involve
a journal of mathematics

Weighted persistent homology

Gregory Bell, Austin Lawson, Joshua Martin,
James Rudzinski and Clifford Smyth

msp

2019 vol. 12, no. 5

INVOLVE 12:5 (2019)
dx.doi.org/10.2140/involve.2019.12.823

Weighted persistent homology

Gregory Bell, Austin Lawson, Joshua Martin,
James Rudzinski and Clifford Smyth

(Communicated by Józef H. Przytycki)

We introduce weighted versions of the classical Čech and Vietoris–Rips complexes. We show that a version of the Vietoris–Rips lemma holds for these weighted complexes and that they enjoy appropriate stability properties. We also give some preliminary applications of these weighted complexes.

1. Introduction

Topological data analysis (TDA) provides a means for the power of algebraic topology to be used to better understand the shape of a data set. In the traditional approach to TDA, isometric balls of a fixed radius $r > 0$ are centered at each data point in some ambient Euclidean space. One then constructs the nerve of the union of these balls and computes the simplicial homology of this nerve. Computationally, this approach is infeasible for large data sets or high-dimensional data, so instead one computes the so-called Vietoris–Rips complex, which is the flag complex over the graph obtained by placing an edge between any pair of vertices that are at distance no more than $2r$ from each other. The key idea of TDA is to allow the radius of these balls to vary and to compute simplicial homology for each value of this radius to create a topological profile of the space. This profile is encoded in either a barcode or a persistence diagram. Topological features such as holes or voids that exist for a relatively large interval of radii are said to persist and are believed to be more important than more transient features that exist for very short intervals of radii. (There are, however, important exceptions to this rule of thumb; see [Bendich et al. 2016].)

In the traditional model, the radius of each ball is the same and can be modeled by the linear function of time $r(t) = rt$. In this paper, we consider a model of computing persistent homology in which the radius of each ball is allowed to be a

MSC2010: primary 55N35; secondary 55U99, 68U10.
Keywords: persistent homology, weighted persistent homology, stability, Vietoris–Rips complex, Čech complex, interleaving, bottleneck distance, persistence diagram.
Smyth was supported by NSA MSP Grant H98230-13-1-0222 and by a grant from the Simons Foundation (Grant Number 360486, CS).

different monotonic function $r_x(t)$ at each point x. In this way we can emphasize certain data points by assigning or *weighting* them with larger and/or more quickly growing balls and de-emphasize others by weighting them with smaller and/or more slowly growing balls. This is appropriate in the case of a noisy data set, for instance, as an alternative to throwing away data that fails to meet some threshold of significance. Various other methods of enhancing persistence with weights have been considered; see, e.g., [Buchet et al. 2016; Edelsbrunner and Morozov 2013; Petri et al. 2013; Ren et al. 2017; 2018].

The weighted model we propose fits into the framework of generalized persistence in the sense of [Bubenik et al. 2015]. We show that it enjoys many of the properties familiar from the techniques of traditional persistent homology. We prove a weighted Vietoris–Rips lemma (Theorem 3.2) that relates our weighted Čech and Rips complexes in the same way that they are related in the case of isometric balls. We also show that the persistent homology computed over weighted complexes is stable with respect to small perturbations of the rates of growth and/or the points in the data set (Theorem 4.1). Moreover, packages for computing persistent homology such as Javaplex [Adams et al. 2014] and Perseus [Mischaikow and Nanda 2013] are capable of handling our weighted persistence with the same complexity as unweighted persistence by merely adjusting inputs to the package functions.

As a proof of concept, we apply our methods to the Modified National Institute of Standards and Technology (MNIST) data set of handwritten digits translated into pixel information. Our method proves more effective than isometric persistence in finding the number 8 from among these handwritten digits. (We chose 8 for its unique 1-dimensional homology among these digits.) We found our methods to be 95.8% accurate as opposed to isometric persistence's 92.07% accuracy. This experiment was chosen to demonstrate the performance of weighted persistence over usual persistence, but it should be noted that neither method approaches the accuracy of state-of-the art computer vision and we make no claim that we are improving on known methods.

In Section 2, we provide the background definitions that are needed for what follows and describe our weighted persistence model. In Section 3 we prove the weighted Vietoris–Rips lemma and indicate how persistent homology packages can be used to compute weighted persistence. In Section 4 we establish our stability results. Our experiments on MNIST data appear in Section 5. We end with some remarks and questions for further study.

2. Preliminaries

We begin by defining some terminology and setting our notation. We will assume some familiarity with simplicial homology and the basic ideas of topological data analysis. For details, we refer to [Edelsbrunner and Harer 2010; Rotman 1988].

In algebraic topology, simplicial homology is a tool that assigns to any simplicial complex K a collection of $\mathbb{Z}$-modules $H_0(K), H_1(K), \ldots$, called *homology groups*, in such a way that the rank of $H_n(K)$ describes the number of "n-dimensional holes" in K. For our purposes, we replace the standard definition in terms of $\mathbb{Z}$-modules with vector spaces (usually over the field with two elements, for ease of computation). We therefore refer to *homology vector spaces* instead of homology groups. We do not attempt to define $H_n(K)$ here, but instead refer to any text in algebraic topology, such as [Rotman 1988].

Let $\mathcal{U}$ be a collection of sets. We define the *nerve* $\mathcal{N}(\mathcal{U})$ to be the abstract simplicial complex with vertex set $\mathcal{U}$ with the property that the subset $\{U_0, U_1, \ldots, U_n\}$ of $\mathcal{U}$ spans an n-simplex in $\mathcal{N}$ whenever $\bigcap_{i=0}^{n} U_i \neq \varnothing$.

Let (X, d) be a metric space. We define $B_r(x) = \{y \in X \mid d(x, y) < r\}$ and $\bar{B}_r(x) = \{y \in X \mid d(x, y) \leq r\}$ to be the open and closed balls of radius r about x, respectively. (Note that we're abusing notation since in a general metric space $\bar{B}_r(x)$ is not necessarily the closure of the open ball, usually denoted by $B_r(x)$). We most often consider examples where X is a subset of $\mathbb{R}^d$ and $d(x, y) = \|x - y\|$ is the Euclidean distance between x and y. For a real number $r \geq 0$, we define the *Čech complex of X at scale r* by $\check{\mathrm{C}}\mathrm{ech}(r) = \mathcal{N}\{\bar{B}_r(x) \mid x \in X\}$.

We generalize this construction by allowing the radius of the ball around each element x to depend on x. Let $\boldsymbol{r} : X \to [0, \infty)$ be any function. We define the *weighted $\boldsymbol{r}$-*Čech *complex* $\check{\mathrm{C}}\mathrm{ech}(\boldsymbol{r})$ of X by $\check{\mathrm{C}}\mathrm{ech}(\boldsymbol{r}) = \mathcal{N}\{\bar{B}_{\boldsymbol{r}(x)}(x)\}$.

In practice, it is difficult to determine whether an intersection of balls is nonempty. A much simpler construction to use is the Vietoris–Rips complex. For a given parameter $r \geq 0$ the *Vietoris–Rips complex* is the flag complex of the 1-skeleton of the Čech complex; i.e., a collection of $n + 1$ balls forms an n-simplex in the Vietoris–Rips complex if and only if the balls are pairwise intersecting. For the Vietoris–Rips complex we identify each ball with its center, so that the *Vietoris–Rips complex at scale r* is $\mathrm{VR}(r) = \{\sigma \subset X \mid \mathrm{diam}(\sigma) \leq 2r\}$. Similarly, if $\boldsymbol{r} : X \to [0, \infty)$, the *weighted $\boldsymbol{r}$-Vietoris–Rips complex* is

$$\mathrm{VR}(\boldsymbol{r}) = \{\sigma \subset X \mid d(x, y) \leq \boldsymbol{r}(x) + \boldsymbol{r}(y) \text{ for all } x, y \in \sigma \text{ with } x \neq y\}.$$

Fix $\boldsymbol{r} : X \to [0, \infty)$ and consider the simplicial complex $\check{\mathrm{C}}\mathrm{ech}(\boldsymbol{r})$ (or $\mathrm{VR}(\boldsymbol{r})$). Using simplicial homology with field coefficients, one can associate homology vector spaces $H_*(\check{\mathrm{C}}\mathrm{ech}(\boldsymbol{r}))$ to these simplicial complexes. Whenever $t_0 \leq t_1$ there is a natural inclusion map of simplicial complexes given by $\iota : \check{\mathrm{C}}\mathrm{ech}(t_0\boldsymbol{r}) \to \check{\mathrm{C}}\mathrm{ech}(t_1\boldsymbol{r})$ (or the corresponding inclusion of the Vietoris–Rips complexes). By functoriality, there is an induced linear map on homology $\iota_* : H_* \check{\mathrm{C}}\mathrm{ech}(t_0\boldsymbol{r})) \to H_* \check{\mathrm{C}}\mathrm{ech}(t_1\boldsymbol{r})$.

Let $X \subset \mathbb{R}^d$ be finite. Although we defined the weighted complexes above for any function $\boldsymbol{r} : X \to [0, \infty)$, we want to study the persistence properties of these weighted complexes. For example, in the case of the weighted Čech complex, we

want to study the evolution of homology as the radii of the balls grow to infinity. One straightforward way to do this would be to simply scale our weighted complexes linearly in the same way that one usually scales the isometric balls in persistent homology. We prefer a more flexible approach, which we describe in terms of radius functions.

Let $\mathcal{C}^1_+ = \mathcal{C}^1_+([0,\infty))$ denote the collection of differentiable bijective functions $\phi : [0,\infty) \to [0,\infty)$ with positive first derivative. By a *radius function* on X we mean a function $\boldsymbol{r} : X \to \mathcal{C}^1_+$. We denote the image function $\boldsymbol{r}(x)$ by $\boldsymbol{r}_x$.

For $t \geq 0$, we define the Čech *and Vietoris–Rips complexes at scale* t by

$$\check{\mathrm{C}}\mathrm{ech}_{\boldsymbol{r}}(t) = \mathcal{N}\{\bar{B}_{\boldsymbol{r}_x(t)}(x)\}$$

and

$$\mathrm{VR}_{\boldsymbol{r}}(t) = \{\sigma \subset X \mid d(x,y) \leq \boldsymbol{r}_x(t) + \boldsymbol{r}_y(t) \text{ for all } x, y \in \sigma \text{ with } x \neq y\},$$

respectively. We define the *entry function*,

$$f_{X,\boldsymbol{r}}(y) = \min_{x \in X}\{\boldsymbol{r}_x^{-1}(d(y,x))\}. \tag{1}$$

This function captures the scale t at which the point $y \in \mathbb{R}^d$ is first captured by some ball $\bar{B}_{\boldsymbol{r}_x(t)}(x)$; we have $f_{X,\boldsymbol{r}}(y) = t$ if and only if $y \in \bar{B}_{\boldsymbol{r}_x(t)}(x)$ for some x in X and $y \notin \bigcup_{x\in X} B_{\boldsymbol{r}_x(t)}(x)$. Thus we have the following proposition.

Proposition 2.1. *Let X be a finite subset of some Euclidean space $\mathbb{R}^d$. Suppose that $\boldsymbol{r}$ and $f_{X,\boldsymbol{r}}$ are defined as above. Then,*

$$f^{-1}_{X,\boldsymbol{r}}([0,t]) = \bigcup_{x\in X} \bar{B}(x, \boldsymbol{r}_x(t)).$$

It follows from the nerve lemma, see for example [Hatcher 2002, Corollary 4G.3], that $\check{\mathrm{C}}\mathrm{ech}_{\boldsymbol{r}}(t)$ is homotopy equivalent to $f^{-1}_{X,\boldsymbol{r}}([0,t])$.

3. A weighted Vietoris–Rips lemma

The Vietoris–Rips complex is much easier to compute than the Čech complex in high dimensions. To determine whether $n+1$ balls form an n-simplex in the Čech complex, we must check whether the balls intersect, a computationally complex problem. To determine whether $n+1$ balls $B_{r_i}(x_i)$ form a simplex in the Vietoris–Rips complex is computationally easy; only $\binom{n+1}{2}$ conditions $d(x_i,x_j) \leq r_i + r_j$ need be checked. Furthermore, if there are m points in X, it may be necessary to check all 2^m subcollections of balls to determine the Čech complex, whereas determining the Rips complex will only require checking $\binom{m}{2}$ pairs of points.

Our weighted Čech and Vietoris–Rips complexes are similar in spirit to weighted alpha complexes [Edelsbrunner and Harer 2010, III.4]. Both constructions seek to

permit "balls" with different sizes. Our constructions are simpler from a conceptual standpoint since the alpha complexes are built as subcomplexes of the Delaunay complex, which comes from the Voronoi diagram. Moreover, our complexes are computationally simple; indeed our method of finding weighted Vietoris–Rips complexes requires only marginally more computation than the unweighted Vietoris–Rips complex.

In particular, Javaplex and Perseus can compute regular (unweighted) persistent homology given input of a distance matrix M with $M_{i,j} = d(x_i, x_j)$. Inputting $M_{i,j} = d(x_i, x_j)/(r_i + r_j)$ allows these packages to compute the persistent homology with $r_{x_i}(t) = r_i t$ in the same time.

In computational problems it is common to use the Vietoris–Rips complex instead of the Čech complex to simplify the calculational overhead. The following theorem justifies this decision by saying that the Vietoris–Rips complex is "close" to the weighted Čech complex.

The classical Vietoris–Rips lemma can be stated as follows:

Theorem 3.1 [de Silva and Ghrist 2007]. *Let X be a set of points in $\mathbb{R}^d$ and let $t > 0$. Then*

$$\mathrm{VR}(t') \subseteq \check{\mathrm{C}}\mathrm{ech}(t) \subseteq \mathrm{VR}(t)$$

whenever $0 < t' \leq t(\sqrt{2d/(d+1)})^{-1}$.

The main result of this section is an extension of this result to the weighted case.

Theorem 3.2 (weighted Vietoris–Rips lemma). *Let X be a set of points in $\mathbb{R}^d$. Let $\boldsymbol{r} : X \to (0, \infty)$ be the corresponding weight function and let $t > 0$. Then*

$$\mathrm{VR}(t'\boldsymbol{r}) \subseteq \check{\mathrm{C}}\mathrm{ech}(t\boldsymbol{r}) \subseteq \mathrm{VR}(t\boldsymbol{r})$$

whenever $0 < t' \leq t(\sqrt{2d/(d+1)})^{-1}$.

Proof. The second containment $\check{\mathrm{C}}\mathrm{ech}(t\boldsymbol{r}) \subseteq \mathrm{VR}(t\boldsymbol{r})$ follows from the fact that the weighted Vietoris–Rips complex is the flag complex of the weighted Čech complex.

To show that $\mathrm{VR}(t'\boldsymbol{r}) \subset \check{\mathrm{C}}\mathrm{ech}(t\boldsymbol{r})$, we suppose there is some finite collection $\sigma = \{x_k\}_{k=0}^{\ell} \subseteq \mathbb{R}^d$ with $\ell > 0$ that is a simplex in $\mathrm{VR}(t'\boldsymbol{r})$ and show that this is also a simplex in $\check{\mathrm{C}}\mathrm{ech}(t\boldsymbol{r})$. We have $\|x_i - x_j\|_2 \leq t'(\boldsymbol{r}(x_i) + \boldsymbol{r}(x_j))$ whenever $i \neq j$.

Define a function $f : \mathbb{R}^d \to \mathbb{R}$ by

$$f(y) = \max_{0 \leq j \leq \ell} \left\{ \frac{\|x_j - y\|_2}{\boldsymbol{r}(x_j)} \right\}.$$

Clearly, f is continuous and $f(y) \to \infty$ as $\|y\|_2 \to \infty$. Thus f attains a minimum (say at y_0) on some compact set containing $\mathrm{Conv}(\{x_k\}_{k=0}^{\ell})$. (Here $\mathrm{Conv}(S)$ is the convex hull of the set $S \subseteq \mathbb{R}^d$.) We must have $\|x_i - y_0\|_2 / r(x_i) = f(y_0)$ for at least

one of the vertices x_i. By reordering the vertices, we may assume that

$$f(y_0) = \frac{1}{\boldsymbol{r}(x_j)}\|x_j - y_0\|_2 \quad \text{if } 0 \le j \le n,$$

$$f(y_0) > \frac{1}{\boldsymbol{r}(x_j)}\|x_j - y_0\|_2 \quad \text{if } n < j \le \ell.$$

Let

$$g(y) = \max_{0\le j\le n}\left\{\frac{1}{\boldsymbol{r}(x_j)}\|x_j - y\|_2\right\},$$

$$h(y) = \max_{n< j\le \ell}\left\{\frac{1}{\boldsymbol{r}(x_j)}\|x_j - y\|_2\right\}.$$

Now we wish to show that $y_0 \in \mathrm{Conv}(\{x_j\}_{j=0}^n)$. To this end we apply the separation theorem [Matoušek 2002] to obtain: either $y_0 \in \mathrm{Conv}(\{x_j\}_{j=0}^n)$ or there is a $v \in \mathbb{R}^d$ and a $C < 0$ such that $v x_j \ge 0$ for all $0 \le j \le n$ and $v y_0 < C$. Thus if $y_0 \notin \mathrm{Conv}(\{x_j\}_{j=0}^n)$ there is a $v \in \mathbb{R}^d$ such that $v(x_j - y_0) > 0$ for $0 \le j \le n$. We suppose that there is such a v and derive a contraction.

Since

$$\|x_j - (y_0 + \lambda v)\|_2^2 = \|x_j - y_0\|_2^2 - 2\lambda v(x_j - y_0) + \lambda^2\|v\|_2^2$$

for each $0 \le j \le n$, it follows that $g(y_0 + \lambda v) < f(y_0)$ for all $0 < \lambda < \lambda_1$, where

$$\lambda_1 = \min_{0\le j\le n}\left\{\frac{2v(x_j - y_0)}{\|v\|_2^2}\right\}.$$

Since $h(y)$ is continuous and $h(y_0) < f(y_0)$, there exists a λ_2 such that $h(y_0+\lambda v) < f(y_0)$ for $0 < \lambda < \lambda_2$. Thus, there exists a $\lambda > 0$ such that

$$f(y_0 + \lambda v) = \max\{g(y_0 + \lambda v), h(y_0 + \lambda v)\} < f(y_0),$$

contradicting the minimality of y_0.

By Carathéodory's theorem [Matoušek 2002] and reordering of vertices if necessary, y_0 is a convex combination of some subcollection of vertices $\{x_j\}_{j=0}^m$, where $m \le \min\{d, n\}$. It is not possible that $m = 0$. If so, then $y_0 = x_0$ and

$$f(y_0) = \frac{1}{\boldsymbol{r}(x_0)}\|x_0 - y_0\|_2 = 0$$

and f is identically zero. Since σ has dimension at least 1, it contains a vertex $x_1 \ne x_0$. It follows that

$$f(y_0) = f(x_0) > \frac{1}{\boldsymbol{r}(x_1)}\|x_1 - x_0\|_2 > 0,$$

which is a contradiction.

Let $\hat{x}_j = x_j - y_0$ for all $0 \le j \le m$. Note that

$$\|\hat{x}_j\|_2^2 = \boldsymbol{r}(x_j)^2 f(y_0)^2. \tag{2}$$

Since $y_0 \in \mathrm{Conv}(\{x_j\}_{j=0}^m)$, we know $y_0 = \sum_{j=0}^m a_j x_j$ for some set of nonnegative real numbers $a_0, \ldots, a_m$ that sum to 1. Thus $\sum_{j=0}^m a_j \hat{x}_j = 0$. By relabeling, we may assume that $a_0 \boldsymbol{r}(x_0) \ge a_j \boldsymbol{r}(x_j)$ when $j > 0$. Necessarily $a_0 > 0$ (otherwise $a_j = 0$ for all $0 \le j \le m$, a contradiction). Then,

$$\hat{x}_0 = -\sum_{j=0}^m \frac{a_j}{a_0} \hat{x}_j$$

and so

$$\boldsymbol{r}(x_0)^2 f(y_0)^2 = \|\hat{x}_0\|_2^2 = -\sum_{j=0}^m \frac{a_i}{a_0} \hat{x}_0 \hat{x}_j.$$

Among the indices $1, 2, \ldots, m$, there is some j_0 such that

$$\frac{1}{d} \boldsymbol{r}(x_0)^2 f(y_0)^2 \le \frac{1}{m} \boldsymbol{r}(x_0)^2 f(y_0)^2 \le -\frac{a_{j_0}}{a_0} \hat{x}_0 \hat{x}_{j_0}. \tag{3}$$

We must have $a_{j_0} > 0$. (Otherwise, $f(y_0) = 0$, which, as shown earlier, is a contradiction.) By reordering, we may assume $j_0 = 1$. Putting (1) and (2) together, we find

$$\begin{aligned} f(y_0)^2 \left(\boldsymbol{r}(x_0)^2 + \frac{2a_0 \boldsymbol{r}(x_0)^2}{a_1 d} + \boldsymbol{r}(x_1)^2 \right) & \\ = f(y_0)^2 \boldsymbol{r}(x_0)^2 + \frac{2a_0 f(y_0)^2 \boldsymbol{r}(x_0)^2}{a_1 d} + f(y_0)^2 \boldsymbol{r}(x_1)^2 & \\ \le \|\hat{x}_0\|_2^2 - 2\hat{x}_0 \hat{x}_1 + \|\hat{x}_1\|_2^2 & \\ = \|\hat{x}_0 - \hat{x}_1\|_2^2 & \\ = \|x_0 - x_1\|_2^2 & \\ \le (t'(\boldsymbol{r}(x_0) + \boldsymbol{r}(x_1)))^2. & \end{aligned}$$

We will now show that

$$\frac{f(y_0)^2}{t'} \le \frac{(\boldsymbol{r}(x_0)^2 + \boldsymbol{r}(x_1)^2)^2}{\boldsymbol{r}(x_0)^2 + 2a_0 \boldsymbol{r}(x_0)^2/(a_1 d) + \boldsymbol{r}(x_1)^2} \le \frac{2d}{d+1}.$$

It suffices to show, after cross-multiplying the right-hand inequality, that

$$\left(d - 1 + 4\frac{a_0}{a_1} \right) \boldsymbol{r}(x_0)^2 - 2(d+1) \boldsymbol{r}(x_0) \boldsymbol{r}(x_1) + (d-1) \boldsymbol{r}(x_1)^2 \ge 0.$$

Since

$$\frac{a_0}{a_1} \ge \frac{\boldsymbol{r}(x_1)}{\boldsymbol{r}(x_0)}$$

we get

$$\begin{aligned}\left(d-1+4\frac{a_0}{a_1}\right)\boldsymbol{r}(x_0)^2-2(d+1)\boldsymbol{r}(x_0)\boldsymbol{r}(x_1)+(d-1)\boldsymbol{r}(x_1)^2\\ \geq\left(d-1+4\frac{\boldsymbol{r}(x_1)}{\boldsymbol{r}(x_0)}\right)\boldsymbol{r}(x_0)^2-2(d+1)\boldsymbol{r}(x_0)\boldsymbol{r}(x_1)+(d-1)\boldsymbol{r}(x_1)^2\\ =(d-1)(\boldsymbol{r}(x_0)-\boldsymbol{r}(x_1))^2\geq 0,\end{aligned}$$

as desired. Our assumption that $t' \leq t(\sqrt{2d/(d+1)})^{-1}$ implies $f(y_0) \leq t$ and thus

$$y_0 \in \bigcap_{i=0}^{\ell} \bar{B}_{t\boldsymbol{r}(x_i)}(x_i).$$

Therefore $\sigma \in \check{\text{C}}\text{ech}(t\boldsymbol{r})$ and we are done. □

4. Stability

In this section we discuss the stability of our weighted persistence. Let X and Y be finite subsets of $\mathbb{R}^d$ with corresponding radii functionals $\boldsymbol{r} : X \to \mathcal{C}^1_+$ and $\boldsymbol{s} : Y \to \mathcal{C}^1_+$. Informally, we show that if $(X, \boldsymbol{r})$ and $(Y, \boldsymbol{s})$ are "close", i.e., are small perturbations of each other, then the corresponding entry functions $f_{X,\boldsymbol{r}}$ and $f_{Y,\boldsymbol{s}}$, see (1), are also "close" and hence the associated persistence diagrams must also be "close". We'll now make the definitions of these various types of closeness precise.

Let $\eta \subseteq X \times Y$ be a relation such that for every $x \in X$ there is a $y \in Y$ with $(x, y) \in \eta$ and for every $y \in Y$ there is an $x \in X$ with $(x, y) \in \eta$. We measure the closeness of X and Y with respect to η by

$$\|\eta\| := \max_{(x,y)\in\eta} d(x, y).$$

If L is any compact set and $h : L \to \mathbb{R}$ is continuous let

$$\|h\|_L := \max_{x\in L} |h(x)|.$$

Let K be a compact subset of $\mathbb{R}^d$ that contains $X \cup Y$. The closeness of $\boldsymbol{r}$ and $\boldsymbol{s}$ is measured by

$$D(\boldsymbol{r}, \boldsymbol{s})_{\eta,K} := \max_{(x,y)\in\eta} \|\boldsymbol{r}_x^{-1} - \boldsymbol{s}_y^{-1}\|_{[0,\text{diam}(K)]}.$$

The closeness of $f_{X,\boldsymbol{r}}$ and $f_{Y,\boldsymbol{s}}$ is measured by $\|f_{X,\boldsymbol{r}} - f_{Y,\boldsymbol{s}}\|_K$. We also define $S(\boldsymbol{r})_K := \max_{x\in X} \|(\boldsymbol{r}_x^{-1})'\|_{[0,\text{diam}(K)]}$.

As is common, we measure the closeness of persistence diagrams by the bottleneck distance. We'll give the definition of this metric in the remarks leading up to Theorem 4.5.

Theorem 4.1. *In the above notation we have the following bound on entry functions* (*see* (1)):

$$\|f_{X,\boldsymbol{r}} - f_{Y,\boldsymbol{s}}\|_K \leq D(\boldsymbol{r},\boldsymbol{s})_{\eta,K} + \|\eta\| \max(S(\boldsymbol{r})_K, S(\boldsymbol{s})_K)$$

Proof. There is some point z in the compact set K and some points $x \in X$ and $y \in Y$ so that

$$\|f_{X,\boldsymbol{r}} - f_{Y,\boldsymbol{s}}\|_K = |f_{X,\boldsymbol{r}}(z) - f_{Y,\boldsymbol{s}}(z)| = |\boldsymbol{r}_x^{-1}(d(z,x)) - \boldsymbol{s}_y^{-1}(d(z,y))|.$$

We first suppose $\boldsymbol{r}_x^{-1}(d(z,x)) \geq \boldsymbol{s}_y^{-1}(d(z,y))$. Let $x' \in X$ such that $(x', y) \in \eta$. Since $f_{X,\boldsymbol{r}}$ is a minimum, $\boldsymbol{r}_{x'}^{-1}(d(z,x')) \geq \boldsymbol{r}_x^{-1}(d(z,x))$ and we have

$$\begin{aligned}
&\|f_{X,\boldsymbol{r}} - f_{Y,\boldsymbol{s}}\|_K \\
&\qquad \leq |\boldsymbol{r}_{x'}^{-1}(d(z,x')) - \boldsymbol{s}_y^{-1}(d(z,y))| \\
&\qquad \leq |\boldsymbol{r}_{x'}^{-1}(d(z,x')) - \boldsymbol{s}_y^{-1}(d(z,x'))| + |\boldsymbol{s}_y^{-1}(d(z,x')) - \boldsymbol{s}_y^{-1}(d(z,y))|. \qquad (4)
\end{aligned}$$

Since $d(z,x') \in [0, \operatorname{diam}(K)]$,

$$|\boldsymbol{r}_{x'}^{-1}(d(z,x')) - \boldsymbol{s}_y^{-1}(d(z,x'))| \leq D(\boldsymbol{r},\boldsymbol{s})_{\eta,K}.$$

Since $|d(z,x') - d(z,y)| \leq d(x',y) \leq \|\eta\|$ we apply the mean value theorem to obtain the bound

$$|\boldsymbol{s}_y^{-1}(d(z,x')) - \boldsymbol{s}_y^{-1}(d(z,y))| \leq \|\eta\| \, \|(\boldsymbol{s}_y^{-1})'\|_{[0,\operatorname{diam}(K)]} \leq \|\eta\| \max(S(\boldsymbol{r})_K, S(\boldsymbol{s})_K).$$

Together, these last two bounds give the bound of the theorem. A similar argument gives the same bound if $\boldsymbol{r}_x^{-1}(d(z,x)) \leq \boldsymbol{s}_y^{-1}(d(z,y))$. □

If one has free choice of the perturbed set $(Y, \boldsymbol{s})$ it is clear that $\|f_{X,\boldsymbol{r}} - f_{Y,\boldsymbol{s}}\|_K$ can be made arbitrarily large. This could be done, say by adding a point to Y that is arbitrarily far from any point in X or by making one $\boldsymbol{s}_y$ arbitrarily larger than any $\boldsymbol{r}_x$. The upper bound of Theorem 4.1 is also a bound on how extreme such perturbations may be.

We have the following immediate corollary of Theorem 4.1.

Corollary 4.2. *If the radii functions are all linear, i.e., if there are positive constants r_x and s_y for all $x \in X$ and $y \in Y$ such that $r_x(t) = r_x t$ and $s_y(t) = s_y t$, then*

$$\|f_{X,r} - f_{Y,s}\|_K \leq \operatorname{diam}(K) \max_{(x,y)\in\eta} \left| \frac{1}{r_x} - \frac{1}{s_y} \right| + \|\eta\| \max\left(\max_{x\in X} \frac{1}{r_x}, \max_{y\in Y} \frac{1}{s_y} \right).$$

For our next two corollaries, let X and Y have the same cardinality and let $m : X \to Y$ be a bijection. We now consider each point $x \in X$ as being perturbed to a point $m(x) \in Y$ and hence set $\eta = \{(x, m(x)) : x \in X\}$. We have the following point-stability result in which the points are perturbed but the weight functions stay the same.

Corollary 4.3 (point-stability). *If only the locations of the points are perturbed and the radius functions stay the same, i.e., $\boldsymbol{s}_{m(x)}(t) = \boldsymbol{r}_x(t)$ for all $x \in X$, then*

$$\|f_{X,\boldsymbol{r}} - f_{Y,s}\|_K \leq \max_{x\in X} d(x, m(x)) \|(\boldsymbol{r}_x^{-1})'\|_{[0,\mathrm{diam}(k)]}.$$

Proof. We follow the proof of Theorem 4.1. Take $x' \in X$ such that $m(x') = y$. Then $\boldsymbol{S}_y = \boldsymbol{r}_{x'}$ and the first term in the upper bound in inequality (4) is 0. Since the second term in that upper bound is bounded above by $d(x', m(x'))\|(\boldsymbol{r}_{x'}^{-1})'\|_{[0,\mathrm{diam}(K)]}$, the bound of the corollary holds. □

The next corollary is a weight-function stability result concerning a case in which the points stay the same ($Y = X$ and $m(x) = x$) but the weight functions are perturbed.

Corollary 4.4 (weight-function stability). *If only the radii functions are perturbed and the points stay the same, then*

$$\|f_{X,\boldsymbol{r}} - f_{X,s}\|_K \leq \max_{x\in X} \|\boldsymbol{r}_x^{-1} - \boldsymbol{s}_x^{-1}\|_{[0,\mathrm{diam}(K)]}.$$

Proof. Again following the proof of Theorem 4.1 we take $x' = m(x') = y$. Now the second term in the upper bound of (4) is 0 and the first term is $|\boldsymbol{r}_y^{-1} - \boldsymbol{s}_y^{-1}|$, where $t = d(z, y)$. The corollary follows. □

We now show the stability of the persistence diagrams of $f_{X,\boldsymbol{r}}$ under perturbations of X and $\boldsymbol{r}$. Let $f : K \to [0, \infty)$ be a real-valued function on a compact set $K \subseteq \mathbb{R}^d$. The *persistence diagram* of f, $\mathrm{dgm}(f)$, is a multiset of points in $[0, +\infty]^2$ recording the appearance and disappearance of homological features in $f^{-1}([0, t])$ as t increases. Each point (b, d) in the diagram tracks a single homological feature, recording the scale $t = b$ at which the feature first appears and the scale $t = d$ at which it disappears [Edelsbrunner and Harer 2010]. It should also be noted that if one considers the birth-death pair as an interval, we obtain the *barcode* as seen in [Zomorodian and Carlsson 2005] (see Figures 2 and 3). Given two functions $f, g : K \to [0, \infty]$, let $P = \mathrm{dgm}(f)$ and $Q = \mathrm{dgm}(g)$ be the corresponding persistence diagrams (where as usual we include all points along the diagonal in P and Q). We let N denote the set of all bijections from P to Q. We recall that the *bottleneck distance* between the diagrams [Edelsbrunner and Harer 2010] is given by

$$d_B(\mathrm{dgm}(f), \mathrm{dgm}(g)) = \inf_{\gamma\in N} \sup_{x\in P} \|x - \gamma(x)\|_\infty.$$

Theorem 4.5 [Cohen-Steiner et al. 2007, Theorem 6.9]. *Suppose $\mathcal{X}$ is a triangulable space and that $f : \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{X} \to \mathbb{R}$ are tame, continuous functions. If $|f - g|$ is bounded, then for each n*

$$d_B(\mathrm{dgm}_n(f), \mathrm{dgm}_n(g)) \leq \|f - g\|_\infty,$$

where d_B denotes the bottleneck distance and $\mathrm{dgm}_n(f)$ *denotes the n-th persistence diagram of the filtration of* f.

We refer to [Edelsbrunner and Harer 2010] for the technical definitions of tame and triangulable. Note that as our spaces are nerves of balls around finite collections of points, they are finite simplicial complexes. Hence they are triangulable and only admit tame functions. Thus for our setting we get the following corollary.

Corollary 4.6. *Let X and Y be finite subsets of $\mathbb{R}^d$ and let $\boldsymbol{r} : X \to \mathcal{C}^1_+$ and $\boldsymbol{s} : Y \to \mathcal{C}^1_+$. Suppose that $\eta \subseteq X \times Y$ is a relation as above and K is a compact subset of $\mathbb{R}^d$ containing X and Y. Then for each n,*

$$d_B(\mathrm{dgm}_n(f_{X,\boldsymbol{r}}), \mathrm{dgm}_n(f_{Y,\boldsymbol{s}})) \leq D(\boldsymbol{r}, \boldsymbol{s})_{\eta,K} + \|\eta\| \max(S(\boldsymbol{r})_K, S(\boldsymbol{s})_K).$$

5. MNIST 8's recognition

In this section, we give an application of weighted persistence to a simple computer vision problem. We apply our methods to the Modified National Institute of Standards and Technology (MNIST) data set of handwritten digits. We should emphasize that this application is simply a proof of concept; our methods to detect the handwritten number 8 fall well short of state-of-the-art methods [Cireşan et al. 2012].

The MNIST data set consists of handwritten digits (0 through 9) translated into pixel information. Each data point contains a label and 784 other values ranging from 0 to 255 that correspond to a 28 by 28 grid of pixels. The values 0 through 255 correspond to the intensity of the pixels in gray-scale with 0 meaning completely black and 255 meaning completely white. Considering the digits from 0 through 9, unweighted persistence would easily be able to classify these numbers as having zero, one, or two holes, provided they are written precisely; however, real handwritten digits present a challenge. Consider an 8 as in Figure 1. Unweighted persistence would pick up on two holes, but one of those holes might be slightly too small and ultimately considered insignificant; see Figure 3. Our methods are able to pick up on both holes and would count them as significant; see Figure 2. We chose to work with the digit 8 due to its unique homology.

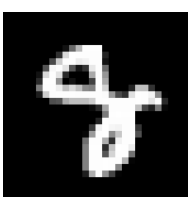

Figure 1. An 8 converted to a 28 by 28 grid of pixels.

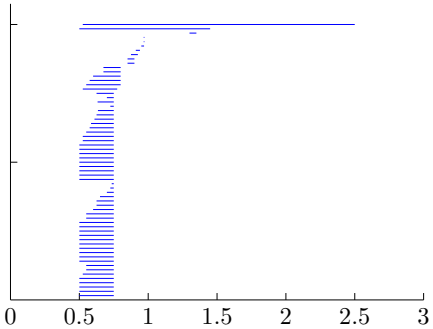

Figure 2. Weighted persistence on the image from Figure 1 produces a barcode that clearly has two long bars in dimension 1.

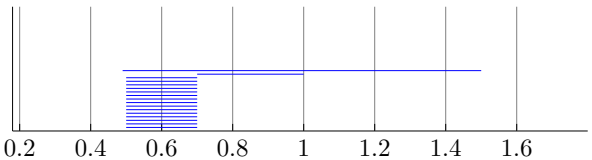

Figure 3. Unweighted persistence on the image from Figure 1 produces a barcode that has one long bar (in 1-homology). The second-longest bar is hard to distinguish (in length) from the rest.

To begin, we convert each 28 by 28 to a set of points in the plane. We treat the location of a value in the matrix as a location in the plane. That is, the value in the i-th row, j-th column corresponds to the point (i, j). The weight on each point is exactly its corresponding pixel intensity. Using this set of points and corresponding weights we calculate persistent homology via weighted Rips complexes. We test this method's performance against the unweighted case where all nonzero pixel values have the uniform weight of 1; again we calculate persistence in this case via Rips complexes.

We compare weighted persistence to unweighted persistence by measuring the accuracy of classifying 8's. Notice in the barcodes that the deciding factor in determining an 8 is the ability to distinguish the length of the second longest bar from the length of the third longest and smaller bars. For this reason, we consider the ratio of the third longest bar to the second longest bar. We will say (arbitrarily) that a barcode represents an 8 if this ratio is less than $\frac{1}{2}$. For each of the 42,000 handwritten digits in the MNIST data set, we compute both weighted and unweighted persistence and collect the predictions. We obtain the confusion matrices as in Table 1.

Notice that the weighted persistence has an accuracy rate of 95.8% whereas unweighted persistence had an accuracy of 92.07%. A full summary can be seen

	weighted persistence		unweighted persistence	
	predicted not 8	predicted 8	predicted not 8	predicted 8
not 8	36487	1450	35869	2068
is 8	633	3430	1261	2802

Table 1. The confusion matrices show that weighted persistence outperforms its unweighted counterpart.

	weighted persistence	unweighted persistence
accuracy	0.9504	0.9207
sensitivity	0.9618	0.9455
specificity	0.8442	0.6896
pos. predicted value	0.9829	0.9660
neg. predicted value	0.7029	0.5754
prevalence	0.9033	0.9033
balanced accuracy	0.9030	0.8176

Table 2. Weighted and unweighted persistence compared.

in Table 2. We view this result as promising for potential future applications of weighted persistence.

6. Concluding remarks and open questions

The method of weighted persistence satisfies the appropriate Vietoris–Rips lemma, is stable under small perturbations of the points, or the weights, or both, and can be successfully applied to data such as the MNIST data set to improve upon usual persistence. Furthermore, it is just as easy to calculate weighted persistence for balls growing at linear rates as it is to calculate regular persistence. We conclude the paper with some further observations and questions.

One can imagine weighted persistence as interpolating between two extreme approaches to a data set that is partitioned into data D and noise N. More precisely, we consider a noisy data set X. Various methods exist to filter X into data D and noise N. Traditional persistence can be applied to $D \cup N$ in two ways. We can either assign the same radius to every point of $D \cup N$ or we can throw the points of N out entirely and compute persistence on D alone. Using weighted persistence, we can assign the radius 0 to each point of N and compute weighted persistence of $D \cup N$. It is easy to see that this will differ from persistence of D itself only in dimension 0. By gradually increasing the N-radii from 0 to 1, our stability results can be interpreted as producing a continuum of barcodes/persistence diagrams that

interpolate between the usual persistence applied to D and the usual persistence applied to $D \cup N$ (in dimensions above 0); see [Lawson 2016].

As mentioned in the Introduction, weighted persistence fits into the framework of generalized persistence in the sense of [Bubenik et al. 2015]. This direction was explored in detail in [Martin 2016].

Finally, it would be interesting to apply weighted persistence to the MNIST data set to determine its effectiveness in distinguishing the 1-homology of the other nine digits. One complication is that the number 4 presents an interesting challenge since it is appropriate to write it both as a simply connected space and as a space with nontrivial H_1. Distinguishing 1-homology creates three clusters of digits from which we could use other machine-learning techniques to create an ensemble and make accurate predictions.

References

[Adams et al. 2014] H. Adams, A. Tausz, and M. Vejdemo-Johansson, "javaPlex: a research software package for persistent (co)homology", pp. 129–136 in *International Congress on Mathematical Software 2014* (Seoul, 2014), edited by H. Hong and C. Yap, Lecture Notes in Comput. Sci. **8592**, 2014. Software available at http://appliedtopology.github.io/javaplex. Zbl

[Bendich et al. 2016] P. Bendich, J. S. Marron, E. Miller, A. Pieloch, and S. Skwerer, "Persistent homology analysis of brain artery trees", *Ann. Appl. Stat.* **10**:1 (2016), 198–218. MR

[Bubenik et al. 2015] P. Bubenik, V. de Silva, and J. Scott, "Metrics for generalized persistence modules", *Found. Comput. Math.* **15**:6 (2015), 1501–1531. MR Zbl

[Buchet et al. 2016] M. Buchet, F. Chazal, S. Y. Oudot, and D. R. Sheehy, "Efficient and robust persistent homology for measures", *Comput. Geom.* **58** (2016), 70–96. MR Zbl

[Cireşan et al. 2012] D. Cireşan, U. Meier, and J. Schmidthuber, "Multi-column deep neural networks for image classification", pp. 3642–3649 in *2012 IEEE Conference on Computer Vision and Pattern Recognition* (Providence, RI, 2012), IEEE, Piscataway, NJ, 2012.

[Cohen-Steiner et al. 2007] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, "Stability of persistence diagrams", *Discrete Comput. Geom.* **37**:1 (2007), 103–120. MR Zbl

[Edelsbrunner and Harer 2010] H. Edelsbrunner and J. L. Harer, *Computational topology: an introduction*, Amer. Math. Soc., Providence, RI, 2010. MR Zbl

[Edelsbrunner and Morozov 2013] H. Edelsbrunner and D. Morozov, "Persistent homology: theory and practice", pp. 31–50 in *European Congress of Mathematics* (Kraków, 2012), edited by R. Latała et al., Eur. Math. Soc., Zürich, 2013. MR Zbl

[Hatcher 2002] A. Hatcher, *Algebraic topology*, Cambridge Univ. Press, 2002. MR Zbl

[Lawson 2016] A. Lawson, *Multi-scale persistent homology*, Ph.D. thesis, University of North Carolina at Greensboro, 2016, available at https://search.proquest.com/docview/1806825235.

[Martin 2016] J. Martin, *Multiradial (multi)filtrations and persistent homology*, master's thesis, University of North Carolina at Greensboro, 2016, available at https://search.proquest.com/docview/1816997091.

[Matoušek 2002] J. Matoušek, *Lectures on discrete geometry*, Graduate Texts in Math. **212**, Springer, 2002. MR Zbl

[Mischaikow and Nanda 2013] K. Mischaikow and V. Nanda, "Morse theory for filtrations and efficient computation of persistent homology", *Discrete Comput. Geom.* **50**:2 (2013), 330–353. MR Zbl

[Petri et al. 2013] G. Petri, M. Scolamiero, I. Donato, and F. Vaccarino, "Topological strata of weighted complex networks", *PLOS One* **8**:6 (2013), art. id. e66506.

[Ren et al. 2017] S. Ren, C. Wu, and J. Wu, "Computational tools in weighted persistent homology", preprint, 2017. arXiv

[Ren et al. 2018] S. Ren, C. Wu, and J. Wu, "Weighted persistent homology", *Rocky Mountain J. Math.* **48**:8 (2018), 2661–2687. MR Zbl

[Rotman 1988] J. J. Rotman, *An introduction to algebraic topology*, Graduate Texts in Math. **119**, Springer, 1988. MR Zbl

[de Silva and Ghrist 2007] V. de Silva and R. Ghrist, "Coverage in sensor networks via persistent homology", *Algebr. Geom. Topol.* **7** (2007), 339–358. MR Zbl

[Zomorodian and Carlsson 2005] A. Zomorodian and G. Carlsson, "Computing persistent homology", *Discrete Comput. Geom.* **33**:2 (2005), 249–274. MR Zbl

Received: 2018-06-09 Revised: 2018-09-20 Accepted: 2018-11-29

gcbell@uncg.edu *Department of Mathematics and Statistics, The University of North Carolina at Greensboro, Greensboro, NC, United States*

azlawson@uncg.edu *Department of Mathematics and Statistics, The University of North Carolina at Greensboro, Mount Airy, NC, United States*

jmmart27@uncg.edu *Department of Mathematics and Statistics, The University of North Carolina at Greensboro, Greensboro, NC, United States*

jerudzin@uncg.edu *Department of Mathematics and Statistics, The University of North Carolina at Greensboro, Greensboro, NC, United States*

cdsmyth@uncg.edu *Department of Mathematics and Statistics, The University of North Carolina at Greensboro, Greensboro, NC, United States*

involve

msp.org/involve

INVOLVE YOUR STUDENTS IN RESEARCH

Involve showcases and encourages high-quality mathematical research involving students from all academic levels. The editorial board consists of mathematical scientists committed to nurturing student participation in research. Bridging the gap between the extremes of purely undergraduate research journals and mainstream research journals, *Involve* provides a venue to mathematicians wishing to encourage the creative involvement of students.

See inside back cover or msp.org/involve for submission instructions. The subscription price for 2019 is US $195/year for the electronic version, and $260/year (+$35, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

Involve peer review and production are managed by EditFlow® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

involve
2019 vol. 12 no. 5