistence modules and persistent homology in section 2, magnitude in section 3, and magnitude homology in section 4. Section 5 introduces the persistent magnitude of persistence modules, and studies its basic properties. Section 6 applies persistent magnitude to the persistent homology of sublevel sets. The final part of the paper studies Rips magnitude: section 7 introduces Rips magnitude and discusses its properties and some basic examples, section 8 computes it in the case of cycle graphs (with various metrics), and section 9 explores the possibility of defining Rips magnitude for infinite metric spaces.

1.4. **Open Questions.** The results obtained in this work raise several natural questions, that we have not yet been able to answer conclusively:

- What is the most general notion of tameness sufficient to develop the theory of persistent magnitude? (Our characterisation using the Laplace transform suggests that one might want to consider a notion of persistence modules of "exponential type", meaning that the rank function is of exponential type.)
- Is there a general definition of Rips magnitude for (a suitable class of) infinite metric spaces? Can we establish asymptotic results similar to the case of the circle for higher-dimensional spheres or other manifolds? (Note that not much seems to be known about Rips filtrations of manifolds beyond the circle [2].)
- The formulas for the Rips magnitudes of cycle graphs and Euclidean cycles seem reminiscent of the sum of divisors functions from number

theory. Are there interesting connections between Rips magnitude of cycles and analytic number theory?

2. BACKGROUND ON PERSISTENCE MODULES AND PERSISTENT HOMOLOGY

2.1. **Persistence modules.** Here we review some standard material on persistence modules, mostly following [17, 14]. For a survey explaining the basic ideas and historical origin of persistence, see [21]. A modern exposition of the main ideas including the structure and stability theorems for persistence modules can be found in [17]. For further background on persistence modules from the category theoretical perspective, see [14]. A slightly more algebraic perspective, with a view towards multi-parameter persistence, can be found in [38]. An account of some aspects of homological algebra for persistence modules can be found in [13].

Throughout the paper, we will work with vector spaces over a fixed field $\Bbbk$. The category of vector spaces over $\Bbbk$ will be denoted by **Vect**. In the most general setting, persistence modules can be considered over an arbitrary small category, see e.g. [11, 14]; however, we will restrict attention to the case of $(\mathbb{R}, \leq)$-indexed persistence modules, as this is entirely sufficient for our purposes. Here $(\mathbb{R}, \leq)$ denotes either the poset $\mathbb{R}$ equipped with the partial order $\leq$, or the associated category with objects $\mathbb{R}$ and a unique morphism $x \to y$ whenever $x \leq y$.

Definition 2.1. A *persistence module* is a functor $M \colon (\mathbb{R}, \leq) \to \textbf{Vect}$. A *morphism* of persistence modules is a natural transformation of such functors.

Remark 2.2. The category $\textbf{PersMod} = \textbf{Vect}^{(\mathbb{R}, \leq)}$ of persistence modules has the structure of an abelian category. In particular, morphisms of persistence modules have well-defined kernels, and cokernels. These are again persistence modules and can be computed object-wise. The zero object of this abelian category is the persistence module $0 \colon (\mathbb{R}, \leq) \to \textbf{Vect}$ all of whose components are 0.

Remark 2.3. In some cases, we will also consider *graded persistence modules*, which are functors $M \colon (\mathbb{R}, \leq) \to \textbf{GrVect}$, where $\textbf{GrVect}$ is the category of $\mathbb{N}_0$-graded vector spaces over $\Bbbk$. Most of the content of this section generalises to the graded case in a completely straightforward way, so to avoid too much duplication, we only state it for the ungraded case.

To be able to extract any sort of useful information from persistence modules, we need to understand their structure. One way of doing this is by decomposing them into indecomposable summands. The indecomposables relevant in our case are known as interval modules.

Definition 2.4. A persistence module $M \colon (\mathbb{R}, \leq) \to \textbf{Vect}$ is *indecomposable* if $M \cong M_1 \oplus M_2$ implies that either $M_1 \cong 0$ or $M_2 \cong 0$.

Definition 2.5. Let $J \subseteq \mathbb{R}$ be an interval. The *interval module* $\Bbbk J \colon (\mathbb{R}, \leq) \to \textbf{Vect}$ is defined as

$$\Bbbk J(x) = \begin{cases} \Bbbk; & \text{if } x \in J, \\ 0; & \text{otherwise,} \end{cases}$$

and

$$\Bbbk J(x \leq y) = \begin{cases} \mathrm{id}_{\Bbbk}; & \text{if } x, y \in J, \\ 0; & \text{otherwise.} \end{cases}$$

One of the main features of persistence modules over $(\mathbb{R}, \leq)$ that makes them useful in TDA is that they can frequently be decomposed as direct sums of interval modules. When such a decomposition exists, it is unique [6]. The following version of the decomposition theorem is originally due to Crawley-Boevey [19]. In the case of persistence modules over $(\mathbb{R}, \leq)$, it can be stated as follows.

Theorem 2.6. *Suppose $M \colon (\mathbb{R}, \leq) \to \mathbf{Vect}$ is a persistence module such that $M(x)$ is finite dimensional for every $x \in \mathbb{R}$. Then M has a decomposition into interval modules.*

Whenever a persistence module $M \colon (\mathbb{R}, \leq) \to \mathbf{Vect}$ decomposes as a sum of interval modules, we can represent it using a *persistence barcode*. This is defined as the multiset of all intervals that occur in the decomposition. Sometimes we represent these intervals as pairs (a, b) where a is the startpoint and b is the endpoint of an interval in the decomposition. (These points are sometimes decorated to preserve information regarding which types of intervals the points correspond to, see [17] for details.) The multiset of such pairs is called the *persistence diagram* corresponding to M. The notion of persistence diagram can be generalised to some cases where the interval decomposition does not exist [17].

We will often concentrate on the case of *finitely presented* persistence modules. Note that a persistence module is finitely presented if and only if it is isomorphic to a finite direct sum of half-open interval modules $\Bbbk[a, b)$, where $-\infty < a < b \leq \infty$.

Definition 2.7. The *tensor product* of two persistence modules M and N is given by

$$(M \otimes N)(s) = \mathrm{colim}_{s_1 + s_2 \leq s} M(s_1) \otimes N(s_2).$$

Thus $(M \otimes N)(s)$ is the quotient of $\bigoplus_{s_1 + s_2 = s} M(s_1) \otimes M(s_2)$ obtained as follows. Suppose given u_1, u_2 with $u_1 + u_2 \leq s$. Then for any pair v_1, v_2 with $v_1 + v_2 = s$ and $u_1 \leq v_1$, $u_2 \leq v_2$, we have a composite

$$M(u_1) \otimes N(u_2) \to M(v_1) \otimes N(v_2) \hookrightarrow \bigoplus_{s_1 + s_2 = s} M(s_1) \otimes M(s_2). \qquad (1)$$

Then $(M \otimes N)(s)$ is the largest quotient of $\bigoplus_{s_1 + s_2 = s} M(s_1) \otimes M(s_2)$ with the property that for all u_1, u_2 all such composites (1) coincide, regardless of the choice of v_1, v_2. See Section 3.2 of [13] or Section 2.2 of [43].

The operation of tensoring with a fixed persistence module is right exact but not exact, and therefore induces *derived functors* denoted by $M, N \mapsto \mathrm{Tor}_i(M, N)$ for $i \geq 0$, with $\mathrm{Tor}_0(M, N) = M \otimes N$.

For finitely presented persistence modules the tensor products and Tor-functors can be described explicitly. In order to do this, it suffices to explain what happens for interval modules. Given interval modules $\Bbbk[a, b)$ and $\Bbbk[c, d)$, we have

$$\Bbbk[a, b) \otimes \Bbbk[c, d) = \Bbbk[a + c, \min(a + d, b + c)),$$
$$\mathrm{Tor}_1(\Bbbk[a, b), \Bbbk[c, d)) = \Bbbk[\max(a + d, b + c), b + d),$$
$$\mathrm{Tor}_i(\Bbbk[a, b), \Bbbk[c, d)) = 0 \text{ for } i \geq 2.$$

See Example 7.1 of [13].

2.2. **Persistent homology.** Persistence modules have an important role in TDA, where they are used in order to study data sets in the form of finite metric spaces, also known as *point clouds*. The idea is to take a finite metric space X and convert it into a simplicial complex (or topological space or other topological object) Y equipped with an $(\mathbb{R}, \leq)$-filtration, i.e. a system of subsets $Y_r \subseteq Y$ for $r \in \mathbb{R}$, such that $\bigcup_{r \in \mathbb{R}} Y_r = Y$ and $Y_r \subseteq Y_{r'}$ for $r < r'$. There are many such constructions, and they are often based on the principle that Y_r should capture the behaviour of X 'at length scale r'. Given such an $(\mathbb{R}, \leq)$-filtered complex Y, the assignment $r \mapsto Y_r$ defines a functor from $(\mathbb{R}, \leq)$ into simplicial complexes (or topological spaces, or other appropriate codomain). So taking the homology of the Y_r then produces a graded persistence module

$$r \longmapsto H_*(Y_r).$$

These persistence modules are called the *persistent homology* of the original object X. Once the persistent homology of X has been obtained, the resulting barcode is then analysed. The bars are regarded as features of the metric space X. Longer or *persistent* bars are often regarded as genuine features, while shorter bars are often regarded as noise, though there are other interpretations of the barcode.

Here we will describe some important examples of this general construction, starting with the Vietoris-Rips filtration and the Čech filtration.

Definition 2.8. Suppose (X, d) is a finite metric space. We define the *Vietoris-Rips complex* $\mathcal{R}(X)$ of (X, d) to be the $(\mathbb{R}, \leq)$-filtered simplicial complex with vertex set X, in which the simplices of the r-th filtration step $\mathcal{R}_r(X)$ are defined by the rule

$$\sigma \in \mathcal{R}_r(X) \Leftrightarrow \mathrm{diam}\, \sigma \leq r.$$

In some cases, we consider X as a subspace of some larger metric space Y, e.g. $Y = \mathbb{R}^n$. In this case we can define the corresponding Čech complex as follows:

Definition 2.9. Suppose (Y, d) is a metric space and $X \subseteq Y$ is a finite subset. We define the *Čech complex* $\check{\mathcal{C}}(X)$ associated to X to be the $(\mathbb{R}, \leq)$-filtered simplicial complex with vertex set X, in which the simplices of the r-th filtration step $\check{\mathcal{C}}_r(X)$ are defined by the rule

$$\sigma \in \check{\mathcal{C}}_r(X) \iff \bigcap_{x \in \sigma} B(x, r) \neq \emptyset,$$

where $B(x, r)$ denotes the open ball in Y with centre x and radius r.

Note that both the Vietoris-Rips and the Čech complex are filtrations of the simplex spanned by the vertices of X.

A related source of persistence modules are sublevel set filtrations. These are associated to a function $f\colon X \to \mathbb{R}$, where X is a topological space. They are motivated by ideas of Morse theory, where $X = M$ is assumed to be a smooth manifold and f is a Morse function (smooth function whose critical points are nondegenerate).

Definition 2.10. Let $f\colon X \to \mathbb{R}$ be a (continuous) function on a topological space X. The *sublevel set filtration* associated to (X, f) is the family $(X^a)_{a\in\mathbb{R}}$ where $X^a = f^{-1}(-\infty, a]$, which can also be viewed as a functor $S\colon (\mathbb{R}, \leq) \to$ **Top**. Composing this functor with k-th singular homology yields a persistence module $H_k \circ S$ which is called the *k-th sublevel set persistent homology of (X, f)*.

Other examples of persistence modules that have been used are lower star filtrations of simplicial complexes, alpha (or Delaunay) complexes, wrap complexes, witness complexes, and many more besides [22, 20, 8].

In order to ensure stability of persistence modules arising in applications despite the noise arising from imprecise measurements, it is important to be able to use approximation techniques. This is done using the notion of ϵ-interleavings. These provide a way to formalise the intuitive notion of approximate isomorphism of persistence modules and can be used to define a notion of distance on the category of persistence modules. For details, see [17, 14, 38].

3. Background on magnitude of metric spaces

In this section we will introduce the magnitude of metric spaces. This is a numerical invariant of metric spaces developed by Tom Leinster in [34], building on earlier work defining numerical invariants of categories [33]. Despite these abstract origins, magnitude turns out to be an interesting invariant containing meaningful geometric information. Here we will introduce the basics and attempt to give readers an impression of magnitude's interest and reach. Readers who wish to know more are strongly recommended to take a look at Leinster's original paper [34] and Leinster and Meckes's survey [35]. We note here that magnitude of metric spaces is just one instance of a more general invariant of enriched categories. The latter is developed in section 1 of [34], and we will not say anything about it here.

Here, and in the rest of the paper, we will use the symbol $|X|$ to denote the magnitude of an object X. To avoid notational clashes, we will use the symbol $\#X$ to denote the cardinality of a finite set X.

3.1. **Magnitude of finite metric spaces.** We begin with the magnitude of finite metric spaces. This is based almost entirely on section 2 of [34].

Definition 3.1 (Magnitude via weightings). Let (X, d) be finite metric space. A *weighting* on X is a function $w\colon X \to \mathbb{R}$ such that the equality

$$\sum_{y\in X} e^{-d(x,y)} w(y) = 1$$

is satisfied for every $x \in X$. If X admits a weighting, then we define the *magnitude* of X to be

$$|X| = \sum_{x \in X} w(x).$$

This is independent of the choice of weighting. If no weighting exists, then the magnitude of X is not defined.

Remark 3.2 (Magnitude via matrices). Suppose that X is a finite metric space with elements $x_1, \ldots, x_n$, and let Z_X denote the $n \times n$ matrix with $(Z_X)_{ij} = e^{-d(x_i, x_j)}$. If it happens that Z_X is invertible, then the magnitude of X is defined and is given by the formula

$$|X| = \sum_{i,j=1}^{n} (Z_X^{-1})_{ij}. \tag{2}$$

It can happen that $|X|$ is defined (using weightings) in cases where Z_X is not invertible. (See Lemma 1.1.4 of [34].)

For $t > 0$, we let tX be the metric space X rescaled by t, so that $d_{tX}(x, y) = t d_X(x, y)$. There is no simple relationship between $|tX|$ and $|X|$, and as a consequence we gain information by considering all rescalings at once, as in the following definition.

Definition 3.3 (The magnitude function). Let X be a finite metric space. Its *magnitude function* is the (partially defined) function from $(0, \infty)$ to $\mathbb{R}$ given by

$$t \mapsto |tX|.$$

Example 3.4 (Magnitude of the one-point space). Let X denote the space consisting of a single point x. Then Z_X is the 1×1 matrix (1), so that $Z_X^{-1} = (1)$ and formula (2) gives us $|X| = 1$.

Example 3.5 (Magnitude of two-point spaces). Let $X = \{x_1, x_2\}$ be the two-point space in which $d_X(x_1, x_2) = d$ for some $d > 0$. Then

$$Z_X = \begin{pmatrix} 1 & e^{-d} \\ e^{-d} & 1 \end{pmatrix}$$

so that

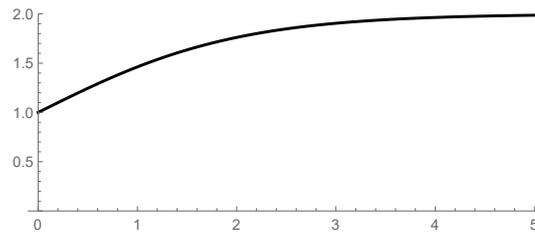$$Z_X^{-1} = \frac{1}{1 - e^{-2d}} \begin{pmatrix} 1 & -e^{-d} \\ -e^{-d} & 1 \end{pmatrix}$$

and consequently

$$|X| = \frac{2 - 2e^{-d}}{1 - e^{-2d}} = \frac{2}{1 + e^{-d}}.$$

The same computation shows that the magnitude function of X is given by

$$|tX| = \frac{2}{1 + e^{-dt}}$$

with graph:

We see in this case that $|tX|$ varies between 1 and 2, tending to 1 as $t \to 0$ and to 2 as $t \to \infty$. This suggests that magnitude is an 'effective number of points', regarding two points as essentially the same if they are very close, and essentially different if they are very far apart. The latter property generalises.

Proposition 3.6 (Leinster [34, Proposition 2.2.6])**.** *Let X be a finite metric space. Then $|tX| \to \#X$ as $t \to \infty$, where $\#X$ denotes the cardinality of X.*

Example 3.7.
 - It is known that all metric spaces with four points or less have magnitude, but there exist spaces with five or more points that do not have magnitude (See pages 870-871 of [34]).
 - There is a simple formula due to Speyer for the magnitude of *homogeneous* metric spaces, i.e. those that admit a transitive group action (see Proposition 2.1.5 of [34]). This allows one to compute magnitude of many simple spaces, for example complete graphs and cyclic graphs. (Graphs are always regarded as metric spaces by equipping them with the shortest path metric.)
 - The magnitude function of a finite metric space X can take negative values, it can take values greater than $\#X$, and it can have intervals on which it is increasing or decreasing. Example 2.2.7 of [34] gives a demonstration of this on a space X with 5 points.
 - It is not always true that $|tX| \to 1$ as $t \to 0$. An example due to Willerton describes a metric space with 6 points for which $|tX| \to 6/5$ as $t \to 0$. (See Example 2.2.8 of [34].)

Remark 3.8 (Magnitude and data)**.** One may take a data set in the form of a finite subspace of Euclidean space, and take its magnitude or magnitude function, which in this case is always defined. The result is a potentially interesting invariant of such data sets. But for this to be useful, one would like to know that the invariant is stable under perturbations of the data set. In mathematical terms, one would like to know that magnitude is continuous with respect to the Hausdorff metric on subsets of Euclidean space. This is currently unknown, although Meckes has shown that in this situation the function $X \mapsto |X|$ is *lower semicontinuous*, meaning roughly that magnitude may jump upwards but not downwards. (See Theorem 2.6 of [39] and the paragraph that follows it.)

3.2. **Magnitude of compact metric spaces.** Magnitude also makes sense for certain classes of compact, infinite metric spaces. Here we will recall the relevant definition and some of the main results. Good references for this section are section 3 of [34] and the survey [35].

In the following we will consider *positive definite* metric spaces, which are metric spaces X with the property that for every finite subspace F the matrix

Z_F is positive definite. Any subset of Euclidean space, with its induced metric, is positive definite.

Definition 3.9. Let (X, d) be a compact positive definite metric space. *The magnitude* of X is defined by the formula:

$$|X| = \sup\{|W| : W \subseteq X, \ W \ \text{finite.}\}$$

The *magnitude function* of X is defined by $t \mapsto |tX|$ for $t \in [0, \infty)$.

Example 3.10. We let S^1_{eucl} denote the Euclidean circle, i.e. the unit circle in the plane with its induced metric. And we let S^1_{geo} denote the same circle with its geodesic metric of total arclength 2π. Both are positive definite. Then the magnitude function of the Euclidean circle is given by

$$|t \cdot S^1_{\text{eucl}}| = \pi t + O(t^{-1}) \qquad \text{as} \qquad t \to \infty$$

and the magnitude function of the geodesic circle is given by

$$|t \cdot S^1_{\text{geo}}| = \frac{\pi t}{1 - e^{-\pi t}}.$$

See Theorems 13 and 14 of [37]. It has been argued [37] that the linear term πt in these expressions corresponds to half the length of the circle, whereas the absence of the constant term corresponds to the fact that the Euler characteristic of the circle is zero.

Remark 3.11. In general, computing the magnitude of infinite spaces is difficult, and existing computations tend to require a significant amount of analysis. A useful survey on this subject is given in [35]. Important recent progress by Gimperlein and Goffeng [24] shows that for appropriate $X \subseteq \mathbb{R}^{2n+1}$, the asymptotics of the magnitude function as $t \to \infty$ encode geometric properties including volume, surface area and mean curvature.

4. Background on magnitude homology

Singular homology can be regarded as a *categorification* of the Euler characteristic: The Euler characteristic is a function taking values in the set of integers, whereas homology is a functor taking values in the category of graded abelian groups, and the function can be obtained from the functor by taking the alternating sum of the ranks:

$$\chi(X) = \sum_{i=0}^{\infty} (-1)^i \operatorname{rank} H_i(X)$$

This is a classical story, but there are more recent examples of such categorifications, notably Khovanov homology, which categorifies the Jones polynomial, and Knot Floer homology, which categorifies the Alexander polynomial.

Hepworth and Willerton [31] together with Leinster and Shulman [36] introduced *magnitude homology*, a categorification of magnitude. (Precisely, Hepworth and Willerton first introduced magnitude homology in the case of graphs, and Leinster and Shulman later extended this to arbitrary metric spaces and very general enriched categories.) More recently, Nina Otter [42] introduced a persistent version of magnitude homology called *blurred magnitude homology*.

In this section we will introduce magnitude homology and its blurred variant, and we will conclude by giving an explicit formula to extract the magnitude of a space from the barcode of its blurred magnitude homology. It is this story that we will *reverse* in the rest of the paper, using its conclusion as the *definition* of the magnitude of persistence modules. Applying this to persistent homology theories other than blurred magnitude homology, we will then obtain new notions of magnitude of metric spaces.

4.1. **Magnitude homology.** Given a metric space X and elements $x_0, \ldots, x_k \in X$, we define

$$\ell(x_0, \ldots, x_k) = d(x_0, x_1) + d(x_1, x_2) + \cdots + d(x_{k-1}, x_k).$$

We think of this as the *length* of the tuple $(x_0, \ldots, x_k)$.

Definition 4.1 (Hepworth-Willerton [31], Leinster-Shulman [36]). The *magnitude chain complex* of a metric space X consists of the abelian groups

$$\mathrm{MC}_{k,l}(X) = \left\langle (x_0, \ldots, x_k) \in X^{k+1} \ \middle| \ \begin{array}{l} x_0 \neq x_1 \neq \cdots \neq x_k, \\ \ell(x_0, \ldots, x_k) = l \end{array} \right\rangle$$

with $l \in [0, \infty)$ and k a non-negative integer. Here, and in what follows, angled brackets $\langle \ \rangle$ denote free $\mathbb{Z}$-modules. The boundary operators

$$\partial_{k,l} \colon \mathrm{MC}_{k,l}(X) \to \mathrm{MC}_{k-1,l}(X)$$

are defined by the rule

$$\partial_{k,l}(x_0, \ldots, x_k) = \sum_{i=0}^{k} (-1)^i (x_0, \ldots, \widehat{x_i}, \ldots, x_k),$$

where the term $(x_0, \ldots, \widehat{x_i}, \ldots, x_k)$ is omitted if $\ell(x_0, \ldots, \widehat{x_i}, \ldots, x_k) < l$. The *magnitude homology* $\mathrm{MH}_{k,l}(X)$ of X is defined to be the homology of the magnitude chains

$$\mathrm{MH}_{k,l}(X) = H_k(\mathrm{MC}_{*,l}(X))$$

where again k is a non-negative integer and $l \in [0, \infty)$.

Magnitude homology is a categorification of the magnitude, in the sense that the graded Euler characteristic of magnitude homology coincides with the magnitude itself, as shown in the next proposition. This is categorification in the same sense that Khovanov homology categorifies the Jones polynomial, and that knot Floer homology categorifies the Alexander polynomial.

Proposition 4.2 (Hepworth-Willerton [31], Leinster-Shulman [36]). *Let X be a finite metric space. Then*

$$|tX| = \sum_{l \in [0, \infty)} \sum_{k=0}^{\infty} (-1)^k \operatorname{rank}(\mathrm{MH}_{k,l}(X)) e^{-lt}$$

for t sufficiently large.

Remark 4.3. The formula above requires some elaboration. Consider the set of real numbers occuring as $\ell(x_0, \ldots, x_k)$ for $x_0, \ldots, x_k \in X$, $k \geq 0$, with consecutive x_i's distinct. Let us call these ℓ-*values*. Since X is finite, there is a
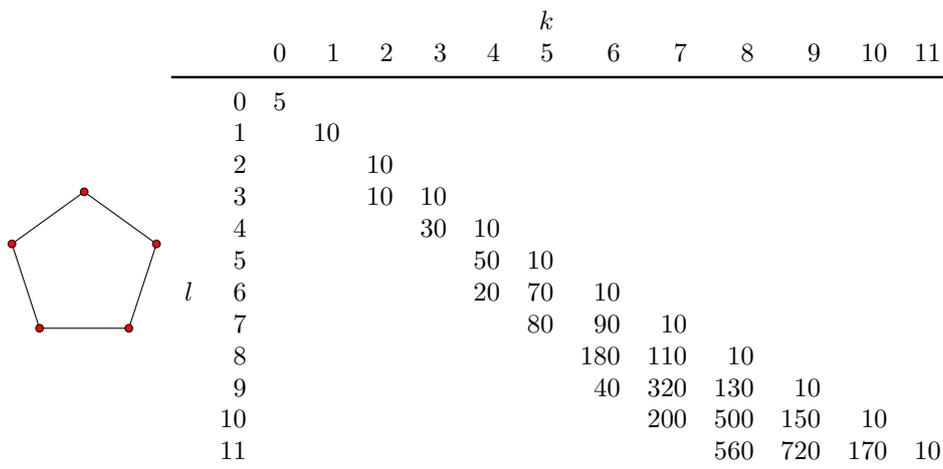
| | | | | | | k | | | | | | |
	0	1	2	3	4	5	6	7	8	9	10	11
0	5											
1		10										
2			10									
3			10	10								
4				30	10							
5					50	10						
6					20	70	10					
7						80	90	10				
8							180	110	10			
9							40	320	130	10		
10								200	500	150	10	
11									560	720	170	10

(The l label appears to the left of row 6.)

FIGURE 1. The ranks of $\mathrm{MH}_{k,l}(C_5)$. (Taken from [31, Table 1])

positive minimum nonzero distance between elements of X, call it $\delta > 0$, and then all ℓ-values satisfy the following inequality:

$$\ell(x_0, \ldots, x_k) \geq \delta k \tag{3}$$

A first consequence of equation (3) is that, for a fixed choice of $l \in [0, \infty)$, the set of k for which $\mathrm{MH}_{k,l}(X) \neq 0$ is bounded above by l/δ. That is because if $\mathrm{MH}_{k,l}(X) \neq 0$ then l must be an ℓ-value $\ell(x_0, \ldots, x_k)$. It follows that in Proposition 4.2 the inner sum is finite for each l.

The second consequence of equation (3) is that for any positive real N, the collection of ℓ-values satisfying $\ell(x_0, \ldots, x_k) \leq N$ is finite (because then $k \leq N/\delta$, and X is finite). It follows that the set of all ℓ-values can be totally ordered $0 = l_0 < l_1 < l_2 < \cdots$. Thus the outer series in Proposition 4.2 can be rewritten as the (infinite) sum over the l_i.

Example 4.4 (Magnitude homology of graphs). A graph can be regarded as a metric space by taking the set of vertices and equipping them with the shortest path metric. This is the original setting of magnitude homology in [31], where a number of explicit examples (done using computer algebra) are described. We include two of these here as an illustration. Figure 1 shows the ranks of the magnitude homology $\mathrm{MH}_{k,l}(C_5)$ of the cyclic graph with 5 vertices, and Figure 2 shows the ranks of the magnitude homology $\mathrm{MH}_{k,l}(Petersen)$ of the Petersen graph. (Note that the images and tables in Figures 1 and 2 are taken directly from [31].) Observe that in each case, the rank of $\mathrm{MH}_{0,0}(G)$ is the number of vertices, and the rank of $\mathrm{MH}_{1,1}(G)$ is the number of oriented edges. These are general features, but the question of what data is encoded in $\mathrm{MH}_{k,l}(G)$ for other choices of k and l remains mysterious. Another general feature visible here is that the nonzero magnitude homology groups lie in a range of pairs (k, j) bounded by two diagonals, one of them the diagonal $k = j$, and the other determined by the diameter of the graph.

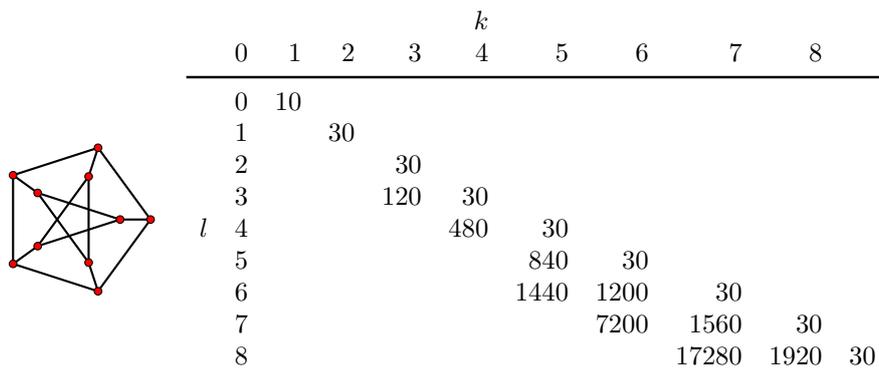Magnitude homology has many good characteristics of homology theories and categorification:

	k								
l	0	1	2	3	4	5	6	7	8
0	10								
1		30							
2			30						
3			120	30					
4				480	30				
5					840	30			
6					1440	1200	30		
7						7200	1560	30	
8							17280	1920	30

FIGURE 2. The ranks of $\mathrm{MH}_{k,l}(Petersen)$. (Taken from [31, A.3.1].)

- Magnitude homology refines magnitude: there are finite metric spaces with the same magnitude but non-isomorphic magnitude homologies [27].
- Magnitude homology can contain torsion [32]. Thus the magnitude homology contains more data than just the ranks $\mathrm{rank}(\mathrm{MH}_{k,l}(X))$.
- Magnitude homology has properties that categorify known properties of the magnitude. In the setting of graphs, a Künneth theorem categorifies the known product formula for magnitude, and a Mayer-Vietoris sequence categorifies the known inclusion-exclusion formula. See [31].
- Magnitude homology contains information about geometric features of a metric space, for example it can precisely detect the property of being Menger convex, and it contains obstructions to the existence of upper bounds on curvature, and to the existence of closed geodesics. See [36], [25], [5].
- Magnitude homology has been computed fully in several interesting examples, including trees, complete graphs, cycle graphs, and the icosahedral graph. See [31] and [27].

4.2. **Blurred magnitude homology.** We now describe some recent work of Nina Otter [42] that connects magnitude homology with persistent homology, specifically the Vietoris-Rips complex. We also give a new result that relates magnitude with barcodes for the first time.

Definition 4.5 (Otter [42]). The *blurred magnitude chain complex* of a metric space X is the chain complex of persistence modules $\mathrm{BMC}_*(X)$ defined by the rule

$$\mathrm{BMC}_k(X)(l) = \left\langle (x_0, \ldots, x_k) \in V^{k+1} \; \middle| \; \begin{array}{l} x_0 \neq x_1 \neq \cdots \neq x_k, \\ \ell(x_0, \ldots, x_k) \leq l \end{array} \right\rangle$$

where l is the persistence parameter and k is a non-negative integer. The boundary operators

$$\partial_k \colon \mathrm{BMC}_k(X) \to \mathrm{BMC}_{k-1}(X)$$

are defined by the rule

$$\partial_{k,l}(x_0, \ldots, x_k) = \sum_{i=0}^{k} (-1)^i (x_0, \ldots, \widehat{x_i}, \ldots, x_k).$$

The *blurred magnitude homology* $\mathrm{BMH}_*(X)$ of X is defined to be the homology of the blurred magnitude chains:

$$\mathrm{BMH}_k(X) = H_k(\mathrm{BMC}_*(X))$$

for k a non-negative integer.

Remark 4.6 (Blurred magnitude homology and the Rips complex)**.** One of the main results of Otter's paper [42] is that it compares the blurred magnitude homology of a metric space X with the homology of its Rips complex. The main idea of this comparison is that there are maps

$$\mathrm{BMC}_k(X)(s) \to C_k(\mathcal{R}^{\mathrm{sim}}(X)(s)) \quad \text{and} \quad C_k(\mathcal{R}^{\mathrm{sim}}(X)(s)) \to \mathrm{BMC}_k(X)(ks),$$

where $\mathcal{R}^{\mathrm{sim}}(X)$ denotes a variant of the Rips chain complex, having the same persistent homology. These comparison maps are a multiplicative version of an interleaving, and although the constant appearing here is the degree k in the chain complex, and in particular is not constant, it is nevertheless sufficient for Otter to prove the following theorem.

Theorem 4.7 (Otter [42, Theorem 32])**.**

$$\lim_{0 \leftarrow \epsilon} \mathrm{BMH}_*(X)(\epsilon) \cong \lim_{0 \leftarrow \epsilon} H_*(\mathcal{R}(X)(\epsilon))$$

The quantity $\lim_{0 \leftarrow \epsilon} H_*(\mathcal{R}_*(X)(\epsilon))$ is the *Vietoris homology* of X, a version of homology developed for metric spaces. In good cases, e.g. when X is a compact Riemannian manifold, it coincides with the singular homology of X. This theorem therefore demonstrates for the first time a concrete connection between magnitude homology and ordinary homology of spaces.

We will now state a new result that gives the relation between magnitude and the barcode decomposition of the blurred magnitude homology. First note that by standard homological algebra, the definitions stated above have the following immediate consequence, which relates ordinary and blurred magnitude homology.

Proposition 4.8. *Let X be a finite metric space and let $0 = l_0 < l_1 < l_2 < \cdots$ be the distinct real numbers occuring as $\ell(x_0, x_1, \ldots, x_k)$ for $x_0, \ldots, x_k \in X$, $k \geq 0$. Then for each $k \geq 0$ and $j > 0$ we have a short exact sequence:*

$$0 \to \mathrm{BMC}_k(X)(l_{j-1}) \to \mathrm{BMC}_k(X)(l_j) \to \mathrm{MC}_{k,l_j}(X) \to 0$$

Consequently in homology there is a long exact sequence:

$$\cdots \to \mathrm{BMH}_k(X)(l_{j-1}) \to \mathrm{BMH}_k(X)(l_j) \to \mathrm{MH}_{k,l_j}(X) \to \cdots$$

Our result can then be stated as follows. Its proof is rather long and technical, thanks to convergence issues.

Theorem 4.9. *Let X be a finite metric space and let $\mathrm{BMH}_*(X)$ denote its blurred magnitude homology. Suppose that $\mathrm{BMH}_*(X)$ has barcode whose bars in*

degree $k \geq 0$ *are* $[a_{k,0}, b_{k,0}), [a_{k,1}, b_{k,1}), \ldots$. *Then the magnitude of* X *is given by the formula*

$$|tX| = \sum_{k=0}^{\infty} \sum_{i=1}^{m_k} (-1)^k (e^{-a_{k,i}t} - e^{-b_{k,i}t})$$

for t *sufficiently large.*

Proof. Throughout this proof we let δ denote the minimum nonzero distance between elements of X, and we let n denote the cardinality of X. We let $l_0 < l_1 < l_2 < \cdots$ be the distinct values of l for which the inner sum appearing in Proposition 4.2 is nonzero, as in Remark 4.3. And we define $D(i, j, k)$ to be 1 if l_j is in $[a_{k,i}, b_{k,i})$, and to be 0 otherwise. We make a standing assumption that t is large enough that $ne^{-\delta t} < 1$; this is the assumption under which t is large enough that the conclusions of Proposition 4.2 hold.

We will use the following fact several times. Let $(x_0, \ldots, x_k)$ be a tuple of elements of X in which consecutive elements are distinct, and suppose that this tuple is a generator of $\mathrm{MC}_{k,l}(X)$ or $\mathrm{BMC}_k(X)(l)$. Then $\ell(x_0, \ldots, x_k) \leq l$, while $\ell(x_0, \ldots, x_k) \geq k\delta$, so that $k\delta \leq l$. It follows that, if k and l do not satisfy this relation, then the homology groups $\mathrm{MH}_{k,l}(X)$ and $\mathrm{BMH}_k(X)(l)$ vanish.

We now have the following computation, whose steps will be justified below.

$$|tX| \overset{1}{=} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (-1)^k \operatorname{rank}(\mathrm{MH}_{k,l_j}(X)) e^{-l_j t}$$

$$\overset{2}{=} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (-1)^k \left[\operatorname{rank}(\mathrm{BMH}_k(X)(l_j)) - \operatorname{rank}(\mathrm{BMH}_k(X)(l_{j-1})) \right] e^{-l_j t}$$

$$\overset{3}{=} \sum_{k=0}^{\infty} (-1)^k \sum_{j=0}^{\infty} \left[\operatorname{rank}(\mathrm{BMH}_k(X)(l_j)) - \operatorname{rank}(\mathrm{BMH}_k(X)(l_{j-1})) \right] e^{-l_j t}$$

$$\overset{4}{=} \sum_{k=0}^{\infty} (-1)^k \sum_{j=0}^{\infty} \operatorname{rank}(\mathrm{BMH}_k(X)(l_j))(e^{-l_j t} - e^{-l_{j+1} t})$$

$$\overset{5}{=} \sum_{k=0}^{\infty} (-1)^k \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} D(i,j,k)(e^{-l_j t} - e^{-l_{j+1} t})$$

$$\overset{6}{=} \sum_{k=0}^{\infty} (-1)^k \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} D(i,j,k)(e^{-l_j t} - e^{-l_{j+1} t})$$

$$\overset{7}{=} \sum_{k=0}^{\infty} (-1)^k \sum_{i=0}^{\infty} (e^{-a_{k,i}t} - e^{-b_{k,i}t})$$

Step 1 is precisely the formula of Proposition 4.2. The series here is absolutely convergent. That is because

$$\sum_{j=0}^{J}\sum_{k=0}^{\infty}\mathrm{rank}(\mathrm{MH}_{k,l_j}(X))e^{-l_j t} = \sum_{k=0}^{\infty}\sum_{j=0}^{J}\mathrm{rank}(\mathrm{MH}_{k,l_j}(X))e^{-l_j t}$$

$$\leq \sum_{k=0}^{\infty}\sum_{x_0,\dots,x_k} e^{-\ell(x_0,\dots,x_k)t}$$

$$\leq \sum_{k=0}^{\infty} n^{k+1}e^{-k\delta t}$$

$$= n \cdot \sum_{k=1}^{\infty}(ne^{-\delta t})^k.$$

Here, in the second line the inner sum is over all tuples $(x_0,\dots,x_k)$ with consecutive elements distinct, and there are at most n^{k+1} of these, where n denotes the cardinality of X. Now we have $ne^{-\delta t} < 1$ by our standing assumption, so that the latter sum converges and is bounded above independent of J. This shows absolute convergence.

To explain step 2, recall that by Proposition 4.8 there is a long exact sequence

$$\cdots \to \mathrm{BMH}_k(X)(l_{j-1}) \to \mathrm{BMH}_k(X)(l_j) \to \mathrm{MH}_{k,l_j}(X) \to \cdots$$

This sequence terminates in both directions, because the relation described in the second paragraph above fails for all three groups when k is large enough. A standard fact from homological algebra then guarantees that

$$\sum_{k=0}^{\infty}(-1)^k\,\mathrm{rank}(\mathrm{MH}_{k,l_j}(X)) =$$

$$\sum_{k=0}^{\infty}(-1)^k\left[\mathrm{rank}(\mathrm{BMH}_k(X)(l_j)) - \mathrm{rank}(\mathrm{BMH}_k(X))(l_{j-1})\right].$$

For step 3, we have exchanged the order of summation. This is valid because the series is absolutely convergent (indeed, it is the same series as the one appearing in step 1).

For step 4, we have 'telescoped' the sum, using the fact that

$$\mathrm{rank}(\mathrm{BMH}_k(X)(l_j))e^{-l_j t} \to 0 \text{ as } j \to \infty.$$

The latter holds because $\mathrm{rank}(\mathrm{BMH}_k(X)(l_j))$ is at most the number of tuples $(x_0,\dots,x_k)$ with consecutive entries distinct and $\ell(x_0,\dots,x_k) \leq l_j$. But then $l_j \geq k\delta$ so that $\mathrm{rank}(\mathrm{BMH}_k(X)(l_j))e^{-l_j t} \leq n^{k+1}e^{-l_j t} \leq n \cdot n^{l_j/\delta}e^{-l_j t} = n \cdot (n^{1/\delta}e^{-t})^{l_j}$. But $(n^{1/\delta}e^{-t}) < 1$ by our standing assumption. Since $l_j \to \infty$ as $j \to \infty$, the claim follows.

Step 5 follows by simply describing $\mathrm{rank}(\mathrm{BMH}_k(X)(l_j))$ as the number of bars in the barcode decomposition for $\mathrm{BMH}_k(X)$ that contain l_j. For step 6 we have again exchanged the order of summation, which is valid because the series consists of non-negative numbers and is convergent. Step 7 is then a direct computation of the series $\sum_{j=0}^{\infty} D(i,j,k)(e^{-l_j t} - e^{-l_{j+1} t})$. $\qquad\square$

5. Magnitude of persistence modules

In Theorem 4.9 in the previous section, we saw a formula expressing the magnitude function of a finite metric space X in terms of the barcode decomposition of its blurred magnitude homology. In this section we will turn that result on its head, and use the formula to *define* a numerical invariant of persistence modules and graded persistence modules, and explore its formal properties.

In subsequent sections we will apply our new invariant to persistent homology groups, in order to obtain new invariants of finite metric spaces (or of whatever input the persistent homology theory accepts).

In this section we will usually work with finitely presented persistence modules, and finitely presented graded persistence modules. In the latter case, we mean that the graded persistence module has finitely many generators and relations in total, so that it is nonzero in only finitely many degrees. Thus, our persistence modules will always be direct sums of finitely many interval modules of the form $\Bbbk[a, b)$ where possibly $b = \infty$. This restriction allows us to work with the most relevant examples such as Rips and Čech complexes while keeping technicalities to a minimum. At the end of the section we will offer two different perspectives on the persistent magnitude, via the derived associated graded module and the Laplace transform. These offer potential for extending the scope of persistent magnitude beyond the present setting.

5.1. **Persistent magnitude.**

Definition 5.1 (Persistent magnitude). Let M be a finitely presented persistence module with barcode decomposition

$$M \cong \bigoplus_{i=1}^{n} \Bbbk[a_i, b_i).$$

The *persistent magnitude* or simply *magnitude* of M is the real number

$$|M| = \sum_{i=1}^{n} (e^{-a_i} - e^{-b_i})$$

where by convention $e^{-\infty} = 0$.

Example 5.2. For interval modules we have $|\Bbbk[a, b)| = e^{-a} - e^{-b}$ and $|\Bbbk[a, \infty)| = e^{-a}$. Thus longer intervals have greater magnitude. This is in line with one of the general philosophies of persistent homology, that longer bars — the features that persist longer — are the more significant, while the shorter bars represent 'noise'. (But note that there are other interpretations of barcodes, especially of the shorter bars. For instance, it is shown in [12] that the short bars in the barcode of points sampled from a disk of constant curvature can be used to infer the curvature.) Note also that interval modules of fixed length have greater magnitude the closer they are to 0, i.e. the sooner they begin.

Proposition 5.3 (Additivity with respect to short exact sequences). *If*

$$0 \to M \to N \to P \to 0$$

is a short exact sequence of finitely presented persistence modules, then $|N| = |M| + |P|$.

We will give a proof of this proposition in section 5.4 below, and another proof in section 5.5.

Definition 5.4 (Persistent magnitude of graded persistence modules)**.** Let M_* be a finitely presented graded persistence module. The *persistent magnitude* of M_* is defined as follows:

$$|M_*| = \sum_i (-1)^i |M_i|$$

If C_* is a chain complex of persistence modules, then we obtain two graded persistence modules, namely C_* itself, and the homology $H_*(C)$. The persistent magnitude of these is related by the following result, whose proof is a standard consequence of additivity with respect to short exact sequences. (Compare with the proof of Theorem 2.44 of [28].)

Proposition 5.5. *Let C_* be a finitely presented chain complex of persistence modules. Then*

$$|H_*(C)| = |C_*|.$$

5.2. **Rescaling and the magnitude function.** In Definition 3.3 the magnitude of a finite metric space was extended from a number to a function by means of rescaling the metric space. We now do the same with persistent magnitude.

Definition 5.6 (Rescaling of persistence modules)**.** Given a persistence module M and a real number $t \in (0, \infty)$, we can define the *rescaled module tM* to be the new persistence module defined by

$$tM(s) = M(s/t)$$

for $s \in [0, \infty)$. More precisely, tM is obtained from M by precomposing with the functor from $[0, \infty)$ to itself that sends s to s/t. This operation extends to graded persistence modules and chain complexes of persistence modules in the evident way.

One can think of the definition of this rescaling operation as saying that features of M that occur at s become features of tM that occur at ts.

Example 5.7 (Rescaling intervals and barcodes)**.** One can check that $t\Bbbk[a, b) = \Bbbk[ta, tb)$. Thus the effect of the rescaling operation on the barcode of a finitely presented persistence module M is to simply rescale it by t: the barcode of tM is obtained from that of M by applying a scale factor of t in the horizontal direction.

Example 5.8 (Rescaling of metric spaces)**.** Rescaling of persistence modules is designed to interact nicely with rescaling of metric spaces. Recall that if X is a metric space and $t \in (0, \infty)$, then the rescaling tX is the metric space with the same underlying set and with metric defined by $d_{tX}(a, b) = td_X(a, b)$. Then one can check directly that the Rips and Čech complexes satisfy

$$tC_*(\mathcal{R}(X)) = C_*\mathcal{R}(tX) \quad \text{and} \quad tC_*(\check{\mathcal{C}}(X)) = C_*(\check{\mathcal{C}}(tX))$$

and similarly for the persistent homology.

Definition 5.9 (The persistent magnitude function)**.** The *persistent magnitude function* or simply *magnitude function* of a finitely presented persistence module M is the function $(0, \infty) \to \mathbb{R}$ defined by

$$t \longmapsto |tM|.$$

If M has direct sum decomposition $M \cong \bigoplus_{i=1}^{n} \Bbbk[a_i, b_i)$, then the magnitude function is given by the formula

$$|tM| = \sum_{i=1}^{n} (e^{-a_i t} - e^{-b_i t})$$

where again by convention $e^{-\infty} = 0$.

The extremal behaviour of the magnitude function singles out two special classes of bars, as we see in the next proposition. Its proof is an immediate consequence of the definitions.

Proposition 5.10. *Let M be a finitely presented persistence module. Then:*
- $\lim_{t \to 0} |tM|$ *is the number of bars in M of the form $\Bbbk[a, \infty)$.*
- *if all bars of M are contained in $[0, \infty)$ (or equivalently if $M(s) = 0$ for $s < 0$), then $\lim_{t \to \infty} |tM|$ is the number of bars in M of the form $\Bbbk[0, b)$, including the case $b = \infty$.*

One can think of this as follows: As $t \to 0$, we are scaling down the barcode of M, so that any finite bars eventually disappear at 0, while any infinite bars remain, but all become indistinguishable; in this limit the magnitude function simply counts the latter. As $t \to \infty$, we are scaling up the barcode of M, so that any bars which begin *after* 0 eventually disappear at infinity, while all bars that begin at 0 remain but become indistinguishable; in this limit the magnitude function again just counts the latter.

Remark 5.11 (Reparameterisation of persistence modules)**.** Occasionally, it is useful to reparameterise a persistence module M by an orientation preserving homeomorphism $h \colon \mathbb{R} \to \mathbb{R}$. In this case, we can define the *reparameterised module hM* by:

$$hM(s) = M(h^{-1}(s))$$

The properties of this definition generalise the ones for rescaling by a positive real number in a natural way. For instance,

$$h\Bbbk[a, b) = \Bbbk[h(a), h(b))$$

and if $M \cong \bigoplus_{i=1}^{n} \Bbbk[a_i, b_i)$ we have

$$|hM| = \sum_{i=1}^{n} (e^{-h(a_i)} - e^{-h(b_i)}).$$

This definition has the following immediate but useful consequence:

Lemma 5.12. *Suppose M is a finitely presented persistence module with magnitude function*

$$|tM| = \sum_{i=1}^{n} \lambda_i e^{-r_i t}$$

and $h\colon \mathbb{R} \to \mathbb{R}$ *is an orientation preserving homeomorphism. Then the reparameterised module* hM *has magnitude function*

$$|thM| = \sum_{i=1}^{n} \lambda_i e^{-h(r_i)t}.$$

Remark 5.13 (Connection with the Euler characteristic of a barcode)**.** In section 6 of [9], Bobrowski and Borman define the *Euler characteristic* of a barcode with no bars of length ∞, or in other words of a finitely presented graded persistence module M with the property that $M(t) = 0$ for t sufficiently large. The definition is given by

$$\chi(M_*) = \sum_{i=1}^{b} (-1)^{|\beta_i|}(b_i - a_i)$$

where $\beta_1, \ldots, \beta_b$ are the bars of M_*, and $\beta_i = [a_i, b_i)$. They then describe a connection between the Euler integral of a function and the Euler characteristic of the barcode of the persistent homology of that function. Observe that the Euler characteristic of a finitely presented persistence module is related to its magnitude as follows:

$$\chi(M_*) = |M_*|'(0)$$

Thus the magnitude of a persistence module encodes its Euler characteristic. (In order to form $|M_*|'(0)$ above we have extended the magnitude function $t \mapsto |tM_*|$ to 0 in the evident way (see the formula in Definition 5.9) and then taken a one-sided derivative.)

5.3. **Persistent magnitude and products.** Now we will explore how the persistent magnitude interacts with the tensor product of persistence modules, which was described in Section 2.

Proposition 5.14 (Persistent magnitude respects products)**.** *Let M and N be finitely presented persistence modules. Then*

$$|M| \cdot |N| = |M \otimes N| - |\operatorname{Tor}_1(M, N)|. \tag{4}$$

Proof. It suffices to prove this when M and N are interval modules. In this case both sides of the equation can be computed directly using the results stated in Section 2. $\qquad\square$

Let us explain why equation (4) states that 'persistent magnitude respects tensor products', since it may look a little odd from that point of view. Homological algebra tells us that to fully understand the tensor product of M and N, we must consider the graded object $\operatorname{Tor}_*(M, N)$. (Serre's intersection formula in algebraic geometry is a good example of this principle in action.) Thus, the 'true' statement that persistent magnitude respects products would be

$$|M| \cdot |N| = |\operatorname{Tor}_*(M, N)| = \sum_{i=0}^{\infty} (-1)^i |\operatorname{Tor}_i(M, N)|.$$

But this reduces to exactly the equation appearing in the proposition.

The usefulness of Proposition 5.14 is that it leads to product preserving properties of the magnitude of persistent homology theories. Indeed, suppose given

a theory that assigns to objects X a graded persistence module $A_*(X)$, and suppose that objects X and Y can be equipped with a product $X \times Y$ in such a way that we have a Künneth theorem:

$$0 \to A_*(X) \otimes A_*(Y) \to A_*(X \times Y) \to \mathrm{Tor}_1(A_*(X), A_{*-1}(Y)) \to 0$$

Examples (and non-examples) of such Künneth theorems are discussed in [13], [43] and [15]. In such a setting, Proposition 5.14 can be used to prove the identity

$$|A_*(X \times Y)| = |A_*(X)| \cdot |A_*(Y)|.$$

Indeed, we have

$$
\begin{aligned}
|A_*(X \times Y)| &= |A_*(X) \otimes A_*(Y)| + |\mathrm{Tor}_1(A_*(X), A_{*-1}(Y))| \\
&= |A_*(X) \otimes A_*(Y)| - |\mathrm{Tor}_1(A_*(X), A_*(Y))| \\
&= |A_*(X)| \cdot |A_*(Y)| + |\mathrm{Tor}_1(A_*(X), A_*(Y))| - |\mathrm{Tor}_1(A_*(X), A_*(Y))| \\
&= |A_*(X)| \cdot |A_*(Y)|
\end{aligned}
$$

where the first equality comes from the short exact sequence (Proposition 5.3) and the third comes from (5.14).

5.4. **Persistent magnitude via derived associated graded modules.** In this subsection we will give a perspective on the definition of persistent magnitude using the homological algebra of the 'associated graded' or 'causal onset' functor. This will allow us to give a proof of Proposition 5.3. It also gives a potential avenue for extending the definition of persistent magnitude beyond the case of finitely presented modules.

As described in section 2 we denote by **PersMod** the category of persistence modules and by **GrMod** the category of $\mathbb{R}$-graded modules. These are abelian categories, and **PersMod** has enough projectives — they are the interval modules $\mathbb{k}[a, \infty)$ for $a \in \mathbb{R}$. For $a \in \mathbb{R}$ we let $\mathbb{k}_a$ denote the object of **GrMod** consisting of $\mathbb{k}$ in grading a and 0 in all other gradings.

Definition 5.15. The *associated graded functor* $\mathrm{Gr} \colon \mathbf{PersMod} \to \mathbf{GrMod}$ is defined by

$$\mathrm{Gr}(M)(s) = \frac{M(s)}{\sum_{s' < s} \mathrm{im}(M(s') \to M(s))}$$

for $s \in \mathbb{R}$. The functor Gr is right exact, and we denote its derived functors by $\mathrm{Gr}_i(M)$, $i \geq 0$, with $\mathrm{Gr}_0(M) = \mathrm{Gr}(M)$.

The terminology above was chosen because if the persistence module M is obtained from an $\mathbb{R}$-filtered vector space in the evident way, then $\mathrm{Gr}(M)$ is nothing other than the $\mathbb{R}$-graded vector space associated to M. The functor $M \mapsto \mathrm{Gr}(M)_s$ appears in [46], where it is denoted by $\mathcal{O}_s$, and called the *causal onset* functor.

Example 5.16. For a free module $\mathbb{k}[a, \infty)$ we have $\mathrm{Gr}_0(\mathbb{k}[a, \infty)) = \mathrm{Gr}(\mathbb{k}[a, \infty)) = \mathbb{k}_a$ and $\mathrm{Gr}_1(\mathbb{k}[a, \infty)) = 0$. And for an interval module $\mathbb{k}[a, b)$ with $a < b$ we have a free resolution

$$\cdots \to 0 \to \mathbb{k}[b, \infty) \to \mathbb{k}[a, \infty) \to \mathbb{k}[a, b) \to 0$$

so that $\mathrm{Gr}_0(\Bbbk[a,b)) = \Bbbk_a$ and $\mathrm{Gr}_1(\Bbbk[a,b)) = \Bbbk_b$. It follows that for finitely presented modules, $\mathrm{Gr}_i(M) = 0$ for $i > 1$.

Definition 5.17 (Graded magnitude)**.** The *graded magnitude* of a finitely presented object of **GrMod**, i.e. a module of the form $\bigoplus_{i=1}^{n} \Bbbk_{a_i}$ is

$$\left| \bigoplus_{i=1}^{n} \Bbbk_{a_i} \right| = \sum_{i=1}^{n} e^{-a_i}.$$

The graded magnitude function is clearly additive with respect to short exact sequences of finitely presented graded modules. The computations in Example 5.16 give the following.

Lemma 5.18. *The persistent and graded magnitude are related as follows. Let M be a finitely presented persistence module. Then*

$$|M| = |\mathrm{Gr}_0(M)| - |\mathrm{Gr}_1(M)|.$$

We are now in a position to prove that magnitude is additive with respect to short exact sequences.

Proof of Proposition 5.3. Like any derived functors, the Gr_i convert a short exact sequence

$$0 \to M \to N \to P \to 0$$

into a long exact sequence

$$0 \to \mathrm{Gr}_1(M) \to \mathrm{Gr}_1(N) \to \mathrm{Gr}_1(P) \to \mathrm{Gr}_0(M) \to \mathrm{Gr}_0(N) \to \mathrm{Gr}_0(P) \to 0,$$

and the statement of the proposition amounts to the claim that the alternating sum of the magnitudes of the modules in this sequence is zero. But this is a standard consequence of additivity with respect to short exact sequences, which in the case of graded magnitude is immediate from the definitions. (Compare with the proof of Theorem 2.44 of [28].) □

Remark 5.19. The proof of Proposition 5.3 shows that the magnitude of a persistence module M depends only on its 'derived associated graded' modules $\mathrm{Gr}_0(M)$ and $\mathrm{Gr}_1(M)$. In simpler terms, the magnitude depends not on the lengths of the bars in the barcode, but only on the collection of start and end points of bars in the barcode. One may then argue that magnitude does not contain any 'persistent' information, only 'graded' information. However, the same comment can apply to *any* invariant of persistence modules that is additive with respect to short exact sequences, thanks to the short exact sequences

$$0 \to \Bbbk[b, \infty) \to \Bbbk[a, \infty) \to \Bbbk[a, b) \to 0.$$

From the point of view of graded modules, persistence modules and persistent homology theories should perhaps then be regarded as an excellent source of interesting examples.

5.5. **Persistent magnitude and the Laplace transform.** Here is an alternative approach to the persistent magnitude using the Laplace transform, that we believe is the 'correct' way to understand persistent magnitude.

Let M be a finitely presented persistence module and let

$$\text{rank}(M)\colon \mathbb{R} \to \mathbb{R}$$

be its associated rank function, i.e. $\text{rank}(M)(s) = \text{rank}(M(s))$ for $s \in \mathbb{R}$. This is a step function, given by the sum of the indicator functions of the bars of M. It has a derivative in the distributional sense, given by

$$\text{rank}(M)' = \sum_{i=1}^{n}(\delta_{a_i} - \delta_{b_i})$$

where $M \cong \bigoplus_{i=1}^{n} \Bbbk[a_i, b_i)$, and where δ_x denotes the Dirac delta distribution supported at x. Recall that the *bidirectional Laplace transform* $\mathcal{L}\{f\}$ of a function or distribution f is given by

$$\mathcal{L}\{f\}(t) = \int_{-\infty}^{\infty} f(s)e^{-st}ds$$

for $t \in [0, \infty)$. Then one can check directly that

$$|tM| = \mathcal{L}\{\text{rank}'(M)\}(t).$$

In particular, the right hand side can be used as an alternative definition of magnitude function, whereas magnitude is recovered by evaluating at 1. For example, if $M = \Bbbk[a, b)$ then $\text{rank}(M)$ is the step function $1_{[a,b)}$ and $\text{rank}'(M) = \delta_a - \delta_b$ so that

$$\mathcal{L}\{\text{rank}'(M)\}(t) = \int_{-\infty}^{\infty} (\delta_a(s) - \delta_b(s))e^{-st}ds = e^{-at} - e^{-bt} = |tM|.$$

From this point of view, the additivity of persistent magnitude with respect to short exact sequences (Proposition 5.3) is an immediate consequence of the fact that if

$$0 \to M \to N \to P \to 0$$

is a short exact sequence of persistence modules, then $\text{rank}(N) = \text{rank}(M) + \text{rank}(P)$.

In the case of a finitely presented graded persistence module M_*, we can associate to it its Euler characteristic curve (see [44, Section 3.2] and [30, 10, 23] for some related work), i.e. $\chi(M_*)(s) = \sum_{i=0}^{\infty}(-1)^i \text{rank}(M_i(s))$ for $s \in \mathbb{R}$. Then we have:

$$|tM_*| = \sum_{i=0}^{\infty}(-1)^i|tM_i| = \sum_{i=0}^{\infty}(-1)^i\mathcal{L}\{\text{rank}'(M_i)\}(t) = \mathcal{L}\{\chi'(M_i)\}(t).$$

In other words, the magnitude of a finitely presented graded persistence module is precisely the Laplace transform of the derivative of its Euler characteristic curve. This provides yet another connection between magnitude and TDA.

More explicitly, whenever M_* is a finitely presented graded persistence module and $r_1 < \ldots < r_n < r_{n+1} = \infty$ is the sequence of all the startpoints and

endpoints in its interval decomposition, we have

$$|tM_*| = \sum_{j=1}^{n} \chi(M_*)(r_j)(e^{-r_j t} - e^{-r_{j+1}t}). \tag{5}$$

One interpretation of this formula is that persistent magnitude of a graded persistence module can be regarded as the 'filtered Euler characteristic' associated to it.

We include this perspective as a useful alternative point of view, as well as a potential avenue for generalising the magnitude from finitely presented modules to more general modules that, despite not being finitely presented, may nevertheless have a 'rank function' or 'rank distribution' that we can then differentiate and subject to the Laplace transform. (Here we recall [17], where persistence modules that do not admit a barcode decomposition are nevertheless equipped with a persistence diagram.)

Remark 5.20 (Multi-parameter persistent magnitude)**.** The theory of persistent homology captures the topology of families governed by a single parameter, but one often encounters richer structures that are governed by multiple parameters. This is the setting of multi-parameter persistence [16].

An *r-parameter persistence module* is a functor $M \colon (\mathbb{R}^r, \leq) \to \mathbf{Vect}$, where $(x_1, \ldots, x_r) \leq (y_1, \ldots, y_r)$ if $x_i \leq y_i$ for all i. Thus a persistence module is a 1-parameter persistence module in the sense of this definition.

Adapting the methodology and results of 1-parameter persistence to the multi-parameter case is a difficult and ongoing problem. For example, it is shown in [16] that there is no discrete invariant of r-parameter persistence modules analogous to the barcode or persistence diagram of a 1-parameter persistence module.

It is therefore interesting to wonder whether an r-parameter generalisation of persistent magnitude is possible. A first guess may be that the r-parameter persistent magnitude function of an r-parameter persistence module M is given by a multi-parameter version of the interpretation given in this section, so that we have a (partially defined) function from $(0, \infty)^r$ to $\mathbb{R}$ defined by

$$(t_1, \ldots, t_r) \mapsto \int_{\mathbb{R}^r} \frac{\partial^r \operatorname{rank}(M)}{\partial s_1 \cdots \partial s_r}(s_1, \ldots, s_r)e^{-(s_1 t_1 + \cdots + s_r t_r)}ds_1 \cdots ds_r \tag{6}$$

where $\operatorname{rank}(M)$ is the function $\mathbb{R}^r \to \mathbb{R}$ giving the rank of M at each point, and the partial derivative is again taken in an appropriate distributional sense.

For example, if we take a "half-open box" $[a_1, b_1) \times \cdots \times [a_r, b_r)$ in $\mathbb{R}^r$, then we may define an r-parameter persistence module $\Bbbk\big([a_1, b_1) \times \cdots \times [a_r, b_r)\big)$ whose value is $\Bbbk$ on points within the box and 0 elsewhere, and whose associated morphisms are the identity maps where possible, and the zero map otherwise. This is the r-parameter analogue of an interval module $\Bbbk[a, b)$. In this case, the rank function of our module is the product of the indicator functions of the intervals $[a_i, b_i)$, and then one may compute the integral in (6) using Fubini's theorem to show that

$$|\Bbbk\big([a_1, b_1) \times \cdots \times [a_r, b_r)\big)| = |\Bbbk[a_1, b_1)| \cdots |\Bbbk[a_r, b_r)|$$

More generally, one can take r_1- and r_2-parameter persistence modules and take their 'exterior tensor product' to obtain an $(r_1 + r_2)$-parameter persistence module, and one can show that the persistent magnitude of this exterior tensor product is the product of the persistent magnitudes of the two factors.

We would expect the definition (6) to have other good formal properties, but the challenge would be to make it applicable, in particular to ensure that it can be computed for the sort of multi-parameter persistence modules that arise in applications.

6. Magnitude and persistent homology of sublevel sets

An important class of filtrations that can be studied by methods of persistent homology are sublevel set filtrations; the study of these is to a large extent inspired by Morse theory. Recall that if $f\colon X \to \mathbb{R}$ is a continuous function, then the *sublevel set persistent homology* of (X, f) is the graded persistence module defined by $s \mapsto H_*(f^{-1}(-\infty, s])$.

Consider the case of a Morse function $f\colon M \to \mathbb{R}$ on a closed smooth manifold. Being Morse, it only has finitely many nondegenerate critical points. The magnitude of the sublevel set persistence module associated to (M, f), which we refer to as the *Morse magnitude of* (M, f) and write $|t(M, f)|_{\text{Morse}}$, has an explicit formula in terms of the critical points.

Theorem 6.1. *Let $f\colon M \to \mathbb{R}$ be a Morse function on a closed smooth manifold M, Let $S\colon (\mathbb{R}, \leq) \to \mathbf{Top}$ be the sublevel set filtration given by $S(s) = f^{-1}(-\infty, s]$. Then the magnitude function of the sublevel set persistent homology $H_*S\colon (\mathbb{R}, \leq) \to \mathbf{GrVect}$ is expressed as follows:*

$$|t(M, f)|_{\text{Morse}} = |t(H_*S)| = \sum_p (-1)^{\text{ind}(p)} e^{-f(p)t}$$

where the sum is over all critical points of f.

Proof. A basic result of Morse theory [41, Theorem 3.1] states that if $a < b$ are real numbers such that $f^{-1}(a, b]$ contains no critical points of f, then $M^b = f^{-1}(-\infty, b]$ deformation retracts onto $M^a = f^{-1}(-\infty, a]$. It follows that the critical values (i.e. the values $f(p)$ of f at the critical points p) are the startpoints and endpoints of the interval decomposition of H_*S. List the critical values as $v_1 < v_2 < \cdots < v_k$. We may now use the description of magnitude as the filtered Euler characteristic (5):

$$|t(H_*S)| = \sum_{i=1}^{k} \chi(M^{v_i})(e^{-v_i t} - e^{-v_{i+1} t}),$$

where v_{k+1} is interpreted as ∞. Another basic result of Morse theory [41, Theorem 3.2, Remark 3.3 & Remark 3.4] states the following. Suppose that b is a critical value of f, and $a < b$ is such that there are no critical values of f in (a, b), and let $p_1, \ldots, p_r$ be the critical points of f with critical value b. Then M^b has a subspace of the form $M^a \cup e^{\text{ind}(p_1)} \cup \cdots \cup e^{\text{ind}(p_r)}$, and M^b deformation retracts onto $M^a \cup e^{\text{ind}(p_1)} \cup \cdots \cup e^{\text{ind}(p_r)}$. Using this result, we then

have $\chi(M^b) = \chi(M^a) + \sum_{j=1}^{r}(-1)^{\mathrm{ind}(p_j)} = \chi(M^a) + \sum_{p:\ \mathrm{ind}(p)=b}(-1)^{\mathrm{ind}(p)}$. It follows that if v is a critical value of f, then

$$\chi(M^v) = \sum_{p:\ f(p) \leq v} (-1)^{\mathrm{ind}(p)}$$

where the sum is over critical points with critical value at most v. We now have

$$
\begin{aligned}
|t(H_*S)| &= \sum_{i=1}^{k} \chi(M^{v_i})(e^{-v_i t} - e^{-v_{i+1}t}) \\
&= \sum_{i=1}^{k} \sum_{p:\ f(p) \leq v_i} (-1)^{\mathrm{ind}(p)}(e^{-v_i t} - e^{-v_{i+1}t}) \\
&= \sum_{p} (-1)^{\mathrm{ind}(p)} \sum_{v_j:\ f(p) \leq v_j} (e^{-v_j t} - e^{-v_{j+1}t}) \\
&= \sum_{p} (-1)^{\mathrm{ind}(p)} e^{-f(p)t}
\end{aligned}
$$

$\square$

Remark 6.2. This could be generalised in a straightforward way to the case of the sublevel set filtration associated to any tame function $f\colon X \to \mathbb{R}$ on a topological space X using the concept of *homological critical value* [14, 18, 26].

Example 6.3 (Distance filtration)**.** Consider a subset $A \subseteq \mathbb{R}^n$ and filter $\mathbb{R}^n$ by $B(A, r) = \bigcup_{a \in A} B(x, r)$. This is the sublevel set filtration associated to the distance function $x \mapsto d(x, A)$. Applying singular homology H_* to this filtration yields a graded persistence module.

For example, for the standard embedding $i\colon S^{n-1} \hookrightarrow \mathbb{R}^n$ we obtain the persistence module consisting of a $\mathbb{k}[0, \infty)$ bar in degree 0 and a $\mathbb{k}[0, 1)$ bar in degree $n - 1$. In particular, the associated magnitude function, which could also be called 'the distance magnitude function' is $|tS^{n-1}|_{\mathrm{dist}} = 1 + (-1)^{(n-1)}(1 - e^{-t})$.

In the case A is finite, the graded persistence module obtained is isomorphic to the Čech persistent homology module associated to A. The corresponding magnitude function could therefore reasonably be called the 'Čech magnitude function' of A and denoted $|tA|_{\check{\mathrm{C}}\mathrm{ech}}$.

7. Rips magnitude

In this section we will apply the persistent magnitude developed earlier to the persistent homology of the Rips complex in order to obtain a new, variant form of magnitude of a finite metric space. Here we explore the basic properties of this new invariant, before going into further detail in later sections.

Definition 7.1 (Rips magnitude)**.** Let X be a finite metric space. Then the chains $C_*(\mathcal{R}(X))$ of its Rips complex are a chain complex of finitely presented persistence modules, concentrated in degrees less than the cardinality of X. The same therefore holds for the homology $H_*(\mathcal{R}(X))$. The *Rips magnitude* of X is defined to be the magnitude of the chains of the Rips complex or equivalently

the magnitude of its homology:

$$|X|_{\text{Rips}} = |C_*(\mathcal{R}(X))| = |H_*(\mathcal{R}(X))|$$

The *Rips magnitude function* of X is defined as $t \mapsto |tX|_{\text{Rips}}$, which is equal to

$$|tX|_{\text{Rips}} = |tC_*(\mathcal{R}(X))| = |tH_*(\mathcal{R}(X))|$$

Proposition 7.2. *The Rips magnitude (function) of a finite metric space X has the following properties.*

(1) *The Rips magnitude is computed by the formula:*

$$|tX|_{\text{Rips}} = \sum_{A \subseteq X, \, A \neq \emptyset} (-1)^{\#A-1} e^{-\operatorname{diam}(A)t} \tag{7}$$

(2) *If $H_*(\mathcal{R}(X))$ has barcode with bars $[a_{k,0}, b_{k,0}), \ldots, [a_{k,m_k}, b_{k,m_k})$ in degree k, then:*

$$|tX|_{\text{Rips}} = \sum_{k=0}^{\#X-1} \sum_{j=0}^{m_k} (-1)^k \left(e^{-a_{k,j}t} - e^{-b_{k,j}t} \right) \tag{8}$$

(3) *If $0 = d_0 < d_1 < \ldots < d_n$ is the set of all pairwise distances between elements of X arranged in a sequence and $d_{n+1} = \infty$, then:*

$$|tX|_{\text{Rips}} = \sum_{j=0}^{n} \chi(\mathcal{R}_{d_j}(X))(e^{-d_j t} - e^{-d_{j+1}t}). \tag{9}$$

(4) $\lim_{t \to 0} |tX|_{\text{Rips}} = 1$ *and* $\lim_{t \to \infty} |tX|_{\text{Rips}} = |X|$.

The fourth part of the proposition suggests that the Rips magnitude is an 'effective number of points', in the same spirit as the magnitude.

Proof. For the first part, we use the description $|tX|_{\text{Rips}} = |tC_*(\mathcal{R}(X))|$. Now $C_*(\mathcal{R}(X))$ has barcode with one bar for each nonempty subset A of X, and this bar lies in degree $\#A - 1$, and has type $[\operatorname{diam}(A), \infty)$. The definition of persistent magnitude gives the result immediately. The second part follows from the description $|tX|_{\text{Rips}} = |tH_*(\mathcal{R}(X))|$. The third part follows from either the first or the second part using the formula (5). The fourth part follows directly from the first, and it can also be deduced from the barcode description given there using Proposition 5.10. $\square$

Example 7.3 (Rips magnitude of the one-point space)**.** Let X denote the space consisting of a single point x. Then it has precisely one nonempty subset, namely X itself, and $\#X = 1$ while $\operatorname{diam}(X) = 0$. Using formula (8) then gives us $|tX|_{\text{Rips}} = 1$.

Example 7.4 (Rips magnitude of two-point spaces)**.** Let $X = \{x_1, x_2\}$ be the two-point space in which $d_X(x_1, x_2) = d$ for some $d > 0$.

Let us compute $|tX|_{\text{Rips}}$ using (7). The nonempty subsets of X are $A_1 = \{x_1\}$, $A_2 = \{x_2\}$ and $A_3 = X$, with $\#A_1 = 1$, $\#A_2 = 1$, $\#A_3 = 2$, $\operatorname{diam}(A_1) = 0$,

$\mathrm{diam}(A_2) = 0$ and $\mathrm{diam}(A_3) = d$. Thus we have

$$|tX|_{\mathrm{Rips}} = (-1)^{1-1} \cdot e^{-0t} + (-1)^{1-1} \cdot e^{-0t} + (-1)^{2-1} \cdot e^{-dt}$$
$$= 1 + 1 - e^{-dt}$$
$$= 2 - e^{-dt}.$$

Let us also compute $|tX|_{\mathrm{Rips}}$ using (8). The Rips-homology $H_*(\mathcal{R}(X))$ has barcode with bars $[0, \infty)$ and $[0, d)$ in degree 0, and no other bars. Thus (8) gives us

$$|tX|_{\mathrm{Rips}} = (e^{-0t} - e^{-\infty t}) + (e^{-0t} - e^{-dt})$$
$$= (1 - 0) + (1 - e^{-dt})$$
$$= 2 - e^{-dt}.$$

(Recall our convention that $e^{-\infty} = 0$.)

Example 7.5. The Rips magnitude is not necessarily increasing or convex, it can attain negative values, and it can attain values greater than the cardinality of the space. For instance, for the complete bipartite graph $K_{5,6}$ (with the graph metric) the Rips magnitude is given by

$$|tK_{5,6}|_{\mathrm{Rips}} = 11 - 30e^{-t} + 20e^{-2t}.$$

with graph:
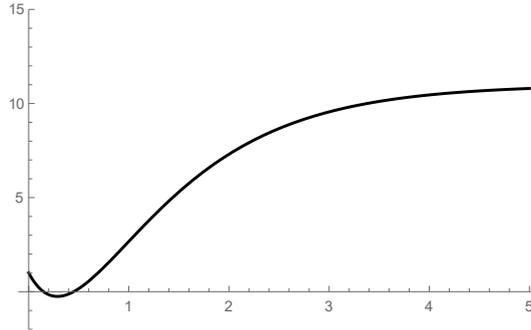
And for the complete tripartite graph $X = K_{4,4,4}$, Rips magnitude is given by

$$|tK_{4,4,4}|_{\mathrm{Rips}} = 12 + 16e^{-t} - 27e^{-2t}$$

with graph:
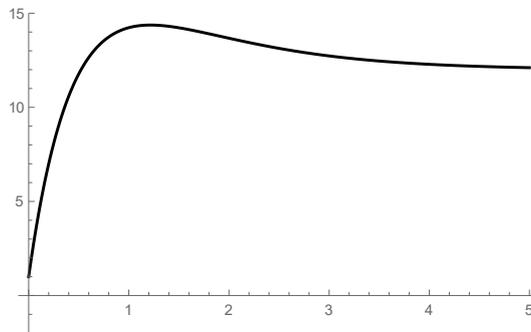
Note that in general, if the metric only assumes integer values, as in the case of a graph metric, the associated Rips magnitude function is a polynomial in $q = e^{-t}$.

Now we consider the case of subsets of the real line, where the computation is a little less trivial but accessible nonetheless.

Proposition 7.6. *Let A be a finite subset of the real line $\mathbb{R}$, with its induced metric. Order the elements of A by size, $a_1 < \ldots < a_n$. Then*

$$|tA|_{\text{Rips}} = n - \sum_{j=1}^{n-1} e^{-(a_{j+1}-a_j)t}.$$

We can think of this result as follows. Take $A_1 = \{a_1\}$, $A_2 = \{a_1, a_2\}$, $A_3 = \{a_1, a_2, a_3\}$ and so on, so that $A = A_n$. Then the proposition tells us that $|tA_1|_{\text{Rips}} = 1$, $|tA_2|_{\text{Rips}} = |tA_1|_{\text{Rips}} + (1 - e^{-(a_2-a_1)t})$, $|tA_3|_{\text{Rips}} = |tA_2|_{\text{Rips}} + (1 - e^{-(a_3-a_2)t})$, and so on. In other words, adding a point at the end increases the Rips magnitude by $1 - e^{-dt}$, where d is the distance of the new end point from the old one. So if d is very large, we increase the Rips magnitude by almost 1, whereas if d is very small, then we increase the Rips magnitude only a tiny amount.

Proof. Given $B \subseteq A$, let $B_{\max}$ and $B_{\min}$ denote the maximum and minimum elements of B respectively. Given $a \le a'$ in A, let $\mathcal{B}_{a,a'}$ denote the set of $B \subseteq A$ for which $B_{\min} = a$ and $B_{\max} = a'$, and note that $\text{diam}(B) = a' - a$ for all $B \in \mathcal{B}_{a,a'}$. Thus equation (7) gives us

$$|tA|_{\text{Rips}} = \sum_{a \le a'} e^{-(a'-a)t} \sum_{B \in \mathcal{B}_{a,a'}} (-1)^{\#B-1}.$$

Now note the following:

- If $a = a'$, then $\mathcal{B}_{a,a'}$ consists of $\{a\}$ alone and $\sum_{B \in \mathcal{B}_{a,a'}} (-1)^{\#B-1} = 1$.
- If a and a' are adjacent elements of A, then $\mathcal{B}_{a,a'}$ consists of $\{a, a'\}$ alone and $\sum_{B \in \mathcal{B}_{a,a'}} (-1)^{\#B-1} = -1$.
- If a and a' are non-adjacent elements of A, then let $A_{a,a'}$ denote the set of those elements of A that lie strictly between a and a'. Then any $B \in \mathcal{B}_{a,a'}$ is the disjoint union of $\{a, a'\}$ with a subset $C \subseteq A_{a,a'}$. Thus $\sum_{B \in \mathcal{B}_{a,a'}} (-1)^{\#B-1} = \sum_{C \subseteq A_{a,a'}} (-1)^{\#C+1} = -\sum_{C \subseteq A_{a,a'}} (-1)^{\#C} = 0$.

We therefore have

$$|tA|_{\text{Rips}} = \sum_{a} e^{-(a-a)t} \cdot 1 + \sum_{\substack{a < a' \\ \text{adjacent}}} e^{-(a'-a)t} \cdot (-1) + \sum_{\substack{a < a' \\ \text{non-adjacent}}} e^{-(a'-a)t} \cdot 0$$

$$= n - \sum_{j=1}^{n-1} e^{-(a_{j+1}-a_j)t}$$

as required. $\qquad\square$

8. Rips magnitude of cycle graphs and Euclidean cycles

Other than the cases treated in the previous section, the topology of Rips complexes seems to be understood at all scales only in the case of finite subsets of the circle. See [2] and [3]. In the case of Riemannian manifolds, Rips complexes are well understood at small scales thanks to a result of Hausmann [29,

Theorem 3.5]. For ellipses, the Rips complex has been studied at a range of length scales in [4]. Recently, it has also been shown that Rips complexes can be understood as nerves of certain covers via Dowker duality [45].

In this section we focus on the circle S^1, which is equipped either with the Euclidean metric obtained from the standard embedding into $\mathbb{R}^2$, and denoted S^1_{eucl}, or the geodesic (arclength) metric of total length 2π, and denoted S^1_{geo}.

We will examine the subsets of equally spaced points in these spaces, whose corresponding Rips filtrations are well understood [1]. Let C_n^{eucl} be the subset of n equidistant points in S^1_{eucl} and let C_n^{geo} be the subset of n equidistant points in S^1_{geo}.

Both of these can be related to cycle graphs. Let C_n be the set of vertices of the n-cycle graph, equipped with the graph metric, where two adjacent vertices are considered to be at a distance of 1. Note that this can be described as the subset of n equidistant points in a geodesic circle of total arclength n. The Rips filtration of C_n was studied in [1] and the Rips filtrations of C_n^{geo} and C_n^{eucl} are just reparameterised versions of it.

More precisely, extend the function $[0, 2] \to [0, \pi]$ given by $r \mapsto 2 \arcsin \frac{r}{2}$ to a homeomorphism $\phi \colon \mathbb{R} \to \mathbb{R}$. Then, by elementary trigonometry, we have the following relations:

$$\mathcal{R}_r(C_n^{\text{geo}}) = \mathcal{R}_{\frac{n}{2\pi}r}(C_n), \qquad \mathcal{R}_r(C_n^{\text{eucl}}) = \mathcal{R}_{\frac{n}{2\pi}\phi(r)}(C_n) \tag{10}$$

and

$$\mathcal{R}_r(C_n^{\text{eucl}}) = \mathcal{R}_{\phi(r)}(C_n^{\text{geo}}). \tag{11}$$

We now state the main results of this section and then proceed to prove them.

Proposition 8.1. *Writing $q = e^{-t}$, we have:*

$$|tC_n|_{\text{Rips}} = \sum_{\substack{odd\ r|n \\ r \neq n}} \frac{n}{r} q^{\frac{n}{r} \frac{r-1}{2}} (1 - q) + q^{\lfloor \frac{n}{2} \rfloor}.$$

This is reminiscent of certain functions that appear in analytic number theory, the simplest of which is probably the sum of divisors function $\sigma_k(n) = \sum_{d|n} d^k$. The appearance of sums over divisors of integers is quite surprising to us and seems to suggest that the Rips magnitudes of n-cycle graphs might be intimately connected to number theory in some way. We feel this connection could be worthy of further study:

Question 8.2. What is the connection between Rips magnitude of cycles and various functions studied in analytic number theory?

Using Lemma 5.12 and equations (10) and (11), Proposition 8.1 immediately allows us to infer the following two corollaries:

Corollary 8.3. *The Rips magnitude of Euclidean cycles is given by[1]:*

$$|tC_n^{\text{eucl}}|_{\text{Rips}} = \sum_{\substack{odd\ r|n \\ r \neq n}} \frac{n}{r} (e^{-\delta_r t} - e^{-\delta_{r,n} t}) + e^{-\delta_n t},$$

---

[1]In fact, $\delta_n = \delta_{n,n}$ for odd n, so the condition $r \neq n$ can be omitted for Euclidean cycles.

where

$$\delta_r = \operatorname{diam}(C_r^{\text{eucl}}) = 2\sin\left(\pi\frac{\lfloor\frac{r}{2}\rfloor}{r}\right) \quad and \quad \delta_{r,n} = 2\sin\left(\pi\left(\frac{1}{n} + \frac{\lfloor\frac{r}{2}\rfloor}{r}\right)\right).$$

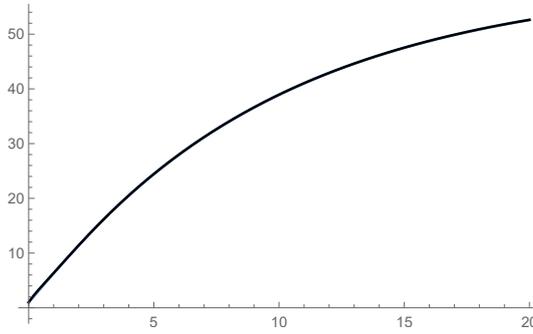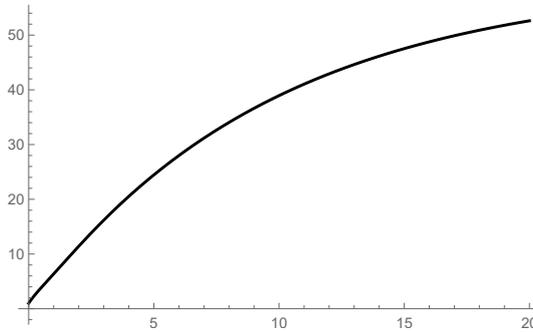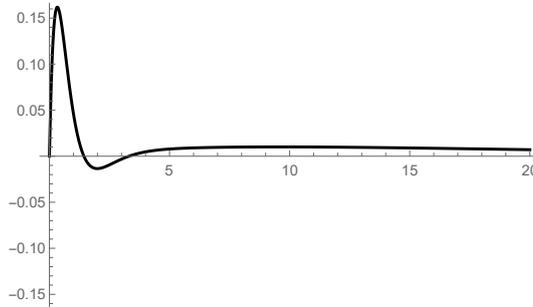**Corollary 8.4.** *The Rips magnitude of geodesic cycles is given by:*

$$|tC_n^{\text{geo}}|_{\text{Rips}} = \sum_{\substack{odd\ r|n \\ r\neq n}} \frac{n}{r}(e^{-\eta_r t} - e^{-\eta_{r,n} t}) + e^{-\eta_n t},$$

where

$$\eta_r = 2\pi\frac{\lfloor\frac{r}{2}\rfloor}{r} \qquad and \qquad \eta_{r,n} = 2\pi\left(\frac{1}{n} + \frac{\lfloor\frac{r}{2}\rfloor}{r}\right).$$

Example 8.5. The graph of $|tC_{60}^{\text{eucl}}|_{\text{Rips}}$

looks deceptively similar to the one for $|tC_{60}^{\text{geo}}|_{\text{Rips}}$

so it is perhaps more instructive to look at the difference $|tC_{60}^{\text{geo}}|_{\text{Rips}} - |tC_{60}^{\text{eucl}}|_{\text{Rips}}$:

Proof of Proposition 8.1. Note that the Rips filtration of C_n only has "jumps" at the integers. More precisely:

$$\mathcal{R}_r(C_n) = \mathcal{R}_{\lfloor r\rfloor}(C_n).$$

So it is sufficient to understand the Rips filtration at integer values of the filtration parameter r. For integer r such that $0 \leq r < \frac{n}{2}$, Adamaszek [1, Corollary 6.7] gives the following description of the homotopy types of the various stages of the Rips filtration:

$$\mathcal{R}_r(C_n) \simeq \begin{cases} \bigvee_{n-2r-1} S^{2l}; & r = \frac{l}{2l+1}n, \\ S^{2l+1}; & \frac{l}{2l+1}n < r < \frac{l+1}{2l+3}n. \end{cases}$$

From this we can immediately infer the Euler characteristics:

$$\chi(\mathcal{R}_r(C_n)) = \begin{cases} n - 2r; & \text{if } \frac{n}{n-2r} \text{ is an odd integer}, \\ 1; & \text{if } n = 2r, \\ 0; & \text{otherwise}. \end{cases}$$

Using (9), this implies

$$|tC_n|_{\mathrm{Rips}} = \sum_{r=0}^{\lfloor \frac{n}{2} \rfloor - 1} \chi(\mathcal{R}_r(X))(e^{-rt} - e^{-(r+1)t}) + \chi(\mathcal{R}_{\lfloor \frac{n}{2} \rfloor}(X))e^{-(\lfloor \frac{n}{2} \rfloor)t})$$

$$= \sum_{\substack{\text{odd } d|n \\ d \neq n}} \frac{n}{d} q^{\frac{n}{d} \frac{d-1}{2}} (1-q) + q^{\lfloor \frac{n}{2} \rfloor}$$

where the indices in the first and second summations are related by $\frac{n}{n-2r} = d$. The claim follows. $\square$

Remark 8.6. Another way to calculate the Euler characteristic of $\mathcal{R}_r(C_n)$ would be from the simplex counts. We have computational evidence that the number of i-simplices in $\mathcal{R}_r(C_n)$ for $r < \operatorname{diam} C_n$ is given by[2]

$$N_{n,r,i} = \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \frac{n}{2k+1} \binom{(2k+1)r - kn + 2k}{2k} \binom{(2k+1)r - kn}{i - 2k}.$$

Remark 8.7. Computational evidence seems to suggest that the Rips magnitude $|tC_n|_{\mathrm{Rips}}$ of a cycle graph is convex if and only if

$$n \in \{1, 2, 3, 4, 5, 6, 9, 10, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 48, 51, 57\}.$$

9. Rips magnitude of infinite metric spaces

We will now try to understand the Rips magnitude of infinite metric spaces. In the original setting of magnitude, positive definite spaces X, such as subspaces of Euclidean space, have the property that if $A \subseteq B \subseteq X$ with A, B finite, then $|A| \leq |B|$. Thus one definition of the magnitude of an infinite metric space X is

$$|X| = \sup_{\substack{A \subseteq X \\ A \text{ finite}}} |A|.$$

Proceeding by analogy, we could attempt to make the definition

$$|X|_{\mathrm{Rips}} = \sup_{\substack{A \subseteq X \\ A \text{ finite}}} |A|_{\mathrm{Rips}}$$

---

[2]Here, we use the **nonstandard (!)** convention that $\binom{n}{k} = 0$ for $n < 0$.

or the stronger definition

$$|X|_{\text{Rips}} = \lim_{\substack{A \to X \\ A \text{ finite}}} |A|_{\text{Rips}},$$

where $A \to X$ in the Hausdorff metric. The issues here are whether the supremum above exists, how one computes its actual values, and the corresponding questions for the limit.

In this section we will explore the above situation in the case of the unit interval, the circle with its Euclidean metric, and the circle with its geodesic metric. In the case of the interval the situation is as good as one can hope for, with the conclusion that the Rips magnitude of the interval 'is' the function $t \mapsto 1 + t$. In the circle cases the situation is more ambiguous, and while the supremum above may well exist (the supremum taken over all *equally spaced* subsets certainly does), the limit does not.

To improve readability, we state the results for each of the cases treated in a separate subsection, while deferring all the proofs to one final subsection.

9.1. **The unit interval.** First let us consider the unit interval. Here we can give the following definitive description of the situation.

Theorem 9.1. *Let $I = [0,1]$ denote the unit interval.*
  (1) *If $A, B \subseteq I$ are finite subsets with $A \subseteq B$, then*

$$|A|_{\text{Rips}} \le |B|_{\text{Rips}}.$$

  (2) *For any $t \in (0, \infty)$ we have*

$$\sup_{\substack{A \subseteq I \\ \text{finite}}} |tA|_{\text{Rips}} = 1 + t$$

  and indeed

$$\lim_{\substack{A \to I \\ A \text{ finite}}} |tA|_{\text{Rips}} = 1 + t$$

  with uniform convergence on compact subsets of $(0, \infty)$.

Thus it seems reasonable in this situation to declare the Rips magnitude of $I = [0,1]$ to be the function $|t[0,1]|_{\text{Rips}} = 1 + t$.

9.2. **Euclidean Circle.** One possible approach to try and make sense of the notion of "Rips magnitude of the Euclidean circle" is by studying the behaviour of $|tC_n^{\text{eucl}}|_{\text{Rips}}$ as $n \to \infty$. As it turns out, however, this behaviour is not as straightforward as one might hope.

For instance, we show that $|tC_n^{\text{eucl}}|_{\text{Rips}}$ does not converge as $n \to \infty$, despite the fact that the Hausdorff distance $d_H(C_n^{\text{eucl}}, S_{\text{eucl}}^1) = 2\sin\left(\frac{\pi}{2n}\right)$ converges to 0 as $n \to \infty$. We show this by first studying the behaviour of sequences of the form $|tC_{mp}^{\text{eucl}}|_{\text{Rips}}$, for fixed $m \in \mathbb{N}$ and where p runs through all primes, and showing that the limit along each such subsequence exists. However, these limits are different for different values of m.

We then show that despite this inconsistency, the $|tC_n^{\text{eucl}}|_{\text{Rips}}$ has finite upper and lower limits ($\liminf$ and $\limsup$) as $n \to \infty$ which can be expressed explicitly. These could be considered to be the 'upper' and 'lower Rips magnitude'. We also show that the 'upper Rips magnitude' is equal to $\sup_{n \in \mathbb{N}} |tC_n^{\text{eucl}}|_{\text{Rips}}$.

(Compare this with the first proposed definition of Rips magnitude at the beginning of this section.)

Theorem 9.2. *For any $m \in \mathbb{N}$:*

$$\lim_{\substack{p \to \infty \\ p \ prime}} |tC_{mp}^{\mathrm{eucl}}|_{\mathrm{Rips}} = e^{-2t} + 2\pi t \sum_{odd \ r|m} \frac{1}{r} e^{-2t\cos\left(\frac{\pi}{2r}\right)} \sin\left(\frac{\pi}{2r}\right)$$

This result means in particular that

$$\lim_{n \to \infty} |tC_n^{\mathrm{eucl}}|_{\mathrm{Rips}}$$

cannot exist. For instance, we have

$$\lim_{\substack{p \to \infty \\ p \ prime}} |tC_p^{\mathrm{eucl}}|_{\mathrm{Rips}} = e^{-2t} + 2\pi t$$

but

$$\lim_{\substack{p \to \infty \\ p \ prime}} |tC_{3p}^{\mathrm{eucl}}|_{\mathrm{Rips}} = e^{-2t} + 2\pi t + \frac{\pi t}{3} e^{-\sqrt{3}t}$$

and the two limits do not coincide.

Theorem 9.3. *The Rips magnitudes of Euclidean cycles C_n^{eucl} satisfy*

$$\liminf_{n \to \infty} |tC_n^{\mathrm{eucl}}|_{\mathrm{Rips}} = e^{-2t} + 2\pi t$$

and

$$\limsup_{n \to \infty} |tC_n^{\mathrm{eucl}}|_{\mathrm{Rips}} = e^{-2t} + 2\pi t \sum_{r \ odd} \frac{1}{r} e^{-2t\cos\left(\frac{\pi}{2r}\right)} \sin\left(\frac{\pi}{2r}\right).$$
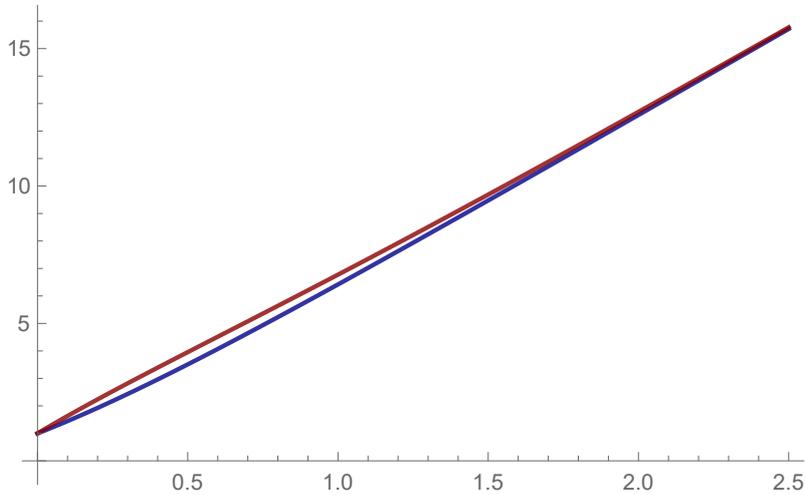
This series converges absolutely since its r-th term is bounded above by $\frac{\pi}{2r^2}$.

In fact, the upper limit is also the supremum of the sequence:

Theorem 9.4. *The Rips magnitudes of Euclidean cycles C_n^{eucl} satisfy*

$$\sup_{n \in \mathbb{N}} |tC_n^{\mathrm{eucl}}|_{\mathrm{Rips}} = e^{-2t} + 2\pi t \sum_{r \ odd} \frac{1}{r} e^{-2t\cos\left(\frac{\pi}{2r}\right)} \sin\left(\frac{\pi}{2r}\right).$$

We include a plot of the $\liminf$ and $\limsup$ for small t for comparison:

It is visible from this graph that the difference is the most pronounced for small values of t. Now, fix $t = \frac{1}{2}$. The behaviour of $|tC_n^{\text{eucl}}|_{\text{Rips}}$ evaluated at $t = \frac{1}{2}$ as n grows larger can be pictured as follows, with n on the horizontal axis and $|tC_n^{\text{eucl}}|_{\text{Rips}}$ on the vertical. The values of the lim inf and lim sup at $t = \frac{1}{2}$ are plotted as the two red lines.
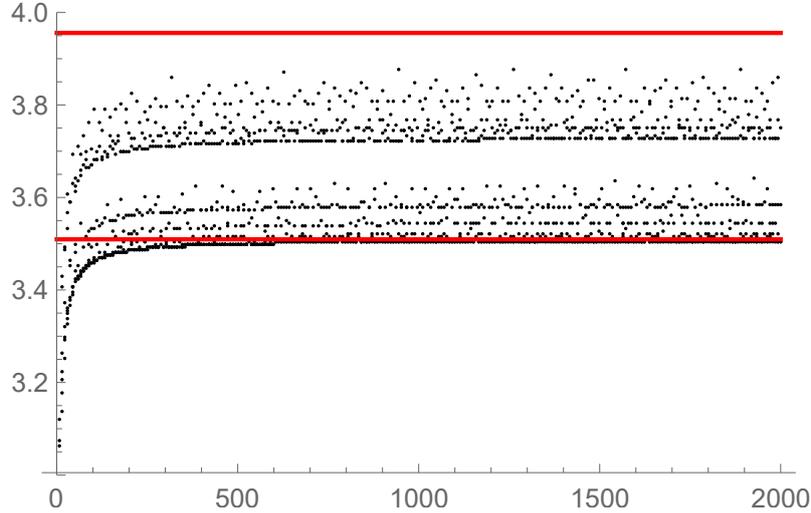
Note that the chaotic behaviour of the graph is reminiscent of various functions from analytic number theory such as for instance the sum of divisors function $\sigma(n) = \sum_{d|n} d$, and is the graphical expression of the behaviour discussed in the paragraph following Proposition 8.1. It does appear as though the points accumulate more along specific lines, which we suspect correspond to subsequences with certain divisibility properties.

Finally, note that restricting to equally spaced subsets of S^1_{eucl} is somewhat unnatural; we leave open the following question, which would define the 'upper and lower Rips magnitude' of the circle intrinsically:

Question 9.5. Does this asymptotic behaviour extend to arbitrary finite subsets of S^1_{eucl}? For instance, given any $\epsilon > 0$, is there a $\delta > 0$ such that for all finite $A \subseteq S^1_{eucl}$ with $d_H(A, S^1_{eucl}) < \delta$ we have

$$e^{-2t} + 2\pi t - \epsilon < |tA|_{\text{Rips}} < e^{-2t} + 2\pi t \sum_{r \text{ odd}} \frac{1}{r} e^{-2t \cos\left(\frac{\pi}{2r}\right)} \sin\left(\frac{\pi}{2r}\right) + \epsilon$$

for all t in a given interval?

9.3. **Geodesic circle.** Finally, we note that the case of the geodesic circle S^1_{geo} of total arclength 2π can be treated using the same methods as we used for S^1_{eucl}. Namely, we restrict attention to equally spaced subsets C_n^{geo} described in Section 8. (Note that C_n^{geo} is just the n-cycle graph C_n rescaled by $\frac{2\pi}{n}$.)

We could calculate the limits along the same subsequences we examined in the case S^1_{eucl} and find that they again exist, but instead we just state the final result regarding the lower and upper limit. In this case it turns out that the lower limit is still finite and can be expressed explicitly, whereas the upper limit becomes infinite. Thus the Rips magnitude of S^1_{eucl} and S^1_{geo} behave quite differently.
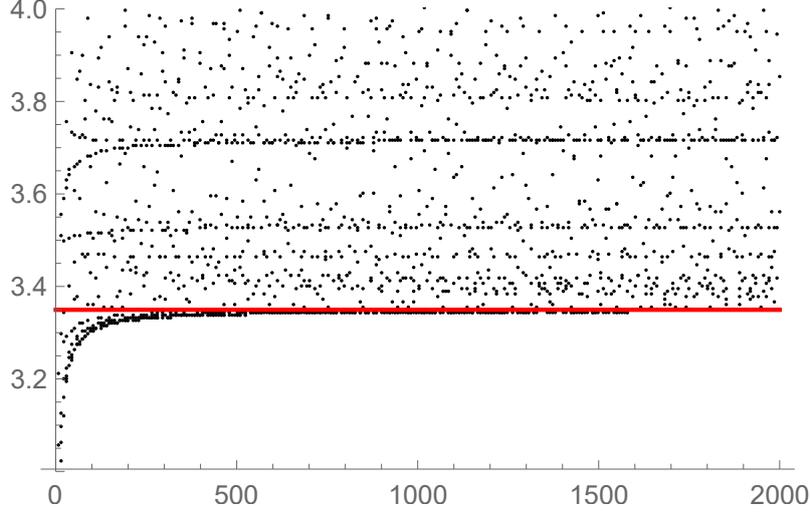
Theorem 9.6. *The Rips magnitudes of geodesic cycles C_n^{geo} satisfy*

$$\liminf_{n \to \infty} |tC_n^{\text{geo}}|_{\text{Rips}} = e^{-\pi t} + 2\pi t$$

and

$$\limsup_{n \to \infty} |tC_n^{\text{geo}}|_{\text{Rips}} = \infty.$$

The behaviour of $|tC_n^{\text{geo}}|_{\text{Rips}}$ evaluated at $t = \frac{1}{2}$ as n grows larger can be pictured as follows; the $\liminf$ evaluated at $t = \frac{1}{2}$ corresponds to the red line and the $\limsup$ is infinite.

We again note that it would be interesting to study the asymptotics over all finite subsets.

9.4. **Proofs of the Results.** Here we prove the results stated in the preceding subsections. We only give a sketch of the proof of Theorem 9.6 as the ideas are analogous to the ones used in the proof of Theorem 9.3.

Proof of Theorem 9.1 part (1). Proposition 7.6 shows that if $A \subseteq [0, 1]$ is finite, with elements $a_1 < \cdots < a_n$, then

$$|A|_{\text{Rips}} = n - \sum_{j=1}^{n-1} e^{-(a_{j+1} - a_j)}.$$

Suppose that $a_0 < a_1$. Then

$$|\{a_0\} \cup A|_{\text{Rips}} = |A|_{\text{Rips}} + (1 - e^{-(a_1 - a_0)}) \geq |A|_{\text{Rips}},$$

and similarly for $|A \cup \{a_{n+1}\}|_{\text{Rips}}$ if $a_{n+1} > a_n$. Now suppose that $a_i < b < a_{i+1}$. Then

$$|A \cup \{b\}|_{\text{Rips}} = |A|_{\text{Rips}} + 1 + e^{-(a_{i+1} - a_i)} - e^{-(a_{i+1} - b)} - e^{-(b - a_i)} \geq |A|_{\text{Rips}},$$

the latter because one can easily see that $1 + e^{-(x+y)} - e^{-x} - e^{-y} \geq 0$ for $x, y \geq 0$. $\qquad\square$

Proof of Theorem 9.1 part (2). Suppose A consisting of $a_1 < \ldots < a_n$ is a finite subset of the interval such that $d_H(A, I) < \delta < 1$. Further define $a_0 = 0$ and $a_{n+1} = 1$. The assumption on the Hausdorff distance implies that $a_{j+1} - a_j < \delta$

for $j = 0, \ldots, n$ and $\delta < 1$ implies $\delta^k \leq \delta$ for $k \in \mathbb{N}$. Consider the function $f \colon (0, \infty) \to \mathbb{R}$ defined by

$$f(t) = (1 + t) - |tA|_{\text{Rips}}$$

$$= (1 + t) - n + \sum_{j=1}^{n-1} e^{-(a_{j+1} - a_j)t}$$

$$= t + \sum_{j=1}^{n-1} (e^{-(a_{j+1} - a_j)t} - 1).$$

We extend the domain of f to $[0, \infty)$ in the evident way. We will show that:

(1) $f(0) = 0$
(2) $f'(0)$ lies in the range $0 \leq f'(0) \leq 2\delta$.
(3) $f''(t)$ lies in the range $0 \leq f''(t) \leq \delta$ for all $t \in [0, \infty)$.

It follows quickly that

$$0 \leq f(t) \leq \delta \cdot (2t + t^2/2)$$

for $t \in [0, \infty)$. Thus, as $\delta \to 0$ we have $f \to 0$ uniformly on any bounded subset of $[0, \infty)$, and the result follows.

It remains to check the three properties. The first and second derivatives of f are as follows.

$$f'(t) = 1 - \sum_{j=1}^{n-1} (a_{j+1} - a_j) e^{-(a_{j+1} - a_j)t}$$

$$f''(t) = \sum_{j=1}^{n-1} (a_{j+1} - a_j)^2 e^{-(a_{j+1} - a_j)t}$$

Then (1) is immediate, while $f'(0) = 1 - \sum_{j=1}^{n-1}(a_{j+1} - a_j) = (1 - a_n) + a_1$ and (2) follows, and $0 \leq f''(t) \leq \sum_{j=1}^{n-1} \delta \cdot (a_{j+1} - a_j) \cdot 1$ and (3) follows. $\square$

Proof of Theorem 9.2. Recall that by Corollary 8.3, the Rips magnitudes of Euclidean cycles are given by:

$$|tC_n^{\text{eucl}}|_{\text{Rips}} = \sum_{\text{odd } r | n} \frac{n}{r} (e^{-\delta_r t} - e^{-\delta_{r,n} t}) + e^{-\delta_n t},$$

where

$$\delta_r = \text{diam}(C_r^{\text{eucl}}) = 2 \sin\left(\pi \frac{\lfloor \frac{r}{2} \rfloor}{r}\right) \quad \text{and} \quad \delta_{r,n} = 2 \sin\left(\pi \left(\frac{1}{n} + \frac{\lfloor \frac{r}{2} \rfloor}{r}\right)\right).$$

For an odd prime p, each odd divisor of mp is of the form r or rp (or both), where r is an odd divisor of m. Therefore, assuming p is large enough, so that $p \nmid m$, we can split the Rips magnitude into three summands:

$$|tC_{mp}^{\text{eucl}}|_{\text{Rips}} = \sum_{\text{odd } r | m} \frac{mp}{r} (e^{-\delta_r t} - e^{-\delta_{r,mp} t}) + \sum_{\text{odd } r | m} \frac{m}{r} (e^{-\delta_{rp} t} - e^{-\delta_{rp,mp} t}) + e^{-\delta_{mp} t}.$$

We can calculate the limit as $p \to \infty$ of each summand individually. To treat the first summand, define a function $\phi \colon \mathbb{R} \to \mathbb{R}$ by the formula

$$\phi(x) = 2\sin\left(\pi\left(x + \frac{r-1}{2r}\right)\right),$$

with derivative

$$\phi'(x) = 2\pi\cos\left(\pi\left(x + \frac{r-1}{2r}\right)\right).$$

Observe that $\delta_r = \phi(0)$ and $\delta_{r,mp} = \phi(\frac{1}{mp})$. Therefore,

$$
\begin{aligned}
\lim_{p\to\infty} \frac{mp}{r}\left(e^{-\delta_r t} - e^{-\delta_{r,mp}t}\right) &= -\frac{1}{r}\lim_{p\to\infty} \frac{e^{-\phi(\frac{1}{mp})t} - e^{-\phi(0)t}}{\frac{1}{mp}} \\
&= -\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}x}\Big|_{x=0} e^{-\phi(x)t} \\
&= \frac{1}{r}e^{-\phi(0)t}\phi'(0)t \\
&= \frac{2\pi t}{r}e^{-2t\sin\left(\pi\frac{r-1}{2r}\right)}\cos\left(\pi\frac{r-1}{2r}\right) \\
&= \frac{2\pi t}{r}e^{-2t\cos\left(\frac{\pi}{2r}\right)}\sin\left(\frac{\pi}{2r}\right).
\end{aligned}
$$

This takes care of the first summand. The second summand vanishes in the limit, because

$$\lim_{p\to\infty} e^{-\delta_{rp}t} = \lim_{p\to\infty} e^{-\delta_{rp,mp}t} = e^{-2t}.$$

Finally, the limit of the third summand is

$$\lim_{p\to\infty} e^{-\delta_{mp}t} = e^{-2t}.$$

$\square$

Proof of Theorem 9.3. Define $\phi_{n,r}(t) = \frac{n}{r}\left(e^{-\delta_r t} - e^{-\delta_{r,n}t}\right)$ for $n \geq r$ and 0 otherwise. Here r will always be assumed to be odd. We can show that for $t > 0$,

$$\phi_{n,r}(t) \geq 0 \qquad \text{and} \qquad \phi_{n+1,r}(t) \geq \phi_{n,r}(t)$$

always hold. The first of these inequalities reduces to the fact that $\delta_r \leq \delta_{r,n}$. For the second inequality, define $\psi(t) = 1 + ne^{(\delta_r - \delta_{r,n})t} - (n+1)e^{(\delta_r - \delta_{r,n+1})t}$ and observe that the inequality reduces to $\psi(t) \geq 0$ for $t > 0$. To show that this is in fact true, one can then verify that:

- $\psi(0) = 0$,
- $\lim_{t\to\infty} \psi(t) > 0$,
- $\psi'(0) = (n+1)\delta_{r,n+1} - n\delta_{r,n} - \delta_r > 0$ and
- $\psi'(t) = 0$ for at most one $t \in (0, \infty)$.

To prove the third bullet point, we can write out the expression explicitly and use the addition theorem for the sine function:

$$2(n+1)\sin\left(\pi\left(\frac{1}{n+1}+\frac{r-1}{2r}\right)\right)-2n\sin\left(\pi\left(\frac{1}{n}+\frac{r-1}{2r}\right)\right)-2\sin\left(\pi\frac{r-1}{2r}\right)$$

$$=2\left[(n+1)\sin\left(\frac{\pi}{n+1}\right)-n\sin\left(\frac{\pi}{n}\right)\right]\cos\left(\pi\frac{r-1}{2r}\right)$$

$$+2\left[(n+1)\cos\left(\frac{\pi}{n+1}\right)-n\cos\left(\frac{\pi}{n}\right)-1\right]\sin\left(\pi\frac{r-1}{2r}\right).$$

One can now show that the expressions in square brackets are positive and conclude that the whole expression is positive. This proves the second inequality.

Using the explicit formula for the Rips magnitude, we now have:

$$|tC_n^{\text{eucl}}|_{\text{Rips}}=\sum_{\text{odd }r|n}\frac{n}{r}(e^{-\delta_r t}-e^{-\delta_{r,n}t})+e^{-\delta_n t}\leq\sum_{\text{odd }r\leq n}\frac{n}{r}(e^{-\delta_r t}-e^{-\delta_{r,n}t})+e^{-\delta_n t}.$$

Therefore,

$$\limsup_{n\to\infty}|tC_n^{\text{eucl}}|_{\text{Rips}}\leq\lim_{n\to\infty}\left(\sum_{\text{odd }r}\phi_{n,r}(t)+e^{-\delta_n t}\right)$$

$$=\sum_{\text{odd }r}\lim_{n\to\infty}\phi_{n,r}(t)+\lim_{n\to\infty}e^{-\delta_n t}$$

$$=\sum_{\text{odd }r}\lim_{n\to\infty}\frac{n}{r}(e^{-\delta_r t}-e^{-\delta_{r,n}t})+e^{-2t}$$

$$=\sum_{\text{odd }r}\frac{2\pi t}{r}e^{-2t\cos\left(\frac{\pi}{2r}\right)}\sin\left(\frac{\pi}{2r}\right)+e^{-2t},$$

where the limit is calculated in the same way as in the proof of the previous theorem. The interchange of sum and limit is justified by the Lebesgue monotone convergence theorem. To prove the reverse inequality, let N be an arbitrary positive integer. Let $m=N!$ and note that m is divisible by every odd $r\leq N$. Therefore, by the previous theorem,

$$e^{-2t}+2\pi t\sum_{\text{odd }r\leq N}\frac{1}{r}e^{-2t\cos\left(\frac{\pi}{2r}\right)}\sin\left(\frac{\pi}{2r}\right)\leq e^{-2t}+2\pi t\sum_{\text{odd }r|m}\frac{1}{r}e^{-2t\cos\left(\frac{\pi}{2r}\right)}\sin\left(\frac{\pi}{2r}\right)$$

$$=\lim_{\substack{p\to\infty\\p\text{ prime}}}|tC_{mp}^{\text{eucl}}|_{\text{Rips}}$$

$$\leq\limsup_{n\to\infty}|tC_n^{\text{eucl}}|_{\text{Rips}}.$$

Taking the limit as $N\to\infty$ establishes the lower bound.

To prove the statement about the lower limit, again start from the explicit formula and note that $r=1$ is a proper odd divisor of any integer $n>1$:

$$|tC_n^{\text{eucl}}|_{\text{Rips}}=\sum_{\text{odd }r|n}\frac{n}{r}(e^{-\delta_r t}-e^{-\delta_{r,n}t})+e^{-\delta_n t}\geq n(e^{-\delta_1 t}-e^{-\delta_{1,n}t})+e^{-\delta_n t}.$$

Therefore

$$\liminf_{n\to\infty}|tC_n^{\text{eucl}}|_{\text{Rips}}\geq\lim_{n\to\infty}\left(n(e^{-\delta_1 t}-e^{-\delta_{1,n}t})+e^{-\delta_n t}\right)=2\pi t+e^{-2t}.$$

To establish the other inequality, note that by the previous theorem, the prime indices yield a subsequence converging to the lower bound. □

Proof of Theorem 9.4. The upper limit of any sequence is a lower bound for its supremum, so

$$\limsup_{n\to\infty} |tC_n^{\text{eucl}}|_{\text{Rips}} \leq \sup_{n\in\mathbb{N}} |tC_n^{\text{eucl}}|_{\text{Rips}}.$$

To show the converse inequality, we first observe that

$$|tC_n^{\text{eucl}}|_{\text{Rips}} \leq |tC_{2n}^{\text{eucl}}|_{\text{Rips}}$$

holds for any odd $n \in \mathbb{N}$. To see this, we look at their difference. Using the facts that n and $2n$ have the same odd divisors, that $\delta_n = \delta_{n,n}$ and $\delta_{n,2n} = \delta_{2n} = 2$, this can be simplified to:

$$|tC_{2n}^{\text{eucl}}|_{\text{Rips}} - |tC_n^{\text{eucl}}|_{\text{Rips}} = \sum_{\substack{\text{odd } r|n \\ r\neq n}} (\phi_{2n,r}(t) - \phi_{n,r}(t)) + e^{-\delta_n t} - e^{-2t},$$

where $\phi_{n,r}$ is as defined in the proof of Theorem 9.3, where we also showed that $\phi_{n+1,r} \geq \phi_{n,r}$ for all n and r. Using this latter fact and the fact that $\delta_n \leq 2$, we now have

$$\phi_{2n,r}(t) - \phi_{n,r}(t) \geq 0 \qquad \text{and} \qquad e^{-\delta_n t} - e^{-2t} \geq 0,$$

so the difference is indeed nonnegative.

Therefore, the supremum may be calculated over the even numbers:

$$\sup_{n\in\mathbb{N}} |tC_n^{\text{eucl}}|_{\text{Rips}} = \sup_{n \text{ even}} |tC_n^{\text{eucl}}|_{\text{Rips}}.$$

Now recall that for even n we have:

$$|tC_n^{\text{eucl}}|_{\text{Rips}} \leq \sum_{\text{odd } r\leq n} \frac{n}{r}(e^{-\delta_r t} - e^{-\delta_{r,n} t}) + e^{-2t}.$$

Furthermore, we saw in the proof of Theorem 9.3 that the expression on the right hand side is increasing in n (because $\phi_{n+1,r} \geq \phi_{n,r}$). Taking the supremum of both sides over all even integers n, we therefore have:

$$\sup_{n \text{ even}} |tC_n^{\text{eucl}}|_{\text{Rips}} \leq \lim_{\substack{n\to\infty \\ n \text{ even}}} \sum_{\text{odd } r\leq n} \frac{n}{r}(e^{-\delta_r t} - e^{-\delta_{r,n} t}) + e^{-2t} \leq \limsup_{n\to\infty} |tC_n^{\text{eucl}}|_{\text{Rips}},$$

where the second inequality is immediate from the proof of Theorem 9.3. □

Sketch proof of Theorem 9.6. By Corollary 8.4, the explicit formulas for Rips magnitudes of geodesic cycles are

$$|tC_n^{\text{geo}}|_{\text{Rips}} = \sum_{\substack{\text{odd } r|n \\ r\neq n}} \frac{n}{r}(e^{-\eta_r t} - e^{-\eta_{r,n} t}) + e^{-\eta_n t},$$

where

$$\eta_r = 2\pi\frac{\lfloor\frac{r}{2}\rfloor}{r} \qquad \text{and} \qquad \eta_{r,n} = 2\pi\left(\frac{1}{n} + \frac{\lfloor\frac{r}{2}\rfloor}{r}\right).$$

A similar procedure as in the Euclidean case now allows for the calculation of limits of various subsequences, as well as the upper and the lower limit. The

main difference is that in this case, the series obtained as the upper limit does not converge anymore:

$$\limsup_{n \to \infty} |tC_n^{\text{geo}}|_{\text{Rips}} = \sum_{\text{odd } r} \frac{2\pi t}{r} e^{-\pi \frac{r-1}{r} t} + e^{-\pi t} = \infty. \qquad \square$$

References

[1] Michał Adamaszek. Clique complexes and graph powers. *Israel J. Math.*, 196(1):295–319, 2013.

[2] MichałAdamaszek and Henry Adams. The Vietoris-Rips complexes of a circle. *Pacific J. Math.*, 290(1):1–40, 2017.

[3] Michał Adamaszek, Henry Adams, Florian Frick, Chris Peterson, and Corrine Previte-Johnson. Nerve complexes of circular arcs. *Discrete Comput. Geom.*, 56(2):251–273, 2016.

[4] Michał Adamaszek, Henry Adams, and Samadwara Reddy. On Vietoris-Rips complexes of ellipses. *J. Topol. Anal.*, 11(3):661–690, 2019.

[5] Yasuhiko Asao. Magnitude homology of geodesic metric spaces with an upper curvature bound. Preprint, available at arXiv:1903.11794, 2019.

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

[7] Juan Antonio Barceló and Anthony Carbery. On the magnitudes of compact sets in Euclidean spaces. *Amer. J. Math.*, 140(2):449–494, 2018.

[8] Ulrich Bauer and Herbert Edelsbrunner. The Morse theory of Čech and Delaunay complexes. *Trans. Amer. Math. Soc.*, 369(5):3741–3762, 2017.

[9] Omer Bobrowski and Matthew Strom Borman. Euler integration of Gaussian random fields and persistent homology. *J. Topol. Anal.*, 4(1):49–70, 2012.

[10] Omer Bobrowski and Matthew Kahle. Topology of random geometric complexes: a survey. *J. Appl. Comput. Topol.*, 1(3-4):331–364, 2018.

[11] Magnus Bakke Botnan and William Crawley-Boevey. Decomposition of persistence modules. To appear in Proc. Amer. Math. Soc, 2018.

[12] Peter Bubenik, Michael Hull, Dhruv Patel, and Benjamin Whittle. Persistent homology detects curvature. *Inverse Problems*, 36(2):025008, jan 2020.

[13] Peter Bubenik and Nikola Milićević. Homological algebra for persistence modules. Preprint, available at arXiv:1905.052744, 2018.

[14] Peter Bubenik and Jonathan A. Scott. Categorification of persistent homology. *Discrete Comput. Geom.*, 51(3):600–627, 2014.

[15] Gunnar Carlsson and Benjamin Filippenko. Persistent homology of the sum metric. *J. Pure Appl. Algebra*, 224(5):106244, 2020.

[16] Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete Comput. Geom.*, 42(1):71–93, 2009.

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

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

[19] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *J. Algebra Appl.*, 14(5):1550066, 8, 2015.

[20] Vin De Silva and Gunnar Carlsson. Topological estimation using witness complexes. In *Proceedings of the First Eurographics Conference on Point-Based Graphics*, SPBG'04, pages 157–166, Aire-la-Ville, Switzerland, Switzerland, 2004. Eurographics Association.

[21] Herbert Edelsbrunner and John Harer. Persistent homology—a survey. In *Surveys on discrete and computational geometry*, volume 453 of *Contemp. Math.*, pages 257–282. Amer. Math. Soc., Providence, RI, 2008.

[22] Herbert Edelsbrunner and John L. Harer. *Computational topology.* American Mathematical Society, Providence, RI, 2010. An introduction.

[23] Brittany Terese Fasy, Samuel Micka, David L Millman, Anna Schenfisch, and Lucia Williams. Challenges in reconstructing shapes from Euler characteristic curves. Preprint, available at arXiv:1811.11337, 2018.

[24] Heiko Gimperlein and Marcus Goffeng. On the magnitude function of domains in euclidean space. To appear in Amer. J. Math., 2017.

[25] Kiyonori Gomi. Magnitude homology of geodesic space. Preprint, available at arXiv:1902.07044, 2019.

[26] Dejan Govc. On the definition of the homological critical value. *J. Homotopy Relat. Struct.*, 11(1):143–151, 2016.

[27] Yuzhou Gu. Graph magnitude homology via algebraic morse theory. Preprint, available at arXiv:1809.07240v1, 2018.

[28] Allen Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.

[29] Jean-Claude Hausmann. On the Vietoris-Rips complexes and a cohomology theory for metric spaces. In *Prospects in topology (Princeton, NJ, 1994)*, volume 138 of *Ann. of Math. Stud.*, pages 175–188. Princeton Univ. Press, Princeton, NJ, 1995.

[30] Teresa Heiss and Hubert Wagner. Streaming algorithm for Euler characteristic curves of multidimensional images. In *Computer analysis of images and patterns. Part I*, volume 10424 of *Lecture Notes in Comput. Sci.*, pages 397–409. Springer, Cham, 2017.

[31] Richard Hepworth and Simon Willerton. Categorifying the magnitude of a graph. *Homology Homotopy Appl.*, 19(2):31–60, 2017.

[32] Ryuki Kaneta and Masahiko Yoshinaga. Magnitude homology of metric spaces and order complexes. Preprint, available at arXiv:1803.04247, 2018.

[33] Tom Leinster. The Euler characteristic of a category. *Doc. Math.*, 13:21–49, 2008.

[34] Tom Leinster. The magnitude of metric spaces. *Doc. Math.*, 18:857–905, 2013.

[35] Tom Leinster and Mark W. Meckes. The magnitude of a metric space: from category theory to geometric measure theory. In *Measure theory in non-smooth spaces*, Partial Differ. Equ. Meas. Theory, pages 156–193. De Gruyter Open, Warsaw, 2017.

[36] Tom Leinster and Michael Shulman. Magnitude homology of enriched categories and metric spaces. Preprint, available at arXiv:1711.00802, 2017.

[37] Tom Leinster and Simon Willerton. On the asymptotic magnitude of subsets of Euclidean space. *Geom. Dedicata*, 164:287–310, 2013.

[38] Michael Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Found. Comput. Math.*, 15(3):613–650, 2015.

[39] Mark W. Meckes. Positive definite metric spaces. *Positivity*, 17(3):733–757, 2013.

[40] Mark W. Meckes. Magnitude, Diversity, Capacities, and Dimensions of Metric Spaces. *Potential Anal.*, 42(2):549–572, 2015.

[41] J. Milnor. *Morse theory*. Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51. Princeton University Press, Princeton, N.J., 1963.

[42] Nina Otter. Magnitude meets persistence. Homology theories for filtered simplicial sets. Preprint, available at arXiv:1807.01540v1, 2018.

[43] Leonid Polterovich, Egor Shelukhin, and Vukašin Stojisavljević. Persistence modules with operators in Morse and Floer theory. *Mosc. Math. J.*, 17(4):757–786, 2017.

[44] Katharine Turner, Sayan Mukherjee, and Doug M. Boyer. Persistent homology transform for modeling shapes and surfaces. *Inf. Inference*, 3(4):310–344, 2014.

[45] Žiga Virk. Rips complexes as nerves and a functorial Dowker-nerve diagram. Preprint, available at arXiv:1906.04028, 2019.

[46] Pawin Vongmasa and Gunnar Carlsson. Exterior critical series of persistence modules. Preprint, available at arXiv:1305.4780, 2014.

Institute of Mathematics, University of Aberdeen, Aberdeen, United Kingdom AB24 3UE

E-mail address: `dejan.govc@abdn.ac.uk`

Institute of Mathematics, University of Aberdeen, Aberdeen, United Kingdom AB24 3UE

E-mail address: `r.hepworth@abdn.ac.uk`