

Submitted exclusively to the London Mathematical Society
doi:10.1112/0000/000000

Estimating multidimensional persistent homology through a finite sampling

Niccolò Cavazza, Massimo Ferri and Claudia Landi

ABSTRACT

An exact computation of the persistent Betti numbers of a submanifold X of a Euclidean space is possible only in a theoretical setting. In practical situations, only a finite sample of X is available. We show that, under suitable density conditions, it is possible to estimate the multidimensional persistent Betti numbers of X from the ones of a union of balls centered on the sample points; this even yields the exact value in restricted areas of the domain.

Similar inequalities are proved for the multidimensional persistent Betti numbers of the ball union and the one of a combinatorial description of it.

1. Introduction

Persistent Topology is an innovative way of matching topology and geometry, and it proves to be an effective mathematical tool in Pattern Recognition [2, 12, 4]. This new research area is experiencing a period of intense theoretical progress, particularly in the form of the multidimensional *persistent Betti numbers* (PBNs; also called *rank invariant* in [5]). In order to express its full potential for applications, it has to interface with the typical environment of Computer Science: It must be possible to deal with a finite sampling of the object of interest, and with combinatorial representations of it.

A predecessor of the PBNs, the *size function* (i.e. PBNs at degree zero) already enjoys such a connection, in that it is possible to estimate it from a finite, sufficiently dense sampling [11], and it is possible to simplify the computation by processing a related graph [8]. Moreover, strict inequalities hold only in “blind strips”, i.e. in the ω -neighborhood of the discontinuity lines, where ω is the modulus of continuity of the measuring (filtering) function. Out of the blind strips, the values of the size function of the original object, of a ball covering of it, and of the related graph coincide.

The present paper extends this result to the PBNs of any degree, by using a recent article by P. Niyogi, S. Smale, and S. Weinberger [14], for a ball covering of the object X — a submanifold of a Euclidean space — with balls centered at dense enough points of X or near X : Theorems 4.1 and 4.3. A combinatorial representation of X , with the corresponding inequalities (Theorem 5.2) is based on a construction by H. Edelsbrunner [9].

Some elementary examples illustrate the results.

2. Preliminary results

In this section, we report some results on the stability of the PBNs and some topological properties of compact Riemannian submanifolds of \mathbb{R}^m . For basic notions on homology and persistent homology we refer the reader to [15] and [10], for classical properties of submanifolds to [13].

2.1. Multidimensional persistent Betti numbers

In this paper we will always work with coefficients in a field \mathbb{K} , so that all homology modules are vector spaces. First we define the following relation \prec (resp. \preceq) in \mathbb{R}^n : if $\vec{u} = (u_1, \dots, u_n)$ and $\vec{v} = (v_1, \dots, v_n)$, we write $\vec{u} \prec \vec{v}$ (resp. $\vec{u} \preceq \vec{v}$) if and only if $u_j < v_j$ (resp. $u_j \leq v_j$) for $j = 1, \dots, n$. We also define Δ^+ as the open set $\{(\vec{u}, \vec{v}) \in \mathbb{R}^n \times \mathbb{R}^n \mid \vec{u} \prec \vec{v}\}$.

As usual, a topological space X is *triangulable* if there is a finite simplicial complex whose underlying space is homeomorphic to X ; for a submanifold of a Euclidean space, it will mean that its triangulation can be extended to a domain containing it.

DEFINITION 2.1 Persistent Betti numbers. *Let X be a triangulable space and $\vec{f} = (f_1, \dots, f_n) : X \rightarrow \mathbb{R}^n$ be a continuous function. \vec{f} is called a measuring function and (X, \vec{f}) a size pair. We denote by $X\langle \vec{f} \preceq \vec{u} \rangle$ the lower level subset $\{p \in X \mid f_j(p) \leq u_j, j = 1, \dots, n\}$. Then, for each $i \in \mathbb{Z}$, the i -th multidimensional persistent Betti number function of (X, \vec{f}) is $\beta_{(X, \vec{f}, i)} : \Delta^+ \rightarrow \mathbb{N}$ defined as $\beta_{(X, \vec{f}, i)}(\vec{u}, \vec{v}) = \dim(\text{Im } f_{\vec{u}}^{\vec{v}})$, with*

$$f_{\vec{u}}^{\vec{v}} : \check{H}_i(X\langle \vec{f} \preceq \vec{u} \rangle) \rightarrow \check{H}_i(X\langle \vec{f} \preceq \vec{v} \rangle),$$

the homomorphism induced by the inclusion map of the sublevel set $X\langle \vec{f} \preceq \vec{u} \rangle \subseteq X\langle \vec{f} \preceq \vec{v} \rangle$. Here \check{H}_i denotes the i -th Čech homology module. We use Čech homology because it guarantees some useful continuity properties [6].

REMARK 2.2. When PBNs were defined in terms of filtering functions [7] these were so taken as to be tame, but in our case, since we need continuous maps, we refer to [6, Sect. 2.1] for the relevant extension.

PBNs give us a way to analyze triangulable spaces through their homological properties. Then it is natural to introduce a distance for comparing them. This has been done, and through this distance it has been possible to prove stability of the PBNs under variations of the measuring function in the one-dimensional [7] and multidimensional case [6].

2.2. Topological properties of compact Riemannian submanifolds of \mathbb{R}^m

As hinted in Section 1, we are interested in getting information, on a submanifold of \mathbb{R}^m , via a finite sampling of it and a related ball covering. To this goal, we now state some properties of compact Riemannian submanifolds of \mathbb{R}^m , especially referred to such an approximating covering. Definition 2.4 and Proposition 2.5 are due to P. Niyogi, S. Smale, and S. Weinberger ([14]). The main idea is that, under suitable hypotheses, it is possible to get, from a sampling of a submanifold, a ball covering whose union retracts on it.

First, notice that the spaces \mathbb{R}^m and \mathbb{R}^n play two different rôles in our arguments: the ambient space of our submanifolds (which will always be \mathbb{R}^m) is endowed with the classical Euclidean norm and has no partial order relation on it. On the other hand, the codomain of the measuring functions (\mathbb{R}^n throughout) is endowed with the max norm and with the partial order relation \preceq , as defined at the beginning of Section 2.1.

DEFINITION 2.3 Modulus of continuity.

Let $\vec{f} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ be a continuous function. Then, for $\varepsilon \in \mathbb{R}^+$, the modulus of continuity $\Omega(\varepsilon)$ of \vec{f} is:

$$\Omega(\varepsilon) = \max_{j=1,\dots,n} \sup \left\{ \text{abs}(f_j(\vec{p}) - f_j(\vec{p}')) \mid \vec{p}, \vec{p}' \in \mathbb{R}^m, \|\vec{p} - \vec{p}'\| \leq \varepsilon \right\}.$$

In other words $\Omega(\varepsilon)$ is the maximum over all moduli of continuity of the single components of \vec{f} .

By normal bundle of radius s at a point $p \in X$, we mean the collection of all vectors of length less than s anchored at p and with direction normal to X (orthogonal to Tan_p , the tangent space of X in p).

The embedding of the open normal bundle of radius s described above is a tubular neighborhood of X in \mathbb{R}^m , $Tub_s = \{p + \eta \in \mathbb{R}^m \mid p \in X, \eta \in Tan_p^\perp, \|\eta\| < s\}$, where Tan_p^\perp denotes the set of vectors normal to Tan_p . Tub_s can also be seen as $\bigcup_{p \in X} (Tan_p^\perp \cap B(p, s))$, where $B(p, s)$ is the ball of radius s centered at p . For a detailed definition and discussion on it we refer to [1].

A condition number $\frac{1}{\tau}$ is associated with a compact Riemannian submanifold X of \mathbb{R}^m .

DEFINITION 2.4.

τ is the largest number such that every open normal bundle B about X of radius s is embedded in \mathbb{R}^m for $s < \tau$.

PROPOSITION 2.5. [14, Prop. 3.1]

Let X be a compact Riemannian submanifold of \mathbb{R}^m . Let $L = \{l_1, \dots, l_k\}$ be a collection of points of X , and let $U = \bigcup_{j=1,\dots,k} B(l_j, \delta)$ be the union of balls of \mathbb{R}^m with center at the points of L and radius δ . Now, if L is such that for every point $p \in X$ there exists an $l_j \in L$ such that $\|p - l_j\| < \frac{\delta}{2}$, then, for every $\delta < \sqrt{\frac{3}{5}}\tau$, X is a deformation retract of U . So they have the same homology.

REMARK 2.6. The proof of Proposition 2.5 gives us a way to construct a retraction $\pi : U \rightarrow X$ and a homotopy $F : U \times I \rightarrow U$ such that $F(q, 0) = q$ and $F(q, 1) = \pi(q)$. Let $\pi_0 : Tub_\tau \rightarrow X$ be the canonical projection from the tubular neighborhood of radius τ of X onto X . Then π is the restriction of π_0 to U for which it holds:

$$\pi(q) = \arg \min_{p \in X} \|q - p\|.$$

Then the homotopy is given by

$$F(q, t) = (1 - t)q + t\pi(q).$$

It is also important to observe that the retraction π moves the points of U less than δ ; this is because the trajectory of $\pi(q)$ always remains inside a ball of U that contains q (q can be contained in the intersection of different balls), for every $q \in U$. In fact $\pi^{-1}(q) = U \cap Tan_q^\perp \cap B_\tau(q)$ (for a complete argument we refer to [14, Sect. 4]).

3. Retracts

Aim of this Section is to yield two rather general results, which will be specialized to Theorem 4.1 and Lemma 5.1. Throughout this Section, Y will be a compact Riemannian (hence triangulable) submanifold of \mathbb{R}^m and V will be a compact, triangulable subspace of \mathbb{R}^m such that Y is a deformation retract of V , with retraction r and homotopy $G : V \times I \rightarrow V$ from the identity of V , 1_V , to r . Moreover $\forall y \in Y, \forall v \in r^{-1}(y), \forall t \in I$ we assume that $(r \circ G)(v, t) = y$.

Let also $\vec{f} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ be a continuous function, and \vec{f}_Y and \vec{f}_V be the restrictions of \vec{f} to Y and V respectively.

LEMMA 3.1. $Y\langle \vec{f}_Y \preceq \vec{x} \rangle$ is a deformation retract of $V\langle \vec{f}_Y \circ r \preceq \vec{x} \rangle$.

Proof. Let $r_{\vec{x}} : V\langle \vec{f}_Y \circ r \preceq \vec{x} \rangle \rightarrow Y\langle \vec{f}_Y \preceq \vec{x} \rangle$ be the restriction of r to $V\langle \vec{f}_Y \circ r \preceq \vec{x} \rangle$. It is well-defined since, by the definition of the two sets, $r_{\vec{x}}(V\langle \vec{f}_Y \circ r \preceq \vec{x} \rangle) \subseteq Y\langle \vec{f}_Y \preceq \vec{x} \rangle$. We now set $G_{\vec{x}} : V\langle \vec{f}_Y \circ r \preceq \vec{x} \rangle \times I \rightarrow V\langle \vec{f}_Y \circ r \preceq \vec{x} \rangle$ as the restriction of G to $V\langle \vec{f}_Y \circ r \preceq \vec{x} \rangle \times I$. This restriction is well-defined, because the path from v to $r_{\vec{x}}(v)$ is all contained in $V\langle \vec{f}_Y \circ r \preceq \vec{x} \rangle$, thanks to the assumptions on G and r . Moreover, it is continuous and for every $v \in V\langle \vec{f}_Y \circ r \preceq \vec{x} \rangle$, $G_{\vec{x}}(v, 0) = v$ and $G_{\vec{x}}(v, 1) = r_{\vec{x}}(v)$. So it is the searched for deformation retraction. \square

REMARK 3.2. Since the homotopy G is relative to Y (i.e. keeps the points of Y fixed throughout), this is what is called a strong deformation retract in [15].

Now let $\varepsilon = \max_{v \in V} \|r(v) - v\|$ and $\vec{\omega}(\varepsilon) = (\Omega(\varepsilon), \dots, \Omega(\varepsilon)) \in \mathbb{R}^n$, where Ω is the modulus of continuity of \vec{f} (Def. 2.3).

LEMMA 3.3.

If (\vec{u}, \vec{w}) is a point of Δ^+ and if $\vec{u} + \vec{\omega}(\varepsilon) \prec \vec{w} - \vec{\omega}(\varepsilon)$, then

$$\beta_{(V, \vec{f}_V, i)}(\vec{u} - \vec{\omega}(\varepsilon), \vec{w} + \vec{\omega}(\varepsilon)) \leq \beta_{(Y, \vec{f}_Y, i)}(\vec{u}, \vec{w}) \leq \beta_{(V, \vec{f}_V, i)}(\vec{u} + \vec{\omega}(\varepsilon), \vec{w} - \vec{\omega}(\varepsilon))$$

Proof. First, observe that there are inclusions

$$\begin{aligned} \gamma : Y\langle \vec{f}_Y \preceq \vec{u} \rangle &\rightarrow Y\langle \vec{f}_Y \preceq \vec{w} \rangle \\ \varphi : Y\langle \vec{f}_Y \preceq \vec{w} \rangle &\rightarrow V\langle \vec{f}_V \preceq \vec{w} + \vec{\omega}(\varepsilon) \rangle \\ \psi : V\langle \vec{f}_V \preceq \vec{u} - \vec{\omega}(\varepsilon) \rangle &\rightarrow V\langle \vec{f}_V \preceq \vec{w} + \vec{\omega}(\varepsilon) \rangle \end{aligned}$$

The fact that r moves every point not more than ε (since $\varepsilon = \max_{v \in V} \|r(v) - v\|$) implies that also the inclusions $\eta : V\langle \vec{f}_V \preceq \vec{u} - \vec{\omega}(\varepsilon) \rangle \rightarrow V\langle \vec{f}_Y \circ r \preceq \vec{u} \rangle$ and $\theta : V\langle \vec{f}_Y \circ r \preceq \vec{u} \rangle \rightarrow V\langle \vec{f}_V \preceq \vec{w} + \vec{\omega}(\varepsilon) \rangle$ make sense. Let further $\bar{r} = r_{\vec{u}} \circ \eta$, where $r_{\vec{u}}$ is as in the proof of Lemma 3.1 with $\vec{x} = \vec{u}$. Now we have the following (not necessarily commutative) diagram:

$$\begin{array}{ccc} V\langle \vec{f}_V \preceq \vec{u} - \vec{\omega}(\varepsilon) \rangle & \xrightarrow{\bar{r}} & Y\langle \vec{f}_Y \preceq \vec{u} \rangle \\ \psi \downarrow & & \downarrow \gamma \\ V\langle \vec{f}_V \preceq \vec{w} + \vec{\omega}(\varepsilon) \rangle & \xleftarrow{\varphi} & Y\langle \vec{f}_Y \preceq \vec{w} \rangle \end{array}$$

The next step is to prove that ψ is homotopic to $\varphi \circ \gamma \circ \bar{r}$. Let $\bar{G} : V\langle \vec{f}_V \preceq \vec{u} - \vec{\omega}(\varepsilon) \rangle \times I \rightarrow V\langle \vec{f}_Y \preceq \vec{w} + \vec{\omega}(\varepsilon) \rangle$ be the composition $\bar{G} = \theta \circ G_{\vec{u}} \circ (\eta \times 1_I)$, where $G_{\vec{u}}$ is as in the proof of Lemma 3.1. Now, for every $v \in V\langle \vec{f}_V \preceq \vec{u} - \vec{\omega}(\varepsilon) \rangle$, we have $\bar{G}(v, 0) = G(v, 0) = v = \psi(v)$ and $\bar{G}(v, 1) = G(v, 1) = r(v) = \bar{r}(v) = \varphi \circ \gamma \circ \bar{r}(v)$.

Since homotopic maps induce the same homomorphisms in homology, we have (setting $\psi^* = \psi_{\vec{u} - \vec{\omega}(\varepsilon)}^{w + \vec{\omega}(\varepsilon)}$ and $\gamma^* = \gamma_{\vec{u}}^{\vec{w}}$, and φ^*, \bar{r}^* are the homology homomorphisms induced by φ and \bar{r} respectively)

$$\begin{aligned} \beta_{(V, \vec{f}_V, i)}(\vec{u} - \vec{\omega}(\varepsilon), \vec{w} + \vec{\omega}(\varepsilon)) &= \dim \operatorname{Im}(\psi^*) = \dim \operatorname{Im}(\varphi^* \circ \gamma^* \circ \bar{r}^*) \leq \\ &\leq \dim \operatorname{Im}(\gamma^*) = \beta_{(Y, \vec{f}_Y, i)}(\vec{u}, \vec{w}) \end{aligned}$$

concluding the first part of the proof.

For the second inequality we use the following commutative (as will be proved) diagram, with analogous definitions of maps γ' , ψ' , φ' and \bar{r}' :

$$\begin{array}{ccc} Y\langle \vec{f}_Y \preceq \vec{u} \rangle & \xrightarrow{\varphi'} & V\langle \vec{f}_V \preceq \vec{u} + \vec{\omega}(\varepsilon) \rangle \\ \gamma' \downarrow & & \downarrow \psi' \\ Y\langle \vec{f}_Y \preceq \vec{w} \rangle & \xleftarrow{\bar{r}'} & V\langle \vec{f}_V \preceq \vec{w} - \vec{\omega}(\varepsilon) \rangle \end{array}$$

Here ψ' is well defined because we are assuming $\vec{u} + \vec{\omega}(\varepsilon) \prec \vec{w} - \vec{\omega}(\varepsilon)$.

Then, passing to homology, we have (with analogous settings for the starred symbols):

$$\begin{aligned} \beta_{(Y, \vec{f}_Y, i)}(\vec{u}, \vec{w}) &= \dim \operatorname{Im}(\gamma'^*) = \dim \operatorname{Im}(\bar{r}'^* \circ \psi'^* \circ \varphi'^*) \leq \\ &\leq \dim \operatorname{Im}(\psi'^*) = \beta_{(V, \vec{f}_V, i)}(\vec{u} + \vec{\omega}(\varepsilon), \vec{w} - \vec{\omega}(\varepsilon)) \end{aligned}$$

To prove the commutativity of the diagram, we observe that \bar{r}' is the identity map on the points of Y . Since $Y\langle \vec{f}_Y \preceq \vec{u} \rangle \subseteq Y$, we have that $\bar{r}' \circ \psi' \circ \varphi'$ is the canonical inclusion of $Y\langle \vec{f}_Y \preceq \vec{u} \rangle$ in $Y\langle \vec{f}_Y \preceq \vec{w} \rangle$. \square

4. Ball coverings

Throughout this Section, X will be a compact Riemannian (triangulable) submanifold of \mathbb{R}^m . As hinted in Section 1, we want to get information on X out of a finite set of points. First, the points will be sampled on X itself, then even in a (narrow) neighborhood. In both cases, the idea is to consider a covering of X made of balls centered on the sampling points.

What we get, is a double inequality which yields an estimate of the PBNs of X within a fixed distance from the discontinuity sets of the PBNs (meant as integer functions on Δ^+) of the union U of the balls of the covering, but even offers the exact value of it at points sufficiently far from the discontinuity sets.

When we foliate the domain Δ^+ of the PBNs as in [3, Sect. 2.1] or [6, Sect. 3] — or simply when $n = 1$ — the discontinuity sets are (possibly infinite) line segments, and the regions where only the inequality holds appear as strips around them (which we colloquially call “blind strips”). The width of such strips is a representation of the approximation error, in that it is directly related to $\Omega(\delta)$, where $1/\delta$ represents the density of the sampling.

4.1. Points on X

Let $\delta < \sqrt{\frac{3}{5}}\tau$ and let $L = \{l_1, \dots, l_k\}$ be a set of points of X such that for every $p \in X$ there exists an $l_j \in L$ for which $\|p - l_j\| < \frac{\delta}{2}$. Let U be the union of the balls $B(l_j, \delta)$ of radius δ centered at l_j , $j = 1, \dots, k$. So all conditions of Proposition 2.5 are satisfied.

THEOREM 4.1.

If (\vec{u}, \vec{v}) is a point of Δ^+ and if $\vec{u} + \vec{\omega}(\delta) \prec \vec{v} - \vec{\omega}(\delta)$, where $\vec{\omega}(\delta) = (\Omega(\delta), \dots, \Omega(\delta)) \in \mathbb{R}^n$, then

$$\beta_{(U, \vec{f}_U, i)}(\vec{u} - \vec{\omega}(\delta), \vec{v} + \vec{\omega}(\delta)) \leq \beta_{(X, \vec{f}_X, i)}(\vec{u}, \vec{v}) \leq \beta_{(U, \vec{f}_U, i)}(\vec{u} + \vec{\omega}(\delta), \vec{v} - \vec{\omega}(\delta))$$

Proof. By Lemma 3.3, with $Y = X$, $V = U$. □

4.2. Points near X

So far we have approximated X by points picked up on X itself, but it is also possible to choose the points near X , by respecting some constraints. Once more, this is possible thanks to a result of [14].

PROPOSITION 4.2. [14, Prop. 7.1] *Let $L = \{l_1, \dots, l_k\}$ be a set of points in the tubular neighborhood of radius s around X and $U = \bigcup_{j=1, \dots, k} B(l_j, \delta)$ be the union of the balls of \mathbb{R}^m centered at the points of L and with radius δ . If for every points $p \in X$, there exist a point $l_j \in L$ such that $\|p - l_j\| < s$, then U is a deformation retract of X , for all $s < (\sqrt{9} - \sqrt{8})\tau$ and $\delta \in \left(\frac{(s+\tau) - \sqrt{s^2 + \tau^2 - 6s\tau}}{2}, \frac{(s+\tau) + \sqrt{s^2 + \tau^2 - 6s\tau}}{2} \right)$.*

Then, as with Theorem 4.1, we have, with an analogous proof:

THEOREM 4.3.

Under the hypotheses of Proposition 4.2, if (\vec{u}, \vec{v}) is a point of Δ^+ and if $\vec{u} + \vec{\omega}(\delta + s) \prec \vec{v} - \vec{\omega}(\delta + s)$, where $\vec{\omega}(\delta + s) = (\Omega(\delta + s), \dots, \Omega(\delta + s)) \in \mathbb{R}^n$, then

$$\beta_{(U, \vec{f}_U, i)}(\vec{u} - \vec{\omega}(\delta + s), \vec{v} + \vec{\omega}(\delta + s)) \leq \beta_{(X, \vec{f}_X, i)}(\vec{u}, \vec{v}) \leq \beta_{(U, \vec{f}_U, i)}(\vec{u} + \vec{\omega}(\delta + s), \vec{v} - \vec{\omega}(\delta + s))$$

4.3. Examples

The following examples show how Theorem 4.1 can be used for applications. Let X be a circle of radius 4 in \mathbb{R}^2 (Figure 1); we observe that τ is exactly the radius of X , so $\tau = 4$. In order to create a well defined approximation we need that $\delta < \sqrt{\frac{3}{5}}\tau$.

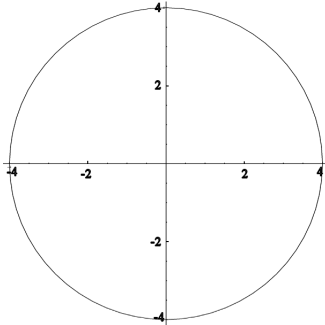


FIGURE 1. The circle of radius 4, X .

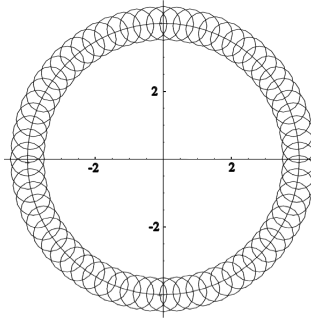


FIGURE 2. The ball union U .

In the first example we have taken $\delta = 0.5$. Now, to satisfy the hypothesis of Theorem 4.1 (that for every $p \in X$ there exist an $l_j \in L$ such that $\|p - l_j\| < \frac{\delta}{2}$), we have chosen 64 points l_j on X . Moreover we have sampled X uniformly, so that there is a point every $\frac{\pi}{32}$ radians (Figure 2). We stick to the monodimensional case, choosing $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, with $f(x, y) = \text{abs}(y)$. U is the resulting ball union.

Figures 3 and 4 represent the PBN functions at degree zero of X and U respectively. Δ^+ is the half-plane above the diagonal line, and the numbers are the values of the PBNs in the triangular regions they are written in. In Figure 3 there is only a big triangle where the

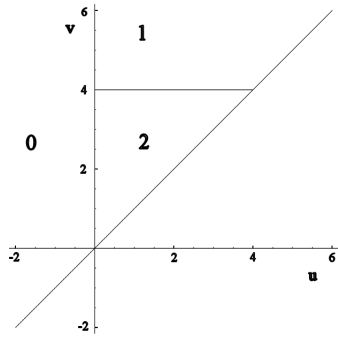


FIGURE 3. The representation of $\beta_{(X, f_X, 0)}$, the 0-PBNs of X .

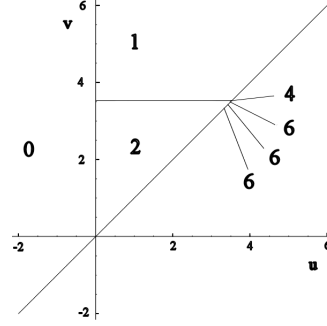


FIGURE 4. The representation of $\beta_{(U, f_U, 0)}$, the 0-PBNs of the ball union U

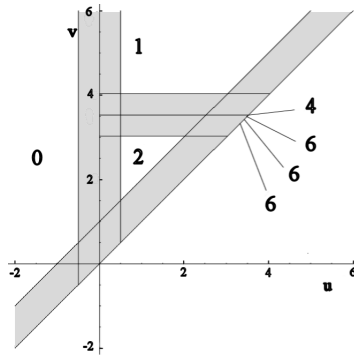


FIGURE 5. The blind strips of $\beta_{(U, f, 0)}$.

value 2 signals the two different connected components generated by f_X . The two connected components collapse to one at value 4. In Figure 4 there is also a big triangle representing the two connected components, but they collapse at value 3.53106. Moreover there are 4 other very small triangles near the diagonal, representing more connected components generated by the approximation error. In the last figure (Figure 5) the blind strips around the discontinuity lines of $\beta_{(U, f_U, 0)}$ are shown. The width of these strips, since $\Omega(\delta) = 0.5$, is equal to $2\Omega(\delta) = 1$. This figure illustrates the idea underlying Theorem 4.1. Taken a point (u, v) outside the strips, the values of the PBNs of U at $(u - \Omega(\delta), v + \Omega(\delta))$ and $(u + \Omega(\delta), v - \Omega(\delta))$ are the same. So also the value of the PBNs of X at (u, v) is determined. Figures 6, 7, 8 depict, in analogous way, the (obviously much simpler) PBNs of degree 1.

For a second example we have chosen the points l_j not necessarily on X . We have satisfied the hypothesis of Proposition 4.2, choosing $s = 0.25$ and $\delta = 0.55$. Then, in order to cover X well, we have chosen a point every $\frac{\pi}{48}$ radians, for a total of 96 points. But this time the points are either 0 or 0.1 or 0.2 away from X . Figure 9 shows the resulting ball union U' . As in the previous case, in the representation of $\beta_{(U', f_{U'}, 0)}$ (Figure 10) there is a big triangle showing two connected components and this time they collapse at value 3.40955. Compared to Figure 4, there are many more small triangles generated by the asymmetry of the sampling. The width of the blind strips in Figure 11 is $2\Omega(\delta + s) = 1.6$, so there is still the central triangle. This means that, although the error in the approximation is much bigger, the blind strips do not cover the entire figure, leaving the topological information intact at least in some small areas of Δ^+ .

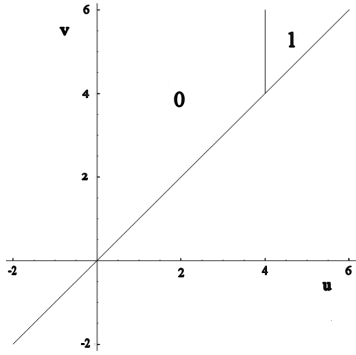


FIGURE 6. The representation of $\beta_{(X, f_X, 1)}$, the 1-PBNs of X .

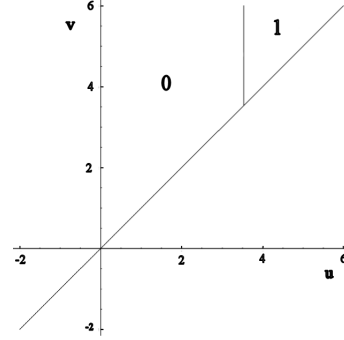


FIGURE 7. The representation of $\beta_{(U, f_U, 1)}$, the 1-PBNs of the ball union U .

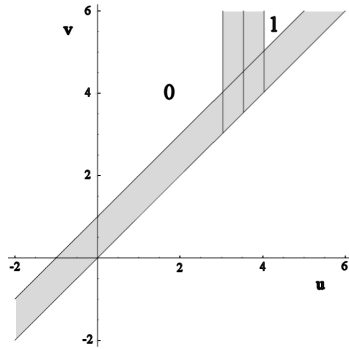


FIGURE 8. The blind strips of $\beta_{(U, f, 1)}$.

5. A combinatorial representation

The ball unions of Section 4, although generated by finite sets, are still continuous objects. It is desirable that the topological information on X , up to a certain approximation, be condensed in a combinatorial object. For size functions (i.e. for PBNs of degree 0) it was a graph; here, it has to be a simplicial complex. We shall build such a complex, by following [9], to which we refer for all definitions not reported here. Please note that [9] uses *weighted* Voronoi cells and diagrams, while we do not need to worry about that, since all of our balls have the same radius; so the customary Euclidean distance can be used instead of the power distance employed in that paper.

Let X , $L = \{l_1, \dots, l_k\}$ and δ be as in Section 4.1 (the case of Section 4.2 is an immediate extension). Moreover, let the points of L be in general position. For each $l_j \in L$, let $B_j = B(l_j, \delta)$ be the ball of radius δ , centered at l_j . The set $B = \{B_1, \dots, B_k\}$ is a ball covering of X ; denote by U the corresponding ball union. Let now V_j be the *Voronoi cell* of B_j , i.e. the set of points of \mathbb{R}^m whose distance from l_j is not greater than the distance from any other $l_{j'}$.

The set $\mathcal{V} = \{V_1, \dots, V_k\}$ is the *Voronoi diagram* of B . From \mathcal{V} we get the collection of cells $\mathcal{Q} = \{V_j \cap B_j \mid j = 1, \dots, k\}$, a decomposition of U .

The *nerve* $N(\mathcal{V})$ of \mathcal{V} is the abstract simplicial complex where vertices are the elements of \mathcal{V} and, for a subset T of $\{1, \dots, k\}$, the set of vertices $\{V_j \mid j \in T\}$ is a simplex if and only if $\bigcap_{j \in T} V_j \neq \emptyset$.

For any $T \subseteq \{1, \dots, k\}$, $T \neq \emptyset$ we denote by σ_T the convex hull of $\{l_j \mid j \in T\}$

The *dual complex* of \mathcal{Q} is $\mathcal{K} = \{\sigma_T \mid \{V_j \cap B_j \mid j \in T\} \in N(\mathcal{Q})\}$ and $\mathcal{S} = |\mathcal{K}|$, union of the simplices of \mathcal{K} , is the *dual shape* of U .

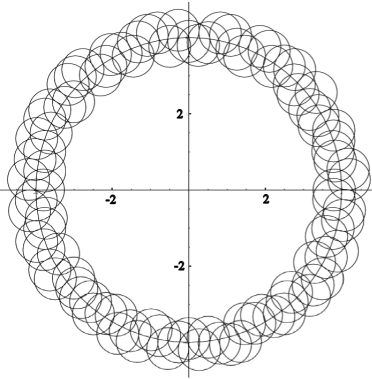


FIGURE 9. The ball union U' .

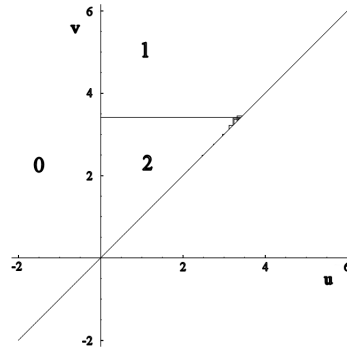


FIGURE 10. The representation of $\beta_{(U', f_{U'}, 0)}$, the 0-PBNs of the ball union U' .

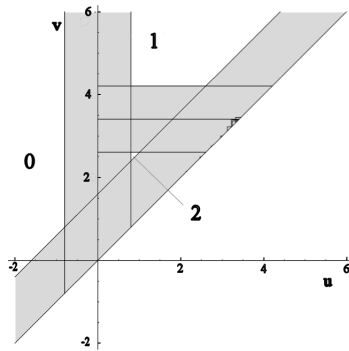


FIGURE 11. The blind strips of $\beta_{(U', f_{U'}, 0)}$.

For a better understanding of the previous part we produce a toy example. Let X be a quarter of circle of radius 4 and U be the union of nine balls of radius 1, with centers near X (Figure 12). The Voronoi Diagram \mathcal{V} associated to this ball covering B is depicted in Figure 13.

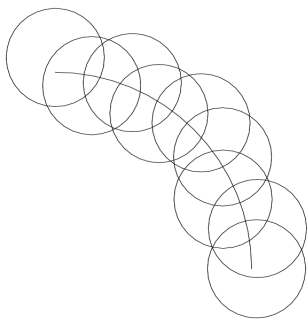


FIGURE 12. A quarter of circle of radius 4 covered by nine balls of radius 1.

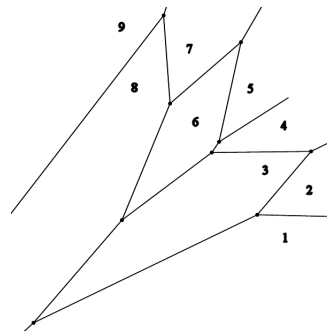


FIGURE 13. The Voronoi Diagram \mathcal{V} of B .

Now the main idea is that we can associate the dual complex \mathcal{K} with the submanifold X . In fact, by [9, Thm. 3.2], its space \mathcal{S} is homotopically equivalent to U and, by transitivity, to X . Moreover, [9, Sect. 3] explicitly builds a retraction r from U to \mathcal{S} and a homotopy H from

the identity of U , to p , such that $\forall y \in \mathcal{S}, \forall v \in p^{-1}(y), \forall t \in I$ we have $(p \circ H)(v, t) = y$. For a complete description of the homotopy H and the retraction p we refer to the original article.

\mathcal{D} , \mathcal{K} and \mathcal{S} are shown in Figures 14, 15 respectively.

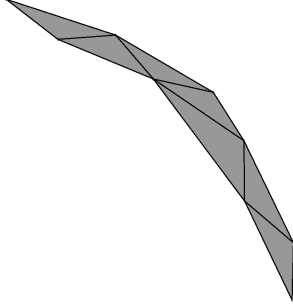


FIGURE 14. The dual complex \mathcal{K} .

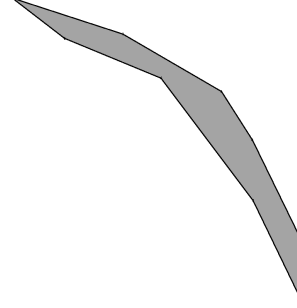


FIGURE 15. The dual shape \mathcal{S} .

5.1. Ball union and dual shape

Let $\vec{f}: \mathbb{R}^m \rightarrow \mathbb{R}^n$ be a continuous function and let $\vec{f}_{\mathcal{S}}$ and \vec{f}_U be the restrictions of \vec{f} to \mathcal{S} and U respectively.

LEMMA 5.1. *If (\vec{u}, \vec{v}) is a point of Δ^+ and if $\vec{u} + \vec{\omega}(\delta) \prec \vec{v} - \vec{\omega}(\delta)$, where $\vec{\omega}(\delta) = (\omega_1(\delta), \dots, \omega_n(\delta)) \in \mathbb{R}^n$, then*

$$\beta_{(U, \vec{f}_U, i)}(\vec{u} - \vec{\omega}(\delta), \vec{v} + \vec{\omega}(\delta)) \leq \beta_{(\mathcal{S}, \vec{f}_{\mathcal{S}}, i)}(\vec{u}, \vec{v}) \leq \beta_{(U, \vec{f}_U, i)}(\vec{u} + \vec{\omega}(\delta), \vec{v} - \vec{\omega}(\delta)).$$

Proof.

By Lemma 3.3, with $Y = \mathcal{S}$, $V = U$. □

Now we can get an estimate of the PBNs of X from the ones of \mathcal{S} . The blind strips will be doubly wide, with respect to the ones previously considered. Still, this can leave some regions of Δ^+ where the computation is exact.

THEOREM 5.2. *If (\vec{u}, \vec{v}) is a point of Δ^+ and if $\vec{u} + 2\vec{\omega}(\delta) \prec \vec{v} - 2\vec{\omega}(\delta)$, where $\vec{\omega}(\delta) = (\omega_1(\delta), \dots, \omega_n(\delta)) \in \mathbb{R}^n$, then*

$$\beta_{(\mathcal{S}, \vec{f}_{\mathcal{S}}, i)}(\vec{u} - 2\vec{\omega}(\delta), \vec{v} + 2\vec{\omega}(\delta)) \leq \beta_{(X, \vec{f}_X, i)}(\vec{u}, \vec{v}) \leq \beta_{(\mathcal{S}, \vec{f}_{\mathcal{S}}, i)}(\vec{u} + 2\vec{\omega}(\delta), \vec{v} - 2\vec{\omega}(\delta)).$$

Proof. By Lemma 4.1,

$$\beta_{(U, \vec{f}_U, i)}(\vec{u} - \vec{\omega}(\delta), \vec{v} + \vec{\omega}(\delta)) \leq \beta_{(X, \vec{f}_X, i)}(\vec{u}, \vec{v}) \leq \beta_{(U, \vec{f}_U, i)}(\vec{u} + \vec{\omega}(\delta), \vec{v} - \vec{\omega}(\delta))$$

Then we have

$$\beta_{(U, \vec{f}_U, i)}(\vec{u} + \vec{\omega}(\delta), \vec{v} - \vec{\omega}(\delta)) \leq \beta_{(\mathcal{S}, \vec{f}_{\mathcal{S}}, i)}(\vec{u} + 2\vec{\omega}(\delta), \vec{v} - 2\vec{\omega}(\delta))$$

by Lemma 5.1 by substituting (\vec{u}, \vec{v}) with $(\vec{u} + 2\vec{\omega}(\delta), \vec{v} - 2\vec{\omega}(\delta))$, and

$$\beta_{(\mathcal{S}, \vec{f}_{\mathcal{S}}, i)}(\vec{u} - 2\vec{\omega}(\delta), \vec{v} + 2\vec{\omega}(\delta)) \leq \beta_{(U, \vec{f}_U, i)}(\vec{u} - \vec{\omega}(\delta), \vec{v} + \vec{\omega}(\delta))$$

by Lemma 5.1 by substituting (\vec{u}, \vec{v}) with $(\vec{u} - 2\vec{\omega}(\delta), \vec{v} + 2\vec{\omega}(\delta))$. □

Acknowledgements

The authors wish to thank P. Frosini for the many helpful suggestions. This work was performed under the auspices of INdAM-GNSAGA, CIRAM, and ARCES.

Niccolò Cavazza and Massimo Ferri
ARCES and Dip. di Matematica,
Univ. di Bologna,
Italia

cavazza@dm.unibo.it
ferri@dm.unibo.it

Claudia Landi
Dip. di Scienze e Metodi dell'Ingegneria,
Univ. di Modena e Reggio Emilia,
Italia

clandi@unimore.it

References

1. R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Found. Comput. Math.*, 9(1):51–77, 2009.
2. S. Biasotti, A. Cerri, P. Frosini, D. Giorgi, and C. Landi. Multidimensional size functions for shape comparison. *J. Math. Imaging Vision*, 32(2):161–179, 2008.
3. F. Cagliari, B. Di Fabio, and M. Ferri. One-dimensional reduction of multidimensional persistent homology. *Proc. Amer. Math. Soc.*, 2010.
4. G. Carlsson. Topology and data. *Bull. Amer. Math. Soc.*, 46(2):255–308, 2009.
5. G. Carlsson and A. Zomorodian. The theory of multidimensional persistence. *Discrete Comput. Geometry*, 42(1):71–93, 2009.
6. A. Cerri, B. DiFabio, M. Ferri, P. Frosini, and C. Landi. Multidimensional persistent homology is stable. Technical report, July 2009.
7. D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1):103–120, 2007.
8. M. d'Amico. A new optimal algorithm for computing size function of shapes. In *CVPRIP Algorithms III, Proceedings International Conference on Computer Vision, Pattern Recognition and Image Processing*, pages 107–110, 2000.
9. H. Edelsbrunner. The union of balls and its dual shape. *Discrete Comput. Geom.*, 13:415–440, 1995.
10. H. Edelsbrunner and J. Harer. Persistent homology—a survey. In *Surveys on discrete and computational geometry*, volume 453 of *Contemp. Math.*, pages 257–282. Amer. Math. Soc., Providence, RI, 2008.
11. P. Frosini. Discrete computation of size functions. *J. of Combin., Inf. System Sci.*, 17(3-4):232–250, 1992.
12. R. Ghrist. Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc. (N.S.)*, 45(1):61–75 (electronic), 2008.
13. M. W. Hirsch. *Differential Topology*, volume 33 of *Graduate Text in Mathematics*. Springer, 1976.
14. P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39(1):419–441, 2008.
15. E. Spanier. *Algebraic topology*. McGraw-Hill, New York, 1966. Series in higher mathematics.