Persistent homology for low-complexity models

Lotz, Martin

2017

MIMS EPrint: **2017.30**

ISSN 1749-9097

PERSISTENT HOMOLOGY FOR LOW-COMPLEXITY MODELS

MARTIN LOTZ

ABSTRACT. We show that recent results on randomized dimension reduction schemes that exploit structural properties of data can be applied in the context of persistent homology. In the spirit of compressed sensing, the dimension reduction is determined by the Gaussian width of a structure associated to the data set, rather than its size. The Gaussian width also turns out to be useful for studying the complexity of other methods for approximating persistent homology.

1. INTRODUCTION

Persistent homology is an approach to topological data analysis (TDA) that allows to infer multi-scale qualitative information from noisy data. Starting from a *point cloud* representing the data, persistent homology extracts topological information about the structure from which the data is assumed to be sampled from (such as number of components, holes, cavities, ...) by associating certain multi-scale invariants, the *barcodes* or *persistence diagrams* to the data. These invariants measure topological features of neighbourhoods of the data at different scales; features that *persist* over large scale ranges are considered to be relevant, while short lived features are considered to be topological noise.

Despite excellent theoretical guarantees and plenty of practical applications, the sometimes large number of data points n and ambient dimension d can cause significant challenges for the computation of persistent homology. Much current work in the field is devoted to addressing these challenges, the underlying rationale being that the true complexity of the data is often smaller than it appears, and that a subset of the data can be enough to extract the relevant topological information.

Our focus is on the analysis of randomized dimension reduction schemes at the point cloud level that depend purely on structural properties of the data, and not on the size of the data set. For example, if we assume that the d-dimensional data points are s-sparse in a suitable basis or frame (for example, in a Fourier or wavelet basis), then we can work in an ambient dimension of $O(s\log(d/s))$, similar to what is common in compressed sensing. This reduction affects the very first part of the persistent homology pipeline, where it can reduce the size of the input. Such a reduction is useful in constructions that depend on the ambient dimension, while the independence of the number of data points is useful in settings such as streaming applications, where one does not know the number of samples in advance. In addition, the notion of intrinsic dimension (Gaussian width) used to study the dimension reduction is likely to be of independent interest as a parameter in the analysis of algorithms.

The point of view taken here differs from the usual manifold assumption, where the data is assumed to lie close to a lower dimensional manifold whose topology one is interested in, and where it is the intrinsic dimension of that manifold that determines the complexity of the problem. In our setting, we do not assume anything about such a manifold, but only consider structural properties of any potential data points. These are related to the type of data we consider and can be known or estimated in advance. As an example, consider image data from the Columbia Object Image Library [NNM$^+$96], which contains photos of objects taken from different angles. The images associated to one single object lie on a one-dimensional structure (a circle). Independent of that, the individual images are compressible and as such lie close to a low-dimensional subspace arrangement. It is this latter structure that determines the target dimension of the random dimension reduction, regardless of the shape of the manifold around which the images cluster, and independent of the number of images present in the data set.

Date: 4th September 2017.

A crucial parameter in this context is the Gaussian width of a set,

$$w(S) = \mathbb{E}[\sup_{\boldsymbol{x}\in S}\langle \boldsymbol{x}, \boldsymbol{g}\rangle],$$

where the expectation is over a standard Gaussian vector $\boldsymbol{g}$. This parameter features prominently in the study of Gaussian processes, in geometric functional analysis, and in compressed sensing [FR13, Chapter 9]. We show that it also determines the dimension in which persistent topological information can be recovered without much loss. Formally, this means that the persistence diagrams for the original and for the projected data are close in some metric, which can be formalized using the interleaving distance on persistence modules. For the precise definition of these and other concepts used in the statement of the result, see Section 2.

Theorem 1.1. *Let $X \subset M \subset \mathbb{R}^d$ with X finite, let $\delta \in (0,1)$, and $T = \{(\boldsymbol{x}-\boldsymbol{y})/\|\boldsymbol{x}-\boldsymbol{y}\| : \boldsymbol{x}, \boldsymbol{y} \in M\}$. Assume that*

$$m \geq \frac{\left(w(T) + \sqrt{2\log(2/\delta)}\right)^2}{\varepsilon^2} + 1.$$

Then for a random $m \times d$ matrix $\boldsymbol{G}$ with normal distributed entries $g_{ij} \sim N(\boldsymbol{0}, 1/m)$, with probability at least $1-\delta$, the persistence modules associated to the Čech, Vietoris-Rips, and α complexes of X and $\boldsymbol{G}X$ are multiplicatively $(1-\varepsilon)^{-1}$-interleaved.

Theorem 1.1 is based on, and recovers as special case, an extension of the Johnson-Lindenstrauss Theorem by Sheehy [She14] (see the example with the Gaussian width of discrete sets below). One crucial difference to the classical approach is that the Gaussian width allows us to do better when the data X has a particularly simple structure. Before proceeding, we present some examples of sets where the Gaussian width is well known. We use the notation $w^2(T) := w(T)^2$ for the square of the Gaussian width.

Discrete set. Let $T = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ be a set of n points with $\|\boldsymbol{x}_i\|_2 = 1$ for $1 \leq i \leq n$. Then

$$w^2(T) \leq 2\log(n).$$

A proof can be found in [BLM13, 2.5].

Spheres and balls. Let $T = S^{m-1}$ be an $(m-1)$-dimensional unit sphere in $\mathbb{R}^d$. Then the invariance property of the Gaussian distribution implies

$$w^2(T) = \mathbb{E}[\|\overline{\boldsymbol{g}}\|]^2 \leq m,$$

where $\overline{\boldsymbol{g}} = (g_1, \dots, g_m)^T$ is the projection of a Gaussian vector $\boldsymbol{g} \in \mathbb{R}^d$ to the first m coordinates.

Sparse vectors. Let $T_s = \{\boldsymbol{x} \in S^{d-1} : |\operatorname{supp}(\boldsymbol{x})| \leq s\}$ be the set of s-sparse unit vectors. As shown by Rudelson and Vershynin [RV08], the squared Gaussian width of this set is bounded by

$$w^2(T_s) \leq C \cdot s\log(d/s),$$

where C is some constant. As the Gaussian width is orthogonally invariant, it is enough to require that the elements of $\boldsymbol{x}$ are sparse in some fixed basis. For example, we could have a collection of compressed images that are sparse in a discrete cosine or wavelet basis, or signals that are sparse in a frequency domain.

Low-rank matrices. Let $M_r = \{\boldsymbol{X} \in \mathbb{R}^{d_1\times d_2} : \|\boldsymbol{X}\|_F = 1,\ \operatorname{rk}(\boldsymbol{X}) \leq r\}$ be the set of matrices of rank at most r and unit Frobenius norm, where $\|\boldsymbol{X}\|_F^2 = \sum_{i,j} X_{ij}^2$. It can be shown that

$$w^2(M_r) \leq C \cdot r(d_1 + d_2)$$

for some constant C, see [RFP10, CP11] for a derivation and more background. Examples of low-rank matrices or approximately low-rank matrices abound, including images, Euclidean distance matrices, correlation matrices, matrices arising from the discretisation of differential equations, or recommender systems. One can also consider low-rank tensors (with respect to several notions of rank).

Linear images. Assume that $T = \boldsymbol{A}S$, where $S \subset \mathbb{R}^d$ and $\boldsymbol{A} \in \mathbb{R}^{m\times d}$. Then the squared Gaussian width of T can be bounded in terms of that of S and the condition number $\kappa(\boldsymbol{A})$ of $\boldsymbol{A}$,

$$\kappa^{-2}(\boldsymbol{A})w^2(S) \leq w^2(T) \leq \kappa^2(\boldsymbol{A})w^2(S).$$

See [ALW17] for a derivation of this in a more general context. This is useful when considering the co-sparse setting [NDEG13], in which the signals of interest are sparse after applying some (not necessarily invertible) linear transform.

Convex cones. Let $T = C \cap S^{d-1}$, where C is a convex cone (a convex set with $\lambda \boldsymbol{x} \in C$ if $\boldsymbol{x} \in C$ and $\lambda \geq 0$). The Gaussian width of $C \cap S^{d-1}$ differs from an invariant of the cone, the statistical dimension $\delta(C)$, by at most one [ALMT14, Prop 10.2]. It is known that $\delta(C) = d/2$ for self-dual cones (this includes the orthant and cone of positive semidefinite matrices), $\delta(C) \approx \log(d)$ for $C = \{\boldsymbol{x} : x_1 \leq \cdots \leq x_d\}$, and $\delta(C) \approx d\sin^2(\alpha)$ for the circular cone of radius α [ALMT14, Chapter 3]. Moreover, approximations are known for the squared Gaussian width of the descent cones of the 1-norm [Sto09] and the nuclear norm [CRPW12], see also [ALMT14, Chapter 4].

1.1. **Efficient projections.** In combination with recent results around the Johnson-Lindenstrauss Theorem, Theorem 1.1 extends (up to constants and logarithmic factors) to a large class of linear maps, including subgaussian matrices [Dir16], sparse Johnson-Lindenstrauss transforms [BDN15], and matrices satisfying a classical Restricted Isometry Property [ORS15, KW11]. Strikingly, in [ORS15] the authors derived a “transfer theorem" that shows that one can use so-called RIP (Restricted Isometry Property) matrices with only minor loss. Such matrices have been studied extensively in compressed sensing [FR13], and include the SORS (subsampled orthogonal with ranodm sign) matrices. These are defined as matrices of the form $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{D}$, where $\boldsymbol{H}$ is an $m \times d$ matrix arising from uniformly sampling m rows from a unitary matrix with entries bounded by $\Delta/\sqrt{d}$, and $\boldsymbol{D}$ is a diagonal matrix with random i.i.d. sign pattern on the diagonal. Using the results of [ORS15] instead of Gordon’s Theorem, we get the following variation of Theorem 1.1.

Theorem 1.2. *Let* $\delta \in (0,1)$, $X \subset \mathbb{R}^d$*, and define* $T = \{(\boldsymbol{x}-\boldsymbol{y})/\|\boldsymbol{x}-\boldsymbol{y}\|_2 : \boldsymbol{x},\boldsymbol{y} \in X\}$*. Then for a suitable constant C and*

$$m \geq C \cdot \Delta^2(1+\log(1/\delta))^2 \log^4(d) \frac{w^2(T)}{\varepsilon^2},$$

for a random $m \times d$ *SORS matrix* $\boldsymbol{A}$*, with probability at least* $1-\delta$*, the persistence modules associated to the Čech, Vietoris-Rips, and* α *complexes of* X *and* $\boldsymbol{A}X$ *are multiplicatively* $(1-\varepsilon)^{-1}$*-interleaved.*

As pointed out in [ORS15], it is likely that the term $\log^4(d)$ can be reduced to $\log(d)$. Important examples of SORS matrices with $\Delta = 1$ are partial Fourier matrices and the Hadamard transform, which allow for fast matrix-vector products. The possibility of computing the dimension reduction efficiently is essential to the applicability of the reduction scheme.

1.2. **An application.** Two key advantages of the proposed dimension reduction scheme are that it is non-adaptive, and the fact that the target dimension of the projection does not depend on the size of the data set. Together with the possibility of using fast projections, the method has potential applications in settings where the data set changes with time, and one would like to update topological information as new data becomes available. More precisely, consider a given data set $X = \{\boldsymbol{x}_1,\dots,\boldsymbol{x}_k\}$, and assume that a new point $\boldsymbol{x}_{k+1}$ becomes available. The most basic operation, updating the distance matrix of the point set, requires $O(kd)$ operations. Assume that we have prior information on the type of data represented by X (for example, that it consists of images that have a certain sparsity structure). If we store projections $\boldsymbol{P}\boldsymbol{x}_1,\dots,\boldsymbol{P}\boldsymbol{x}_k$, where $\boldsymbol{P}$ has $m \ll d$ rows, then updating the distance matrix reduces to $O(km)$ operations after computing $\boldsymbol{P}\boldsymbol{x}_{k+1}$. The cost of this reduction is the added complexity of computing the projection $\boldsymbol{P}\boldsymbol{x}_{k+1}$; when $\boldsymbol{P}$ is a sparse Johnson-Lindenstrauss transform with sparsity s, then the number of operations is sd and the total cost of constructing the distance matrix is $O(km+sd)$. When using a partial Fourier or Hadamard matrix, the cost becomes $O(km+d\log d)$. It follows that if the number of points satisfies $k \gg ds/(d-m)$ in the sparse case, and $k \gg d\log d/(d-m)$ in the Fourier or Hadamard case, then significant gains are possible. The setting most likely to benefit is when the dimension is large, the effective dimension (Gaussian width) is small, and the number of samples is sufficiently large.

1.3. **Relation to previous work.** The application of the Johnson-Lindenstrauss Theorem in relation to persistent homology was introduced by Sheehy [She14], on which our approach is based, and independently by Kerber and Raghvendra [KR14]. In particular, a version of the key Theorem 4.1 appeared in [She14]. These articles formulated their results using a target dimension of order $\log(n)/\varepsilon^2$, where n is the cardinality of the point cloud. The work [KR14] also extends the Johnson-Lindenstrauss Lemma to the setting of distances between points and flats, and presents interesting applications using coresets for projective clustering.

The doubling dimension has been used as a measure of intrinsic dimension in topological data analysis [Oud15]; other work using the doubling dimension as intrinsic dimension of the data is [She13, CK14]. The relation of the Gaussian width to the doubling dimension of a metric space was pointed out by Indyk and Naor [IN07]. We revisit this relation in our context, with matching upper and lower bounds, in Section 5.

The use of the Gaussian width in compressed sensing was pioneered by Rudelson and Vershynin [RV08], and has, in combination with Gordon's inequality, come to play a prominent role as a dimension parameter in the development of the theory [FR13]. The Gaussian width also plays an important role in the analysis of signal recovery by convex optimization, as shown by [Sto09] and generalized in [CRPW12], and a variation of the Gaussian width for convex cones, the statistical dimension, determines the location of phase transitions for success probability of such problems [ALMT14]. As far as we are aware, the Gaussian width has not yet been studied in the context of persistent homology.

1.4. **Outline of contents.** In Section 2 we review in some detail the necessary prerequisites from persistent homology. This section also presents the basic interleaving result of Sheehy that links the Johnson-Lindenstrauss Theorem to the interleaving distance of persistence modules. Section 3 reviews the Johnson-Lindenstrauss Theorem in the version of Gaussian matrices and Gaussian width, which is based on Gordon's inequalities for the expected suprema of Gaussian processes. Section 4 presents a version of the Johnson-Lindenstrauss Theorem for smallest enclosing balls, which slightly improves a corresponding result by Sheehy [She14]. A direct consequence of this result is a proof of Theorem 1.1. Section 5 relates the Gaussian width to another intrinsic dimensionality parameter, the doubling dimension, and shows how proofs based on the doubling dimension can be reformulated using the Gaussian width.

1.5. **Notation and conventions.** For a set $S \subset \mathbb{R}^d$, let $\mathrm{enc}(S)$ denote the smallest enclosing ball, $\boldsymbol{c}_S$ its center and $\rho(S)$ its radius. Denote by $\partial\,\mathrm{enc}(S)$ the boundary and by $\partial S = \partial\,\mathrm{enc}(S) \cap S$ the points of S on the boundary.

Acknowledgments. I would like to thank Uli Bauer, Michael Kerber and Don Sheehy for feedback and some useful comments.

2. OVERVIEW OF PERSISTENT HOMOLOGY

The persistence homology pipeline begins with a point cloud, associates to it a filtration of simplicial complexes, constructs a boundary matrix related to the simplicial filtration, and then computes the persistence barcodes from a matrix reduction. We briefly review the part of the theory that is relevant to our purposes. There are many excellent references for the theory presented here, of which we would like to arbitrarily highlight [Car09], [EH10] and [Oud15], the latter being a good reference for both the module-theoretic perspective and as a survey of modern techniques and applications. For an overview of state-of-the-art software, see [OPT$^+$17].

2.1. **Simplicial complexes and homology.** General references for the material in this section are [Mun84, Hat02] or the relevant chapters in [EH10]. A simplicial complex K is a finite collection of sets σ that is closed under the subset relation. The elements $\sigma \in K$ are called simplices, and a subset $\tau \subseteq \sigma$ (itself a simplex) is called a face of σ. The *dimension* of a simplex is $\dim \sigma = |\sigma| - 1$ (in particular, $\dim \emptyset = -1$). We denote by K_p the set of all p-simplices. The p-skeleton $\bigcup_{k\leq p} K_k$ of K is the sub-complex consisting of all simplices of dimension at most p. A map of simplicial complexes $f\colon K \to L$ is a map $f\colon K_0 \to L_0$ such that $f(\sigma) \in L$ for all $\sigma \in K$. A map $f\colon K \to L$ between simplicial complexes is an isomorphism if it is injective, and $\sigma \in K \Leftrightarrow f(\sigma) \in L$. A *subcomplex* of K is a subset $L \subseteq K$ that is itself a simplicial complex. We can associate to each p-simplex in K a geometric simplex σ in some $\mathbb{R}^d$ (that is, the convex hull of $p + 1$ affinely independent points) in such a way that the face relations remain valid and the intersection of two simplices is either empty or again a simplex. The union of these geometric simplices is a *geometric realization* $|K|$ of the simplicial complex. A map between simplicial complexes $K \to L$ gives rise to a continuous map $|K| \to |L|$ between topological spaces. An important result in algebraic topology states that isomorphic simplicial complexes give rise to homeomorphic realizations. In particular, the homotopy type (loosely speaking, the class of shapes that a set can be continuously deformed into) of a simplicial complex is well defined as the homotopy type of a realization of the complex.

One way to obtain a simplicial complex is as the *nerve* of a cover. Given a set $S \subset \mathbb{R}^d$ and a set of subsets $\mathcal{S} = \{U_i\}_{i\in I}$ such that $S \subseteq \bigcup_{i\in I} U_i$, we define the nerve $\mathcal{N}(\mathcal{S})$ of the cover to be the simplicial complex on the set I defined by

$$\sigma \in \mathcal{N}(\mathcal{S}) \Leftrightarrow \bigcap_{i\in\sigma} U_i \neq \mathbf{0}.$$

The following important result (see, for example, [Hat02, 4.G]) relates the topology of the nerve to that of the cover. The Nerve Theorem also holds for covers with closed balls in Euclidean space [Oud15, Chapter 4.3], the setting in which it will be used in our case.

Theorem 2.1. *(Nerve Theorem) Let $\mathcal{S}$ be a finite collection of open, convex sets in $\mathbb{R}^d$. Then $\mathcal{N}(\mathcal{S})$ is homotopy equivalent to the union $\bigcup_{U\in\mathcal{S}} U$.*

A chain complex $C_p(K)$ is the $\mathbb{F}_2$-vector space generated by the p-simplicies of K. The boundary map maps a p-simplex to the sum of its $(p-1)$-dimensional faces,

$$\partial_p\colon C_p(K) \to C_{p-1}(K), \quad \sigma \mapsto \sum_{\tau\subset\sigma\cap K_{p-1}} \tau.$$

The boundary maps satisfy the fundamental property that for $p \geq 0$, $\partial_p \circ \partial_{p+1} = 0$ (here, we use the convention that $\partial_0 = 0$). If we set $Z_p(K) = \ker \partial_p$ (the set of *cycles*) and $B_p(K) = \operatorname{im} \partial_{p+1}$ (the set of *boundaries*), then the p-th homology vector space is defined as the quotient

$$H_p(K) = Z_p(K)/B_p(K).$$

The p-th Betti number is $\beta_p = \dim H_p(K)$. Homology is functorial, meaning that a map $f\colon K \to L$ induces a morphism $f_*\colon H_p(K) \to H_p(L)$, with the property that an isomorphism of complexes maps to an isomorphism of homology groups.

2.2. **Filtrations.** To capture the topology of the data at different scales, we need to consider sequences of topological spaces and simplicial complexes ordered by inclusion. Such filtrations of topological spaces and of simplicial complexes both give rise to filtrations of homology vector spaces.

Given a metric space (M, d), a continuous function $f\colon M \to \mathbb{R}$ induces a sublevel filtration $\mathcal{M} = \{M_\alpha\}_{\alpha\in\mathbb{R}}$, defined as

$$M_\alpha = \{\boldsymbol{x} \in M : f(\boldsymbol{x}) \leq \alpha\}.$$

The case of interest in topological data analysis is when $M = \mathbb{R}^d$ with the Euclidean distance, X is a finite set $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ (a *point cloud*) and $f = d_X$ is the distance function to X, $d_X \colon X \to \mathbb{R}$, $d_X(\boldsymbol{x}) = \inf_{\boldsymbol{p} \in X} \|\boldsymbol{x} - \boldsymbol{p}\|_2$. The induced filtration consists of the union of closed α-balls around the points of X,

$$X_\alpha = \bigcup_{1 \leq i \leq n} B(\boldsymbol{x}_i, \alpha).$$

A sequence of simplicial subcomplexes

$$\mathcal{K} \colon \emptyset = K^{t_0} \subset K^{t_1} \subset \cdots \subset K^{t_s}$$

also gives rise to a sequence of homology groups

$$H_p(K^{t_0}) \xrightarrow{\iota_0^1} H_p(K^{t_1}) \xrightarrow{\iota_1^2} \cdots \xrightarrow{\iota_{s-1}^s} H_p(K^{t_s}),$$

where ι_i^{i+1} denotes the map induced by the inclusion $K^{t_i} \subset K^{t_{i+1}}$.

The are various ways to associate a simplicial filtration to a topological one, and the resulting homology sequences may or may not be the same. In the following, let $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\} \subset \mathbb{R}^d$ be a point cloud. The nerve of the cover $X_\alpha = d_X((-\infty, \alpha])$ is the Čech filtration,

$$\check{C}_\alpha(X) = \{\sigma \subset [n] : \bigcap_{i \in \sigma} B(\boldsymbol{x}_i, \alpha) \neq \emptyset\}.$$

In other words, a subset $S = \{\boldsymbol{x}_i : i \in \sigma\} \subseteq X$ gives rise to a simplex $\sigma \in \check{C}_\alpha(X)$ if and only if $S \subset B(\boldsymbol{x}, \alpha)$ for some $\boldsymbol{x} \in \mathbb{R}^d$.

The *Delauney* filtration, or α-filtration, is the sequence of simplicial complexes $D_\alpha(X)$ consisting of simplices $\sigma \subset [n]$ such that there exists $\boldsymbol{x} \in \mathbb{R}^d$ with

- $S = \{\boldsymbol{x}_i : i \in \sigma\} \subset B(\boldsymbol{x}, \alpha)$;
- for all $\boldsymbol{p} \notin S$, $\boldsymbol{p} \notin B(\boldsymbol{x}, \alpha)$.

Clearly, $D_\alpha(X) \subseteq \check{C}_\alpha(X)$. The α filtration has the advantage that if the points in X are in general position (meaning that no $d+2$ of them lie on the surface of a sphere), then the simplices have dimension at most d, whereas for the Čech complex they can have dimension up to n. On the other hand, while the complexity of constructing a Čech complex only depends on the size of the data set (or the distance matrix), constructing an α complex has complexity exponential in the ambient dimension d, which makes it practical only for small dimensions.

Finally, one of the most common constructions is the *Vietoris-Rips* complex, $V_\alpha(X)$, where $\sigma \in V_\alpha(X)$ if and only if for all $\boldsymbol{x}, \boldsymbol{y} \in \sigma$, $\|\boldsymbol{x} - \boldsymbol{y}\| \leq 2\alpha$.

A filtration $\mathcal{M}$ gives rise to a filtration in the homology vector spaces: if $\alpha \leq \alpha'$, then we get an induced homomorphism

$$H_p(M_\alpha) \xrightarrow{\iota_\alpha^{\alpha'}} H_p(M_{\alpha'}).$$

The p-th *persistent homology* $H_p(\mathcal{K})$ associated to a (topological or simplicial) filtration is the induced sequence of homology vector spaces and linear maps, and the p-th persistence vector spaces are the images of these homomorphisms,

$$H_p^{\alpha,\alpha'}(\mathcal{M}) = \operatorname{im} \iota_\alpha^{\alpha'} = Z_p(M_\alpha)/(B_p(M_{\alpha'}) \cap Z_p(M_\alpha)).$$

In all the situations of interest to us in this paper, the homology groups are finite-dimensional, and there are only finitely many indices $c_0 < c_1 < \cdots < c_m$, the *critical points*, such that $H_p(K_\alpha) = H_p(K_{\alpha'})$ and $\iota_\alpha^{\alpha'} = \mathrm{id}$ for $\alpha, \alpha' \in [c_i, c_{i+1})$.

While the Nerve Theorem 2.1 guarantees that for each α the homology of the nerve complex is the same as the homology of the cover, it does not automatically follow that the persistent homology of the filtration induced by the cover is the same as the persistent homology of the resulting filtered simplicial complex.

The fact that this is the case is guaranteed by the Persistent Nerve Lemma [Oud15, Lemma 4.12]. The following is a simplified statement. To avoid too many parentheses, we use the notation $\check{C}_\alpha X := \check{C}_\alpha(X)$.

Lemma 2.2. *(Persistent Nerve Lemma) Let* $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ *and* X_α *the cover with closed balls of radius* α*. Let* $\check{C}_\alpha X$ *be the corresponding nerve. Then there is an isomorphism of persistence homology modules* H_p*. Specifically, for every* α *there are isomorphisms* ν_α *such that the following diagram commutes:*

$$\begin{tikzcd}
H_p(X_\alpha) \arrow[r, "\iota_\alpha^{\alpha'}"] \arrow[d, "\nu_\alpha"'] & H_p(X_{\alpha'}) \arrow[d, "\nu'_\alpha"] \\
H_p(\check{C}_\alpha X) \arrow[r, "\iota_\alpha^{\alpha'}"] & H_p(\check{C}_{\alpha'} X)
\end{tikzcd}$$

The Delauney complex can also be related to the offset filtration complex [Oud15, Chapter 3], by considering a deformation retraction of X_α to a geometric realization of D_α, which induces isomorphisms between the homology sequences,

$$\begin{tikzcd}
H_p(X_\alpha) \arrow[r, "\iota_\alpha^{\alpha'}"] \arrow[d, "\nu_\alpha"'] & H_p(X_{\alpha'}) \arrow[d, "\nu'_\alpha"] \\
H_p(D_\alpha X) \arrow[r, "\iota_\alpha^{\alpha'}"] & H_p(D_{\alpha'} X)
\end{tikzcd}$$

While the homology of the Vietoris-Rips complex is not as directly related to the topological filtration via the Nerve Lemma, it approximates the Čech filtration via the observation

$$\check{C}_\alpha(X) \subseteq V_\alpha(X) \subseteq \check{C}_{\sqrt{2}\alpha}(X).$$

For a simplicial filtration one defines the persistent homology vector spaces $H_p^{i,j}(\mathcal{K})$ just as in the case of a topological filtration. A simplex σ is born at time t_i if appears in $H_p(K^{t_i})$ but is not the image of an element in $H_p(K^{t_{i-1}})$. A simplex σ dies at time t_i if $\iota_{t_{i-1}}^{t_i}(\sigma) = 0$. This way each element in the filtration, which corresponds to a topological feature of the realisation of the simplicial complex, comes with an interval $[\alpha, \alpha')$ representing its lifetime, where α' may be ∞. The lifetimes of the various features are recorded in a two-dimensional *persistence diagram*, where each interval $[a, b)$ is represented by a point with coordinates (a, b) with multiplicity (which equals 0 if there is no element whose lifetime matches the interval). Alternatively, one can represent each interval ocurring using *persistence barcodes*, which record the lifetime of each feature as an interval.

2.3. **Persistence modules.** Homology vector spaces associated to a filtration (either simplicial or offset filtration) are examples of *persistence modules*. A persistence module is a sequence of vector spaces $\{U_\alpha\}_{\alpha \in T}$ over a field $\mathbb{F}$, indexed by a partially ordered set T, together with linear maps $u_\alpha^{\alpha'} \colon U_\alpha \to U_{\alpha'}$ whenever $\alpha \leq \alpha'$, and such that $u_\alpha^\alpha = \mathrm{id}_{V_\alpha}$ and $u_{\alpha'}^{\alpha''} \circ u_\alpha^{\alpha'} = u_\alpha^{\alpha''}$ whenever $\alpha \leq \alpha' \leq \alpha''$ (in categorical terms, a persistence module is a functor from a poset to the category of vector spaces over a field). The *direct sum* of persistent modules $\mathbb{U} = \{U_\alpha\}_{\alpha \in T}$ and $\mathbb{V} = \{V_\alpha\}_{\alpha \in T}$, written $\mathbb{W} = \mathbb{U} \oplus \mathbb{V}$, consists of the vector spaces $U_\alpha \oplus V_\alpha$, together with the homomorphisms $u_\alpha^{\alpha'} \oplus v_\alpha^{\alpha'}$. A morphism of persistence modules $\mathbb{U}$ and $\mathbb{V}$ over the same poset is a natural transformation between the persistence modules, i.e., a collection of linear maps $\varphi_\alpha \colon U_\alpha \to V_\alpha$ such that the diagram

$$\begin{tikzcd}
U_\alpha \arrow[r, "u_\alpha^{\alpha'}"] \arrow[d, "\varphi_\alpha"] & U_{\alpha'} \arrow[d, "\varphi_{\alpha'}"] \\
V_\alpha \arrow[r, "v_\alpha^{\alpha'}"] & V_{\alpha'}
\end{tikzcd}$$

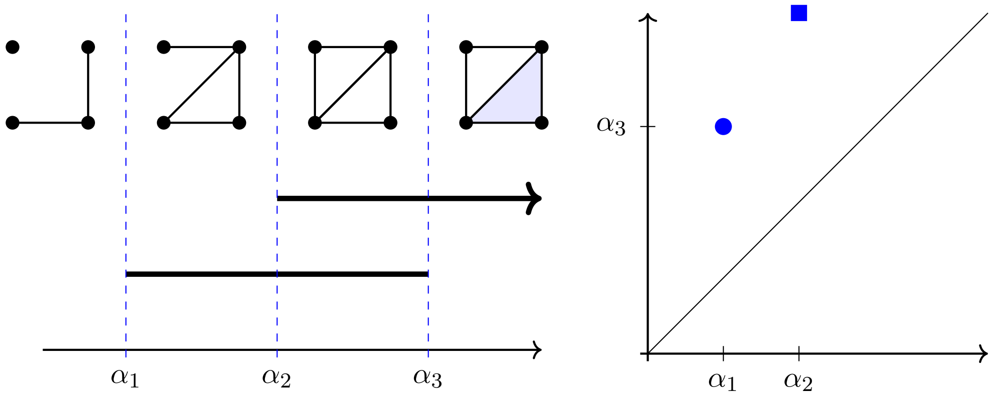

FIGURE 1. Persistence barcode and persistence diagram for the first Betti number of a simplicial filtration. The circle represents a feature that is born at α_1 and dies at α_3, while the square represents a feature that is born at α_2 and lives on forever.

commutes for all $\alpha \leq \alpha' \in T$.

An *interval module* $\mathbb{F}_I$ with interval $I = [a, b) \subset T$ is a persistence module with $U_\alpha = \mathbb{F}$ for $\alpha \in I$ and $u_\alpha^{\alpha'} = \mathrm{id}_{\mathbb{F}}$ for $\alpha, \alpha' \in I$, and $U_\alpha = \mathbf{0}$, $u_\alpha^{\alpha'} = 0$ else. We allow $a = -\infty$ or $b = \infty$. If the vector spaces are all finite-dimensional and/or the index set T is finite or countable, then it follows from classic results in quiver representation theory (see [Oud15, Theorem 1.9]) that a persistence module can be written as a direct sum of interval modules,

$$\mathbb{U} = \bigoplus_{I \in \Gamma} \mathbb{F}_I,$$

where each Γ consists of intervals $[\alpha, \alpha')$ with $\alpha \leq \alpha'$ and $\alpha, \alpha' \in T$. The set of intervals Γ, counted with multiplicity, is the *persistence barcode* of the persistence module. Alternatively, the intervals can be represented by a *persistence diagram* $\mathrm{Dgm}(\mathbb{U})$, which maps each interval $[a, b)$ to a point on the plane $\overline{R}^2 = (\mathbb{R} \cup \{-\infty, +\infty\})^2$, with associated multiplicity.

2.4. **Stability.** One of the strengths of persistent homology is its robustness with respect to noise. In order to measure how changes in the input affect changes in the persistent diagrams, we need to define a notion of distance between persistence diagrams (or between persistence modules). We only discuss the distance in the context of modules, and point out that by the *Isometry Theorem* [CDSGO16] (and its variants) we can relate distances of persistence modules to notions of distance on persistence diagrams, such as the (logarithmic) bottleneck distance.

Given a persistence module $\mathbb{U}$ and $c \geq 1$, define a shifted module $\mathbb{U}[c]$ such that $U[c]_\alpha = U_{c\alpha}$ for each $\alpha \in T$. A *multiplicative c-interleaving* between a pair of persistence modules $\mathbb{U}, \mathbb{V}$ is a pair of morphism of persistence modules $\varphi\colon \mathbb{U} \to \mathbb{V}[c]$ and $\psi\colon \mathbb{V} \to \mathbb{U}[c]$ such that for each α, the following diagrams commute.

$$\begin{tikzcd}
U_{\alpha/c} \arrow[rr, "u^{c\alpha}_{\alpha/c}"] \arrow[dr, "\varphi_{\alpha/c}"'] & & U_{c\alpha} \\
& V_\alpha \arrow[ur, "\psi_\alpha"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
& U_\alpha \arrow[dr, "\varphi_\alpha"] & \\
V_{\alpha/c} \arrow[ur, "\psi_{\alpha/c}"] \arrow[rr, "v^{c\alpha}_{\alpha/c}"] & & V_{c\alpha}
\end{tikzcd}$$

The (multiplicative) *interleaving distance* between two persistence modules $\mathbb{U}$ and $\mathbb{V}$ is the smallest ε such that an additive ε-interleaving exists, and the multiplicative interleaving distance is the smallest c such that a multiplicative c-interleaving exists. One similarly defines an additive interleaving by replacing α/c and $c\alpha$ with $\alpha - \varepsilon$ and $\alpha + \varepsilon$. Note that a multiplicative interleaving is an additive interleaving on a logarithmic scale.

The following Lemma from [She14] reduces the task of finding an interleaving of persistence modules to that of establishing inequalities for smallest enclosing balls. Recall the notation $\rho(S)$ for the radius of a smallest enclosing ball of S.

Lemma 2.3. *Let $X \subset \mathbb{R}^d$ be a finite set and $d_X \colon \mathbb{R}^d \to \mathbb{R}$ the distance function. Assume we have a function $F \colon \mathbb{R}^d \to \mathbb{R}^m$ such that for all subsets $S \subseteq X$ we have*

$$(1-\varepsilon)\rho(S) \leq \rho(F(S)) \leq (1+\varepsilon)\rho(S). \tag{2.1}$$

Then the persistent homology modules associated to the Čech and α (Delauney) filtrations of X and $F(X)$ are multiplicatively $(1-\varepsilon)^{-1}$-interleaved.

Proof. We deal with the Čech-complex. The statement for the α-filtration follows from a standard equivalence [Oud15, Chapter 4]. Recall that a set S defines a simplex $\sigma \in \check{C}_\alpha(X)$ if and only if $\rho(S) \leq \alpha$. Set $c = (1-\varepsilon)^{-1}$. Let $S \subseteq X$ such that $\rho(S) \leq \alpha$, and let $\sigma \in \check{C}_\alpha(X)$ be the associated simplex in the Čech-complex. Then by assumption

$$\rho(F(S)) \leq (1+\varepsilon)\rho(S) \leq c\alpha,$$

so that $F(S)$ gives rise to a simplex $\varphi_\alpha(\sigma) \in \check{C}_{c\alpha}(F(X))$. Conversely, if S' gives rise to a simplex $\sigma \in \check{C}_\alpha(F(X))$ and $S = F^{-1}(S')$, then

$$(1-\varepsilon)\rho(S) \leq \rho(S') \leq \alpha,$$

so that S gives rise to a simplex $\psi_\alpha(\sigma) \in \check{C}_{c\alpha}(X)$. One easily checks that for $\alpha \leq \alpha'$, the following diagrams commute:

$$\begin{tikzcd}
\check{C}_\alpha(X) \arrow[r, "\iota_\alpha^{\alpha'}"] \arrow[d, "\varphi_\alpha"] & \check{C}_{\alpha'}(X) \arrow[d, "\varphi_{\alpha'}"] \\
\check{C}_{c\alpha}(F(X)) \arrow[r, "\iota_{c\alpha}^{c\alpha'}"] & \check{C}_{c\alpha'}(F(X))
\end{tikzcd}
\qquad
\begin{tikzcd}
\check{C}_\alpha(F(X)) \arrow[r, "\iota_\alpha^{\alpha'}"] \arrow[d, "\psi_\alpha"] & \check{C}_{\alpha'}(F(X)) \arrow[d, "\psi_{\alpha'}"] \\
\check{C}_{c\alpha}(X) \arrow[r, "\iota_{c\alpha}^{c\alpha'}"] & \check{C}_{c\alpha'}(X)
\end{tikzcd}$$

The maps φ_α and ψ_α therefore give rise to maps between the filtrations. By the Persistent Nerve Lemma [Oud15, Lemma 4.12], these maps induce maps of persistent homology modules for each p,

$$\begin{tikzcd}
H_p(X_\alpha) \arrow[r, "\iota_\alpha^{\alpha'}"] \arrow[d, "\varphi_\alpha"] & H(X_{\alpha'}) \arrow[d, "\psi_{\alpha'}"] \\
H_p(F(X)_{c\alpha}) \arrow[r, "\iota_{c\alpha}^{c\alpha'}"] & H_p(F(X)_{c\alpha'})
\end{tikzcd}
\qquad
\begin{tikzcd}
H_p(F(X)_\alpha) \arrow[r, "\iota_\alpha^{\alpha'}"] \arrow[d, "\psi_\alpha"] & H_p(F(X)_{\alpha'}) \arrow[d, "\varphi_{\alpha'}"] \\
H_p(X_{c\alpha}) \arrow[r, "\iota_{c\alpha}^{c\alpha'}"] & H_p(X_{c\alpha'})
\end{tikzcd}$$

Moreover, since for any $S \subset X$ we have $S \subseteq F^{-1}(F(S))$, the following identities hold:

$$\psi_\alpha \circ \varphi_{\alpha/c} = \iota_{\alpha/c}^{c\alpha}, \quad \varphi_\alpha \circ \psi_{\alpha/c} = \iota_{\alpha/c}^{c\alpha}.$$

It follows that these maps constitute a c-interleaving, that carries over to a c-interleaving of the persistent homology modules. □

3. GENERAL JOHNSON-LINDENSTRAUSS TRANSFORMS

The classical Johnson-Lindenstrauss Theorem [Mat02, 15.2] shows the existence of a linear map $f \colon \mathbb{R}^d \to \mathbb{R}^m$, such that for all $\boldsymbol{x}, \boldsymbol{y}$ from a finite set $X \subset \mathbb{R}^d$ with $|X| = n$,

$$(1-\varepsilon)\|\boldsymbol{x}-\boldsymbol{y}\| \leq \|f(\boldsymbol{x}) - f(\boldsymbol{y})\| \leq (1+\varepsilon)\|\boldsymbol{x}-\boldsymbol{y}\|,$$

provided $m \geq C \cdot \log(n)/\varepsilon^2$ for some constant C.

This bound is sharp in general [LN14], but it can be refined based on a certain geometric measure of a set related to X. Arguably the most common geometric measure used in this context is the *Gaussian width* of a set T, defined as

$$w(T) = \mathbb{E} \sup_{\boldsymbol{x} \in T} \langle \boldsymbol{g}, \boldsymbol{x} \rangle,$$

where the expectation is taken over a random Gaussian vector in $\mathbb{R}^d$, i.e., $\boldsymbol{g} \in N(\mathbf{0}, \mathbf{1})$. One version of the Johnson-Lindenstrauss Theorem can be stated as follows. In what follows we set $E_m := \mathbb{E}[\|\boldsymbol{g}\|] = \sqrt{2}\Gamma((m+1)/2)/\Gamma(m/2)$, and note that

$$\frac{m}{\sqrt{m+1}} \leq E_m \leq \sqrt{m}.$$

Theorem 3.1. *(Johnson-Lindenstrauss - Gordon version) Let* $\delta \in (0,1)$*,* $X \subset \mathbb{R}^d$*, and define* $T = \{(\boldsymbol{x}-\boldsymbol{y})/\|\boldsymbol{x}-\boldsymbol{y}\|_2 : \boldsymbol{x}, \boldsymbol{y} \in X\}$*. Assume that*

$$m \geq \frac{\left(w(T) + \sqrt{2\log(2/\delta)}\right)^2}{\varepsilon^2} + 1.$$

Then for a random Gaussian $m \times d$ *matrix* $\boldsymbol{G}$*, with entries* $g_{ij} \sim N(\mathbf{0}, 1/E_m^2)$*, we have*

$$(1-\varepsilon)\|\boldsymbol{x}-\boldsymbol{y}\| \leq \|\boldsymbol{G}\boldsymbol{x}-\boldsymbol{G}\boldsymbol{y}\| \leq (1+\varepsilon)\|\boldsymbol{x}-\boldsymbol{y}\|,$$

uniformly for all $\boldsymbol{x}, \boldsymbol{y} \in X$ *with probability at least* $1-\delta$*.*

Remark 3.2. As mentioned after Theorem 1.1, one can generalize this result with minor loss to subgaussian transformations [Dir16], to the setting of the Sparse Johnson Lindenstrauss Transform (SJLT) [BDN15] or more general so-called RIP-matrices [ORS15, KW11], that include, for example, partial Fourier or discrete cosine transforms. We present one such result, which allows to derive Theorem 1.2. Recall that a SORS (subsampled orthogonal with ranodm sign) matrix is defined as a matrix of the form $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{D}$, where $\boldsymbol{H}$ is an $m \times d$ matrix arising from uniformly sampling m rows from a unitary matrix with entries bounded by $\Delta/\sqrt{d}$, and $\boldsymbol{D}$ is a diagonal matrix with random i.i.d. sign pattern on the diagonal.

Theorem 3.3. *(*[ORS15, Theorem 3.3]*) Let* $\delta \in (0,1)$*,* $X \subset \mathbb{R}^d$*, and define* $T = \{(\boldsymbol{x}-\boldsymbol{y})/\|\boldsymbol{x}-\boldsymbol{y}\|_2 : \boldsymbol{x}, \boldsymbol{y} \in X\}$*. Let* $\boldsymbol{A} \in \mathbb{R}^{m\times d}$ *be a SORS matrix. Then for some constant* C *and*

$$m \geq C \cdot \Delta^2 (1+\log(1/\delta))^2 \log^4(d) \frac{w^2(T)}{\varepsilon^2},$$

the matrix $\boldsymbol{A}$ *satisfies*

$$(1-\varepsilon)\|\boldsymbol{x}-\boldsymbol{y}\| \leq \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{A}\boldsymbol{y}\| \leq (1+\varepsilon)\|\boldsymbol{x}-\boldsymbol{y}\|,$$

uniformly for all $\boldsymbol{x}, \boldsymbol{y} \in X$ *with probability at least* $1-\delta$*.*

Keeping in mind that the results presented here also hold in practically relevant settings, we nevertheless restrict the remaining discussion to the Gaussian case to keep the exposition conceptually simple.

The proof of Theorem 3.1 is a well known and direct application of Theorem 3.4, which follows from an inequality of Gordon [Gor88] relating the expected suprema of Gaussian processes, together with concentration of measure for Lipschitz functions. We include the proof for convenience, an accessible derivation of Gordon's inequality itself can be found in the follow-up to [FR13, Theorem 9.21].

Theorem 3.4. *(Gordon) Let* $\boldsymbol{G} \sim N(\mathbf{0}, \mathbf{1})$ *be a Gaussian* $m \times d$ *matrix and let* $T \subseteq S^{d-1}$ *be a subset of the unit sphere. Then*

$$\mathbb{P}\{\min_{\boldsymbol{x}\in T} \|\boldsymbol{G}\boldsymbol{x}\| \leq E_m - w(T) - t\} \leq e^{-t^2/2}$$

$$\mathbb{P}\{\max_{\boldsymbol{x}\in T} \|\boldsymbol{G}\boldsymbol{x}\| \geq E_m + w(T) + t\} \leq e^{-t^2/2}$$

Proof of Theorem 3.1. Set $T = \{(\boldsymbol{x}-\boldsymbol{y})/\|\boldsymbol{x}-\boldsymbol{y}\|_2 : \boldsymbol{x}, \boldsymbol{y} \in X\}$. Then the claim is that with probability at least $1-\delta$, for all $\boldsymbol{p} \in T$,

$$E_m(1-\varepsilon) \leq \|\tilde{\boldsymbol{G}}\boldsymbol{p}\|_2 \leq E_m(1+\varepsilon),$$

where we used the assumption that $\boldsymbol{G} = \frac{1}{E_m}\tilde{\boldsymbol{G}}$ for a standard Gaussian matrix $\tilde{\boldsymbol{G}}$. By the union bound, it suffices to show that

$$\mathbb{P}\{\min_{\boldsymbol{x}\in T} \|\tilde{\boldsymbol{G}}\boldsymbol{x}\| \leq E_m(1-\varepsilon)\} \leq \frac{\delta}{2}$$
$$\mathbb{P}\{\max_{\boldsymbol{x}\in T} \|\tilde{\boldsymbol{G}}\boldsymbol{x}\| \geq E_m(1+\varepsilon)\} \leq \frac{\delta}{2}$$

This is where Gordon's Theorem 3.4 comes into the picture. Set $t = \sqrt{2\log(2/\delta)}$, so that $\delta/2 = e^{-t^2/2}$. The relation between m, ε and δ in the statement of the theorem can be reformulated as

$$w(T) + t \leq \varepsilon\sqrt{m-1} \leq \varepsilon E_m,$$

and including this into the inequalities in Theorem 3.4 finishes the proof. ☐

4. ENCLOSING BALLS

From Lemma 2.3 it follows that what is needed is a version of the Johnson-Lindenstrauss Theorem involving smallest enclosing balls. The following Theorem was independently derived by Sheehy [She17] and improves an earlier version from [She14]. Recall the notation $\rho(S)$ for the radius of the smallest enclosing ball of S.

Theorem 4.1. *Let $S \subset \mathbb{R}^d$ be a finite set and let $\varepsilon \in [0,1)$. Assume that for a map $f\colon \mathbb{R}^d \to \mathbb{R}^m$ and for all $\boldsymbol{x}, \boldsymbol{y} \in S$ we have*

$$(1-\varepsilon)\|\boldsymbol{x}-\boldsymbol{y}\| \leq \|f(\boldsymbol{x}) - f(\boldsymbol{y})\| \leq (1+\varepsilon)\|\boldsymbol{x}-\boldsymbol{y}\|. \tag{4.1}$$

Then

$$(1-\varepsilon)\rho(S) \leq \rho(f(S)) \leq (1+\varepsilon)\rho(S). \tag{4.2}$$

Remark 4.2. The literature on Johnson-Lindenstrauss is not always consistent on whether to use norms or squared norms, which leads to some ambiguity with respect to ε and ε^2. We note that if we had used squared norms in the assumptions of Theorem 4.1, we would also gets the same result.

To prepare for the proof of Theorem 4.1 we first need a few elementary auxiliary results. Lemma 4.3 appears to be folklore. For a set S, the *center* $\boldsymbol{c}_S$ of S is the center of the smallest including ball.

Lemma 4.3. *Let S be a set and let $\boldsymbol{c}_S$ denote the center of S. Then $\boldsymbol{c}_S \in \operatorname{conv}(S)$.*

Proof. Assume $\boldsymbol{c}_S \notin \operatorname{conv}(S)$ and denote by $\boldsymbol{\Pi}(\boldsymbol{c}_S) = \arg\min_{\boldsymbol{x}\in\operatorname{conv}(S)} \|\boldsymbol{c}_S - \boldsymbol{x}\|$ the projection of $\boldsymbol{c}_S$ onto $\operatorname{conv}(S)$. We show that any point in S is closer to $\boldsymbol{\Pi}(\boldsymbol{c}_S)$ than to $\boldsymbol{c}_S$. In fact, for any $\boldsymbol{p} \in S$ we get

$$\begin{aligned}\|\boldsymbol{c}_S - \boldsymbol{p}\|^2 &= \|\boldsymbol{c}_S - \boldsymbol{\Pi}(\boldsymbol{c}_S) + \boldsymbol{\Pi}(\boldsymbol{c}_S) - \boldsymbol{p}\|^2 \\ &= \|\boldsymbol{c}_P - \boldsymbol{\Pi}(\boldsymbol{c}_S)\|^2 + \|\boldsymbol{\Pi}(\boldsymbol{c}_S) - \boldsymbol{p}\|^2 - 2\langle \boldsymbol{c}_S - \boldsymbol{\Pi}(\boldsymbol{c}_S), \boldsymbol{p} - \boldsymbol{\Pi}(\boldsymbol{c}_S)\rangle \\ &\geq \|\boldsymbol{\Pi}(\boldsymbol{c}_S) - \boldsymbol{p}\|^2,\end{aligned}$$

where we used the fact that for a convex set, the inner product of an inward pointing and an outward pointing vector from the boundary is non-positive. ☐

The following elementary observation is just the expression of the sample variance of a point set in terms of pairwise distances.

Lemma 4.4. *Let $\boldsymbol{c} = \sum_{i=1}^k \lambda_i \boldsymbol{x}_i$ be a convex combination of elements of S. Then*

$$\sum_{i=1}^k \lambda_i \|\boldsymbol{x}_i - \boldsymbol{c}\|^2 = \sum_{i<j} \lambda_i\lambda_j \|\boldsymbol{x}_j - \boldsymbol{x}_i\|^2. \tag{4.3}$$

Proof. Using the representation of $\boldsymbol{c}$ as convex combination of the $\boldsymbol{x}_i$, we get

$$\|\boldsymbol{x}_j-\boldsymbol{c}\|^2=\langle\boldsymbol{x}_j-\boldsymbol{c},\boldsymbol{x}_j-\sum_{i=1}^k\lambda_i\boldsymbol{x}_i\rangle=\sum_{i=1}^k\lambda_i\langle\boldsymbol{x}_j-\boldsymbol{c},\boldsymbol{x}_j-\boldsymbol{x}_i\rangle. \tag{4.4}$$

Each summand can be characterized as

$$\langle\boldsymbol{x}_j-\boldsymbol{c},\boldsymbol{x}_j-\boldsymbol{x}_i\rangle=\frac{1}{2}\left(\|\boldsymbol{x}_j-\boldsymbol{c}\|^2+\|\boldsymbol{x}_j-\boldsymbol{x}_i\|^2-\|\boldsymbol{x}_i-\boldsymbol{c}\|^2\right).$$

Plugging this identity into (4.4), using $\sum_{i=1}^k\lambda_i=1$, and combining all terms involving $\|\boldsymbol{x}_j-\boldsymbol{c}\|^2$,

$$\|\boldsymbol{x}_j-\boldsymbol{c}\|^2=\sum_{i=1}^k\lambda_i\|\boldsymbol{x}_j-\boldsymbol{x}_i\|^2-\sum_{i=1}^k\lambda_i\|\boldsymbol{x}_i-\boldsymbol{c}\|^2.$$

Rearranging and summing both sides with weights λ_j establishes the claim. □

Proof of Theorem 4.1. Let $\partial S=\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_k\}$ and let $\boldsymbol{c}_S=\sum_{i=1}^k\lambda_i\boldsymbol{x}_i$ be a convex combination of the center of S. Set $\tilde{\boldsymbol{c}}=\sum_{i=1}^k\lambda_i f(\boldsymbol{x}_i)$. Applying Lemma 4.4 twice, we get

$$\begin{aligned}\rho^2(S)=\sum_{i=1}^k\lambda_i\|\boldsymbol{x}_i-\boldsymbol{c}_S\|^2 &\overset{(4.3)}{=}\sum_{i<j}\lambda_i\lambda_j\|\boldsymbol{x}_j-\boldsymbol{x}_i\|^2\\ &\leq\frac{1}{(1-\varepsilon)^2}\left(\sum_{i<j}\lambda_i\lambda_j\|f(\boldsymbol{x}_j)-f(\boldsymbol{x}_i)\|^2\right)\\ &\overset{(4.3)}{=}\frac{1}{(1-\varepsilon)^2}\left(\sum_{i=1}^k\lambda_i\|f(\boldsymbol{x}_i)-\tilde{\boldsymbol{c}}\|^2\right).\end{aligned}$$

As the function $\boldsymbol{c}\mapsto\sum_{i=1}^k\lambda_i\|f(\boldsymbol{x}_i)-\boldsymbol{c}\|^2$ is minimized at $\tilde{\boldsymbol{c}}$, we can continue the above and conclude

$$\rho^2(S)\leq\frac{1}{(1-\varepsilon)^2}\left(\sum_{i=1}^k\lambda_i\|f(\boldsymbol{x}_i)-\boldsymbol{c}_{f(S)}\|^2\right)\leq\frac{1}{(1-\varepsilon)^2}\rho^2(f(S)).$$

For the right-hand inequality we proceed similarly. □

Proof of Theorem 1.1. The Johnson-Lindenstrauss Theorem 3.1 states that under the assumptions of Theorem 1.1, the pairwise distances are preserved up to multiplicative factors of $1\pm\varepsilon$. Theorem 4.1 states that this implies that smallest enclosing balls are preserved up to the same factors. Finally, Lemma 2.3 asserts that this gives rise to the desired interleaving of persistent modules. This completes the proof. □

Remark 4.5. It was pointed out [Bau17] that Theorem 4.1 also follows from the Kirszbraun intersection property [Gro87].

5. GAUSSIAN WIDTH, ENTROPY, AND DOUBLING DIMENSION

A common notion in the study of metric spaces is the *doubling dimension* [Ass83, Hei12]. This concept has been used to measure the intrinsic dimension of sets in topological data analysis, see [Oud15, Chapter 5] or [She13, CK14] for some examples.

Definition 5.1. The *doubling constant* of a metric space is the smallest number λ such that every ball of radius R can be covered by λ balls of radius $R/2$. The *doubling dimension* is defined as $\dim_d(V)=\log_2(\lambda)$.

It is intuitively clear that the doubling dimension of Euclidean spaces $\mathbb{R}^d$ is of order d, but it can be considerably lower for certain subsets of $\mathbb{R}^d$. The metric spaces we consider here are finite subspaces $X \subset \mathbb{R}^d$. The *diameter* $\mathrm{diam}(X)$ is the largest pairwise distance between points in X. The *spread* Δ of such a set is the ratio of the diameter to the smallest pairwise distance between points in X. If a space X has doubling dimension $\dim_d(X)$, then it is easy to see that any ball of radius R can be covered with $\lambda^{\log_2(R/r)} = (R/r)^{\dim_d(X)}$ balls of radius r. From this it can be deduced that the cardinality $n = |X|$, the spread Δ and the doubling dimension are related as

$$n \leq \Delta^{\dim_d(X)}. \tag{5.1}$$

In [She13], a linear size approximation to the Vietoris-Rips complex has been derived and analysed in terms of the doubling dimension. The key is to approximate the point cloud X by a nested sequence of nets $\mathcal{N}_\alpha$, in such a way that at each relevant level the spread of the net is constant. The approach was further extended in [CK14], who introduce a local version of the doubling dimension.

It turns out that the doubling dimension is closely related to the Gaussian width, a fact pointed out in [IN07]. This relationship provides an alternative way of expressing the cardinality of a set of points in terms of the spread and some intrinsic geometric parameter. To make this relationship precise, we need to introduce the inequalities of Dudley and Sudakov. References are [Tal14] or [BLM13, Chapter 13].

A subset $\mathcal{N}_\alpha \subset X$ is called an α-net, if $\|\boldsymbol{x} - \boldsymbol{y}\| > \alpha$ for distinct $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{N}_\alpha$, and $\mathcal{N}_\alpha$ has maximal cardinality among sets with this property. For $\alpha > 0$ let $N(X, \alpha)$ denote the cardinality of an α-net. The logarithm $H(X, \alpha) = \log N(X, \alpha)$ is often referred to as the *metric entropy* of X. (One version of) Dudley's upper bound on the Gaussian width in terms of metric entropy is given as follows (following [BLM13, Corollary 13.2]).

Theorem 5.2. *(Dudley* [Dud67]*) Let $X \subset \mathbb{R}^d$ be a finite set. Then*

$$w(X) \leq 12 \int_0^{\mathrm{diam}(X)/2} \sqrt{H(X,t)}\, \mathrm{d}t. \tag{5.2}$$

There is a corresponding lower bound, due to Sudakov. A reference is again [BLM13], though explicit constants are never included in the literature.

Theorem 5.3. *(Sudakov* [Sud69]*) Let $X \subset \mathbb{R}^d$ be a finite set with r the smallest distance between points in X. Then*

$$w(X) \geq \frac{3}{5} r \sqrt{\log(|X|)}. \tag{5.3}$$

The upper bound in the following Proposition is from [IN07, (2)], with the difference that, as in Theorem 1.1, we look at the Gaussian width of the set of normalised differences,

$$T = \left\{ \frac{\boldsymbol{x} - \boldsymbol{y}}{\|\boldsymbol{x} - \boldsymbol{y}\|} : \boldsymbol{x}, \boldsymbol{y} \in X \right\}.$$

Contrary to the tradition, the bounds are stated using rather specific constants instead of only "some universal constant C or L". While the precise values given depend on details of the chosen analysis and are not important, for someone looking into actually using dimensionality reduction schemes it may be of interest to know if the "universal constants" are in the tens or in the billions.

Proposition 5.4. *Let $X \subset \mathbb{R}^d$ be a finite set. Then*

$$\frac{9}{25} \cdot \Delta(X)^{-2} \dim_d(X) \leq w^2(T) \leq 228 \cdot \Delta(X)^2 \dim_d(X)$$

Proof. We first relate the Gaussian width of T to that of X. Let $R = \mathrm{diam}(X) = \max_{\boldsymbol{x},\boldsymbol{y}} \|\boldsymbol{x} - \boldsymbol{y}\|$ and $r = \min_{\boldsymbol{x} \neq \boldsymbol{y}} \|\boldsymbol{x} - \boldsymbol{y}\|$, so that $\Delta = R/r$. Note that, by the symmetry of the Gaussian distribution,

$$w(X - X) = \mathbb{E}_{\boldsymbol{g}} \sup_{\boldsymbol{x},\boldsymbol{y} \in X} \langle \boldsymbol{g}, \boldsymbol{x} - \boldsymbol{y} \rangle = 2\, \mathbb{E} \sup_{\boldsymbol{x} \in X} \langle \boldsymbol{g}, \boldsymbol{x} \rangle = 2w(X).$$

Using this, one readily derives the bounds

$$\frac{2}{R}w(X) \leq w(T) \leq \frac{2}{r}w(X). \tag{5.4}$$

The upper bound in the statement of the theorem was given in [IN07], though we recreate the argument with slightly better constants. From (5.1), we get the inequality

$$N(X,\alpha) \leq \left(\frac{R}{\alpha}\right)^{\mathrm{dim}_d(X)},$$

which implies, using Dudley's bound (5.2),

$$\begin{aligned} w(X) &\leq 12\sqrt{\mathrm{dim}_d(X)} \int_0^{R/2} \sqrt{\log(R/t)}\,\mathrm{d}t \\ &\leq 12R\sqrt{\mathrm{dim}_d(X)} \int_0^1 \sqrt{\log(1/t)}\,\mathrm{d}t = 6\sqrt{\pi}R\sqrt{\mathrm{dim}_d(X)}. \end{aligned}$$

Squaring the right-hand side and combining with (5.4) gives the desired bound.

For the lower bound, let $\boldsymbol{p} \in X$ and $B(\boldsymbol{p},\alpha)$ be a ball of radius α such that there is a minimal covering $\mathcal{C}_{\alpha/2}$ of $B(\boldsymbol{p},\alpha) \cap X$ with λ balls $B(\boldsymbol{x}_i,\alpha/2)$, $1 \leq i \leq \lambda$, where λ is the doubling constant. Since the covering is minimal, we have $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \geq \alpha$ for all $1 \leq i,j \leq \lambda$. Set $S = \{\boldsymbol{x}_1,\dots,\boldsymbol{x}_\lambda\}$. Then

$$w(X) \geq w(S) \geq \frac{3}{5}\alpha\sqrt{\log(|S|)} \geq \frac{3}{5}r\sqrt{\mathrm{dim}_d(X)},$$

where the first inequality follows from the monotonicity of the Gaussian width, and the second from Sudakov's inequality. Squaring the right-hand side and combining with (5.4) shows the claimed lower bound. ☐

The "correct" way of bounding the Gaussian width from above and below would be via Talagrand's γ_2 functional. Following [Tal14, Section 2.2], we call a nested sequence of partitions $\mathcal{A} = (\mathcal{A}_n)$ of X *admissible*, if each partition satisfies the cardinality bound $|\mathcal{A}_n| \leq 2^{2^n}$ for $n \geq 1$. For each n and each $\boldsymbol{x} \in X$ there is a unique element $A_n(\boldsymbol{x}) \in \mathcal{A}_n$ which contains $\boldsymbol{x}$. The γ_2 functional is then defined as

$$\gamma_2(X) = \inf_{\mathcal{A}} \sup_{\boldsymbol{x}\in X} \sum_{n\geq 0} 2^{n/2}\mathrm{diam}(A_n(\boldsymbol{x})),$$

where the infimum is over all admissible partition sequences. Talagrand's Majorizing Measures Theorem [Tal14, Theorem 2.4.1] gives upper and lower bounds on the Gaussian width in terms of this functional. For some universal constant C, we have

$$\frac{1}{C}\gamma_2(X) \leq w(X) \leq C\gamma_2(X).$$

Note that we can alternatively represent an admissible sequence of partitions as a hierarchical tree, where each level approximates the data set more accurately.

We suspect that proofs based on net-trees, such as those of the statements in [She13, CK14], could be rewritten in terms of the Gaussian width by associating to the point cloud (and all the associated nets) a stochastic process $g_{\boldsymbol{x}} = \langle \boldsymbol{g}, \boldsymbol{x}\rangle$ for $\boldsymbol{x} \in T$ (or $\boldsymbol{x} \in X$).

REFERENCES

[ALMT14] Dennis Amelunxen, Martin Lotz, Michael B. McCoy, and Joel A. Tropp. Living on the edge: phase transitions in convex programs with random data. *Information and Inference*, 3(3):224–294, 2014.

[ALW17] Dennis Amelunxen, Martin Lotz, and Jake Walvin. Effective condition number bounds for convex regularization. *arXiv preprint arXiv:1707.01775*, 2017.

[Ass83] Patrice Assouad. Plongements lipschitziens dans $\mathbb{R}^n$. *Bulletin de la Société Mathématique de France*, 111:429–448, 1983.

[Bau17] Ulrich Bauer. Private communication, April 2017, 2017.

[BDN15] Jean Bourgain, Sjoerd Dirksen, and Jelani Nelson. Toward a unified theory of sparse dimensionality reduction in euclidean space. *Geometric and Functional Analysis*, 25(4):1009–1088, 2015.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[Car09] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

[CDSGO16] Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve Oudot. *The structure and stability of persistence modules*. Springer, 2016.

[CK14] Aruni Choudhary and Michael Kerber. Local doubling dimension of point sets. *arXiv preprint arXiv:1406.4822*, 2014.

[CP11] Emmanuel J Candes and Yaniv Plan. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory*, 57(4):2342–2359, 2011.

[CRPW12] Venkat Chandrasekaran, Ben Recht, Pablo A. Parrilo, and Alan S. Willsky. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6):805–849, 2012.

[Dir16] Sjoerd Dirksen. Dimensionality reduction with subgaussian matrices: a unified theory. *Foundations of Computational Mathematics*, 16(5):1367–1396, 2016.

[Dud67] Richard M Dudley. The sizes of compact subsets of hilbert space and continuity of gaussian processes. *Journal of Functional Analysis*, 1(3):290–330, 1967.

[EH10] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

[FR13] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*, volume 336 of *Applied and Numerical Harmonic Analysis*. Birkhäuser, Basel, 2013.

[Gor88] Yehoram Gordon. On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric aspects of functional analysis (1986/87)*, volume 1317 of *Lecture Notes in Math.*, pages 84–106. Springer, Berlin, 1988.

[Gro87] Mikhael Gromov. Monotonicity of the volume of intersection of balls. In *Geometrical aspects of functional analysis*, pages 1–4. Springer, 1987.

[Hat02] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

[Hei12] Juha Heinonen. *Lectures on analysis on metric spaces*. Springer Science & Business Media, 2012.

[IN07] Piotr Indyk and Assaf Naor. Nearest-neighbor-preserving embeddings. *ACM Transactions on Algorithms (TALG)*, 3(3):31, 2007.

[KR14] Michael Kerber and Sharath Raghvendra. Approximation and streaming algorithms for projective clustering via random projections. *arXiv preprint arXiv:1407.2063*, 2014.

[KW11] Felix Krahmer and Rachel Ward. New and improved johnson–lindenstrauss embeddings via the restricted isometry property. *SIAM Journal on Mathematical Analysis*, 43(3):1269–1281, 2011.

[LN14] Kasper Green Larsen and Jelani Nelson. The johnson-lindenstrauss lemma is optimal for linear dimensionality reduction. *arXiv preprint arXiv:1411.2404*, 2014.

[Mat02] Jiří Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[Mun84] James R Munkres. *Elements of Algebraic Topology*. Advanced book classics. Perseus Books, 1984.

[NDEG13] S. Nam, M. E. Davies, M. Elad, and R. Gribonval. The cosparse analysis model and algorithms. *Appl. Comput. Harmon. Anal.*, 34(1):30–56, 2013.

[NNM$^+$96] Sameer A Nene, Shree K Nayar, Hiroshi Murase, et al. Columbia object image library (coil-20). 1996.

[OPT$^+$17] Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6(1):17, 2017.

[ORS15] Samet Oymak, Benjamin Recht, and Mahdi Soltanolkotabi. Isometric sketching of any set via the restricted isometry property. *arXiv preprint arXiv:1506.03521*, 2015.

[Oud15] Steve Y Oudot. *Persistence theory: from quiver representations to data analysis*, volume 209. AMS, 2015.

[RFP10] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

[RV08] Mark Rudelson and Roman Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61(8):1025–1045, 2008.

[She13] Donald R Sheehy. Linear-size approximations to the vietoris–rips filtration. *Discrete & Computational Geometry*, 49(4):778–796, 2013.

[She14] Donald R Sheehy. The persistent homology of distance functions under random projection. In *Proceedings of the thirtieth annual symposium on Computational geometry*, page 328. ACM, 2014.

[She17] Don Sheehy. Private communication, April 2017, 2017.

[Sto09] Mihailo Stojnic. Various thresholds for ℓ_1-optimization in compressed sensing. *preprint*, 2009. arXiv:0907.3666.

[Sud69] Vladimir N Sudakov. Gaussian measures, Cauchy measures and ε-entropy. *Soviet Math. Doklady*, 10:310–313, 1969.

[Tal14] Michel Talagrand. *Upper and lower bounds for stochastic processes: modern methods and classical problems*, volume 60. Springer Science & Business Media, 2014.

SCHOOL OF MATHEMATICS, THE UNIVERSITY OF MANCHESTER