

# PERSISTENT HOMOLOGY OF SEMI-ALGEBRAIC SETS

SAUGATA BASU AND NEGIN KARISANI

**ABSTRACT.** We give an algorithm with singly exponential complexity for computing the *barcodes* up to dimension  $\ell$  (for any fixed  $\ell \geq 0$ ) of the filtration of a given semi-algebraic set by the sub-level sets of a given polynomial. Our algorithm is the first algorithm for this problem with singly exponential complexity, and generalizes the corresponding results for computing the Betti numbers up to dimension  $\ell$  of semi-algebraic sets with no filtration present.

## CONTENTS

1. Introduction	1
1.1. Background	1
1.2. Persistent Homology	2
1.3. Summary of the main contributions	3
1.4. Definition of complexity	5
1.5. Prior and Related Work	5
2. Precise definitions and statements of the main results	6
2.1. Persistent homology and barcodes	6
2.2. Semi-algebraic filtrations	10
3. Continuous to finite filtration	14
3.1. Reduction to the case of a finite filtration	14
3.2. Persistent multiplicities for finite filtration	22
4. Algorithms and proof of Theorem 1	24
4.1. Preliminaries	24
4.2. Some preliminary algorithms	25
4.3. Algorithm for computing simplicial replacement	27
4.4. Algorithm for reducing to a finite filtration	28
4.5. Computing barcodes of semi-algebraic filtrations	29
5. Future work and open problems	31
Acknowledgements	32
References	32

## 1. INTRODUCTION

**1.1. Background.** Let  $R$  be a real closed field and  $D$  an ordered domain contained in  $R$  which we fix for the rest of the paper. The algorithmic problem of computing

---

*Date:* **May 5, 2022.**

*2020 Mathematics Subject Classification.* Primary 14F25, 55N31; Secondary 68W30.

*Key words and phrases.* semi-algebraic sets, simplicial complex, persistent homology, barcodes.

Basu was partially supported by NSF grants CCF-1618918, DMS-1620271 and CCF-1910441.

the ranks of the homology groups of a given semi-algebraic set described by a quantifier-free formula, whose atoms are of the form  $P > 0, P \in \mathbb{D}[X_1, \dots, X_k]$ , as input has attracted a lot of attention over the years.<sup>1</sup> Closed and bounded semi-algebraic subsets  $S \subset \mathbb{R}^k$  are semi-algebraically triangulable – and moreover given a description of the semi-algebraic set by a quantifier-free formula, such a triangulation can be effectively computed with complexity (measured in terms of the number of polynomials appearing in the description and their degrees) which is doubly exponential in  $k$ . Together with standard algorithms of linear algebra, this gives an algorithm for computing all the Betti numbers of a given semi-algebraic set with doubly exponential complexity.

Classical bounds coming from Morse theory [37, 38, 36, 5, 26] gives singly exponential bounds on the Betti numbers of semi-algebraic sets. More precisely, if  $S \subset \mathbb{R}^k$  is a semi-algebraic set defined by a quantifier-free formula involving  $s$  polynomials of degree at most  $d$ , then the sum of the Betti numbers of  $S$  is bounded by  $(O(sd))^k$ . Additionally, it is known that the problem of computing the zero-th Betti number of  $S$  (i.e. the number of semi-algebraically components of  $S$ ) (using “roadmap” algorithms [17, 29, 2]), as well as the problem of computing the Euler-Poincaré characteristic of  $S$  (using Morse theory [5, 3]), both admit single exponential complexity algorithms. This led to a search for singly exponential complexity algorithm for computing the higher Betti numbers as well. The current state of the art is that there exists singly exponential algorithm for computing the first  $\ell$  Betti numbers of semi-algebraic sets for each fixed  $\ell \geq 0$  [6, 4].

In this paper we study the algorithmic complexity of computing a finer topological invariant of a given semi-algebraic set  $S$  than its Betti numbers – namely, *the barcode of a filtration* of the given semi-algebraic set. This finer invariant (unlike the Betti numbers) has both a discrete as well as a continuous part and is attached to a filtration of a semi-algebraic set by the sub-level sets of a semi-algebraic function. The problem reduces to the problem of computing the Betti numbers in the case when the given filtration is trivial.

**1.2. Persistent Homology.** One of the recent developments in the area of applied topology is the introduction of the notion of *persistent homology of filtrations*. We initiate in this paper the study of the algorithmic problem of computing the *persistent homology groups* (cf. Definition 2.1) of filtrations of semi-algebraic sets by polynomial functions. Persistent homology is a central object in the emerging field of topological data analysis [22, 39, 27], but has also found applications in diverse areas of mathematics and computations as well (see for example [25, 32]).

One can associate persistent homology groups to any filtration of topological spaces, and they generalize ordinary homology groups of a space  $X$  – which corresponds to the trivial (i.e. constant) filtration on  $X$ . To the best of our knowledge the algorithmic problem of computing persistent homology groups of semi-algebraic sets equipped with a filtration by the sub-level sets of a polynomial (or more generally continuous semi-algebraic functions) have not been considered from an algorithmic viewpoint. The output of a persistent homology computation is usually expressed in the form of “barcodes” [27].

---

<sup>1</sup>Here and everywhere else in the paper all homology groups considered are with rational coefficients.

We will define barcodes for semi-algebraic filtrations precisely later (see Definition 2.5). A basic example which is a starting point of persistent homology theory in the field of topological data analysis is the following one.

1.2.1. *Čech complex of a finite set of points.* Let  $X$  be a (finite) subset of  $\mathbb{R}^k$  (with its Euclidean metric). In practice,  $X$  may consist of a finite set of points (often called “point-cloud data”) which approximates some subspace or sub-manifold  $M$  of  $\mathbb{R}^k$ . The topology (in particular, the homology groups) of the manifold  $M$  is not reflected in the set of points  $X$  (which is a discrete topological space under the subspace topology induced from that of  $\mathbb{R}^k$ ). Now for  $r \geq 0$ , let  $X_r$  denote the union of closed Euclidean balls,  $\overline{B_k(x, r)}$ , of radius  $r$  centered at the points  $x \in X$ . Notice that each  $X_r$  is a semi-algebraic set indexed by  $r$ . In particular,  $X_0 = X$ . Also, for  $0 \leq r \leq r'$ , we have that  $X_r \subset X_{r'}$ . Thus,  $(X_r)_{r \geq 0}$  is an increasing family of semi-algebraic sets indexed by  $r \geq 0$ . Thus, this is an example of a semi-algebraic filtration (see Remark 7). The main rationale for considering this filtration is that nerve complex of the family of convex sets  $\overline{B_k(x, r)}, x \in X$  approximates homotopically the underlying manifold  $M$ , and each homology class of  $M$  would show up in the homology of  $X_r$  for some values of  $r$ . The barcode of the filtration  $(X_r)_{r \geq 0}$  is a tool for filtering out spurious homology (noise) from that which genuinely reflects the topology of  $M$  (see [27, 23]). The barcode of the above filtration thus plays an important role in topological data analysis. In particular they capture information about the homology of the underlying manifold  $M$ . It also serves as a “signature” for topological data (such as point cloud data). In the semi-algebraic world they play a similar role – for example, as a measure of topological similarity of two given semi-algebraic sets which is much finer (because of the presence of the continuous parameters) to just the sequence of Betti numbers.

As stated earlier, the main goal in this paper is to design an efficient (singly exponential complexity algorithm) that takes as input a quantifier-free formula describing a closed semi-algebraic set  $S \subset \mathbb{R}^k$  as well as a polynomial  $P \in \mathbb{R}[X_1, \dots, X_k]$ , and outputs the barcodes up to dimension  $\ell$  for some fixed  $\ell \geq 0$  of the filtration of  $S$  by the sub-level sets of the function  $P$  on  $S$ , thereby generalizing the algorithm in [6] for computing the first  $\ell$  Betti numbers with a similar complexity. There are several intermediate steps needed to achieve this goal. These intermediate steps have been used recently in other applications (that we mention in Section 1.3 below) and hence could be of independent interest. We outline them below.

**1.3. Summary of the main contributions.** We summarize the main contributions of the paper as follows.

1. We reformulate the definition of barcodes in order to treat continuous as well as finite filtrations in a uniform manner. This is important in the current application since we consider filtrations of semi-algebraic sets by polynomial functions which are by nature examples of continuous filtrations (since they are indexed by  $\mathbb{R}$ ). However, we show that the barcode of this continuous filtration is equal to another finite one (see Propositions 3.1 and 3.3). In order for such an equality to make sense it is important that persistent homology of a filtration should be defined in a uniform way for arbitrary ordered index set. It is possible to have a completely categorical description of persistent homology which applies to very general filtration [13]. We avoid categorical language and give an *elementary definition of barcodes* directly in terms of sub-quotients of homology groups (see

Definition 2.5). We remark here that the basic theory of persistent homology with real parameters in the multi-persistence setting was developed independently (in different ways) in [31] and in [34, 35]. However, we prefer to give a self-contained description which applies directly to the one-dimensional semi-algebraic setting and is suitable from our algorithmic view-point.

2. We give a definition of barcodes for semi-algebraic maps which are not necessarily proper (Definition 2.8) generalizing the one for proper maps – and we believe that this could form the basis of generalizing the results of the current papers to arbitrary semi-algebraic sets and maps. Similar ideas appear in [33, Examples 15.11 and 15.14], but our definition is adapted towards applications in real algebraic geometry.
3. By an application of a standard theorem in real algebraic geometry (Hardt triviality theorem [30]) we can deduce that the topological type of the sub-level sets of a filtration of a semi-algebraic set by a semi-algebraic function changes at only finitely many values of the function. This implies that the barcode of the original filtration is equal to that of a finite filtration (after proper definition of barcodes encompassing both the finite and the continuous case as mentioned earlier). However, an algorithm based on Hardt triviality theorem would inevitably lead to a doubly exponential sized filtration – since the proof of this theorem (see for example proof of [9, Theorem 5.46]) depends on taking semi-algebraic triangulations for which only a doubly exponential complexity algorithm is known to exist. Another important contribution of the current paper is an algorithm with singly exponential complexity (see Algorithm 3 below) for reducing a given continuous filtration of a semi-algebraic set by a polynomial to a filtration of simplicial complexes indexed by a *finite* subset of  $\mathbb{R}$ , such that the barcode of this finite filtration is equal to that of the continuous filtration in dimensions up to  $\ell$ . The two main ingredients for this algorithms are:
  - (a) mathematical techniques introduced in [10] for bounding the number of homotopy types of fibers of a semi-algebraic map;
  - (b) a recent algorithm for efficiently computing simplicial replacements of semi-algebraic sets [7, Theorem 1].

We note that Algorithm 3 has other applications as well. For example, it plays a key role in a recent work on computing a homology basis of the first homology group of a given semi-algebraic set with singly exponential complexity [8].

4. The last (and perhaps the most important) contribution is an algorithm with a singly exponential complexity that computes the barcodes of a semi-algebraic filtration up to dimension  $\ell$  for any fixed  $\ell \geq 0$ . After having reduced to the case of finite semi-algebraic filtration using Algorithm 3, we then compute the barcode of this finite filtration of finite simplicial complexes (cf. Algorithms 4 and 5) using Definition 2.5 and standard algorithms from linear algebra.

We remark that it is plausible that after ensuring the finiteness of the filtration, the last step of computing the barcode could be achieved by an appropriate extension of the algorithm for computing the first few Betti numbers of semi-algebraic sets described in [6]. However, this extension would be non-trivial and we prefer to use directly Algorithm 3 in [7] for which no extension is needed.

We prove the following theorem stated informally below. The formal statement appears later in the paper.

**Theorem** (cf. Theorem 1). *There exists an algorithm (Algorithm 5) that takes as input a description of a closed and bounded semi-algebraic set  $S \subset \mathbb{R}^k$ , and a polynomial  $P \in \mathbb{R}[X_1, \dots, X_k]$ , and outputs the “barcodes” (cf. Definition 2.5 below) in dimensions 0 to  $\ell$  of the filtration of  $S$  by the sub-level sets of the polynomial  $P$ . The complexity of this algorithm is bounded singly exponentially in  $k$  (as a function of the number and degrees of polynomials appearing in the description of  $S$ ).*

The importance of the assumption that the input semi-algebraic subset be closed and bounded is discussed in Section 2.2.1.

**1.4. Definition of complexity.** We will use the following notion of “complexity” in this paper. We follow the same definition as used in the book [9].

**Definition 1.1** (Complexity of algorithms). In our algorithms we will usually take as input quantifier-free first order formulas whose terms are polynomials with coefficients belonging to an ordered domain  $D$  contained in a real closed field  $\mathbb{R}$ . By *complexity of an algorithm* we will mean the number of arithmetic operations and comparisons in the domain  $D$ . If  $D = \mathbb{R}$ , then the complexity of our algorithm will agree with the Blum-Shub-Smale notion of real number complexity [11]. In case,  $D = \mathbb{Z}$ , then we are able to deduce the bit-complexity of our algorithms in terms of the bit-sizes of the coefficients of the input polynomials, and this will agree with the classical (Turing) notion of complexity.

**1.5. Prior and Related Work.** As mentioned earlier, designing algorithms with singly exponential complexity for computing topological invariants of semi-algebraic sets has been at the center of research in algorithmic semi-algebraic geometry over the past decades. We refer the reader to the survey [1] for a history of these developments and contributions of many authors. These algorithms are exact algorithms and work for all inputs. The complexity of an algorithm (see Definition 1.1) is measured in terms of the number of arithmetic operations in the ring  $D$  (and also in terms of the bit sizes if  $D = \mathbb{Z}$ ).

More recently, algorithms for computing Betti numbers of semi-algebraic sets have also been developed in other (more numerical) models of computations [14, 15, 16]. In these papers the authors take a different approach. Working over  $\mathbb{R}$ , and given a well-conditioned semi-algebraic subset  $S \subset \mathbb{R}^k$ , they compute a witness complex whose geometric realization is  $k$ -equivalent to  $S$ . The size of this witness complex is bounded singly exponentially in  $k$ . However, the complexity depends on the condition number of the input (and so this bound is not uniform), and the algorithm will fail for ill-conditioned input when the condition number becomes infinite. This is unlike the kind of algorithms we consider in the current paper, which are supposed to work for all inputs and with uniform complexity upper bounds. So these approaches are not comparable. However, to the best of our knowledge there has not been any attempt to extend the numerical algorithms mentioned above for computing Betti numbers to computing persistent homology of semi-algebraic filtrations.

The rest of the paper is organized as follows. In Section 2, we give the precise statements of the main result after introducing the necessary definitions. In

Section 3, we prove the key proposition (Proposition 3.3) which allows us to efficiently reduce to the case of finite filtrations starting with a continuous one. In Section 4, after introducing certain necessary preliminaries, we describe our algorithm for computing barcodes of semi-algebraic filtrations and analyze its complexity (thereby proving Theorem 1). Finally, in Section 5 we state some open questions and directions for future work in this area.

## 2. PRECISE DEFINITIONS AND STATEMENTS OF THE MAIN RESULTS

In this section, we define precisely persistent homology and barcodes of filtrations in Section 2.1. Then in Section 2.2 we define semi-algebraic filtrations and state the main algorithmic result of the paper (Theorem 1).

**2.1. Persistent homology and barcodes.** Let  $T$  be an ordered set, and  $\mathcal{F} = (X_t)_{t \in T}$ , a tuple of subspaces of  $X$ , such that  $s \leq t \Rightarrow X_s \subset X_t$ . We call  $\mathcal{F}$  a filtration of the topological space  $X$ .

We now recall the definition of the *persistent homology groups* associated to a filtration [23, 39]. Since we only consider homology groups with rational coefficients, all homology groups in what follows are finite dimensional  $\mathbb{Q}$ -vector spaces.

**Notation 1.** For  $s, t \in T, s \leq t$ , and  $p \geq 0$ , we let  $i_p^{s,t} : H_p(X_s) \rightarrow H_p(X_t)$ , denote the homomorphism induced by the inclusion  $X_s \hookrightarrow X_t$ .

**Definition 2.1.** [23] For each triple  $(p, s, t) \in \mathbb{Z}_{\geq 0} \times T \times T$  with  $s \leq t$  the *persistent homology group*,  $H_p^{s,t}(\mathcal{F})$  is defined by

$$H_p^{s,t}(\mathcal{F}) = \text{Im}(i_p^{s,t}).$$

Note that  $H_p^{s,t}(\mathcal{F}) \subset H_p(X_t)$ , and  $H_p^{s,s}(\mathcal{F}) = H_p(X_s)$ .

**Notation 2.** We denote by  $b_p^{s,t}(\mathcal{F}) = \dim_{\mathbb{Q}}(H_p^{s,t}(\mathcal{F}))$ .

Persistent homology measures how long a homology class persists in the filtration, in other words considering the homology classes as topological features, it gives an insight about the time (thinking of the indexing set  $T$  of the filtration as time) that a topological feature appears (or is born) and the time it disappears (or dies). This is made precise as follows.

**Definition 2.2.** For  $s \leq t \in T$ , and  $p \geq 0$ ,

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

*Remark 1.* Note that the homology classes that are born at time  $s$ , and those that are born at time  $s$  and dies at time  $t$ , as defined above are not subspaces of  $H_p(X_s)$ . In order to be able to associate a “multiplicity” to the set of homology classes which are born at time  $s$  and dies at time  $t$  we interpret them as classes in certain *subquotients* of  $H_*(X_s)$  in what follows.

First observe that it follows from Definition 2.1 that for all  $s' \leq s \leq t$  and  $p \geq 0$ ,  $H_p^{s',t}(\mathcal{F})$  is a subspace of  $H_p^{s,t}(\mathcal{F})$ , and both are subspaces of  $H_p(X_t)$ . This is because the homomorphism  $i_p^{s',t} = i_p^{s,t} \circ i_p^{s',s}$ , and so the image of  $i_p^{s',t}$  is contained

in the image of  $i_p^{s,t}$ . It follows that, for  $s \leq t$ , the union of  $\bigcup_{s' < s} H_p^{s',t}(\mathcal{F})$  is an increasing union of subspaces, and is itself a subspace of  $H_p(X_t)$ . In particular, setting  $t = s$ ,  $\bigcup_{s' < s} H_p^{s',s}(\mathcal{F})$  is a subspace of  $H_p(X_s)$ .

With the same notation as above:

**Definition 2.3** (Subspaces of  $H_p(X_s)$ ). For  $s \leq t$ , and  $p \geq 0$ , we define

$$\begin{aligned} M_p^{s,t}(\mathcal{F}) &= \bigcup_{s' < s} (i_p^{s,t})^{-1}(H_p^{s',t}(\mathcal{F})), \\ N_p^{s,t}(\mathcal{F}) &= \bigcup_{s' < s \leq t' < t} (i_p^{s,t'})^{-1}(H_p^{s',t'}(\mathcal{F})), \end{aligned}$$

*Remark 2.* The “meaning” of these subspaces are as follows.

- (a) For every fixed  $s \in T$ ,  $M_p^{s,t}(\mathcal{F})$  is a subspace of  $H_p(X_s)$  consisting of homology classes in  $H_p(X_s)$  which are  
“born before time  $s$ , or born at time  $s$  and dies at  $t$  or earlier”
- (b) Similarly, for every fixed  $s \in T$ ,  $N_p^{s,t}(\mathcal{F})$  is a subspace of  $H_p(X_s)$  consisting of homology classes in  $H_p(X_s)$  which are  
“born before time  $s$ , or born at time  $s$  and dies strictly earlier than  $t$ ”

The dimensions of  $M_p^{s,t}(\mathcal{F})$  and  $N_p^{s,t}(\mathcal{F})$  are given in Eqn. (3.6) and (3.7) in Proposition 3.4 below.

We now define certain *subquotients* of the homology groups of  $H_p(X_s)$ ,  $s \in T$ ,  $p \geq 0$ , in terms of the subspaces defined above in Definition 2.3.

**Definition 2.4** (Subquotients associated to a filtration). For  $s \leq t$ , and  $p \geq 0$ , we define

$$\begin{aligned} P_p^{s,t}(\mathcal{F}) &= M_p^{s,t}(\mathcal{F}) / N_p^{s,t}(\mathcal{F}), \\ P_p^{s,\infty}(\mathcal{F}) &= H_p(X_s) / \bigcup_{s \leq t} M_p^{s,t}(\mathcal{F}). \end{aligned}$$

We will call

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

*Remark 3.* Notice that  $M_p^{s,t}(\mathcal{F}) \subset M_p^{s,t'}(\mathcal{F})$  for  $t \leq t'$ , and hence  $\bigcup_{s \leq t} M_p^{s,t}(\mathcal{F})$  is a subspace of  $H_p(X_s)$ , and  $N_p^{s,t}(\mathcal{F})$  is a subspace of  $M_p^{s,t}(\mathcal{F})$ . Therefore, these subquotients are vector spaces and have well defined dimensions.

Finally, we are able to achieve our goal of defining the multiplicity of a bar as the dimension of an associated vector space and define the barcode of a filtration.

**Definition 2.5** (Persistent multiplicity, barcode). We will denote for  $s \in T$ ,  $t \in T \cup \{\infty\}$ ,

$$(2.1) \quad \mu_p^{s,t}(\mathcal{F}) = \dim P_p^{s,t}(\mathcal{F}),$$

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



Finally, we will call the set

$$(2.2) \quad \mathcal{B}_p(\mathcal{F}) = \{(s, t, \mu_p^{s,t}(\mathcal{F})) \mid \mu_p^{s,t}(\mathcal{F}) > 0\}$$

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

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

*Remark 4.* We remark that the definition of multiplicity given up appears in an abstract setting in [20, Corollary 7.3]. Note also that the notion of *persistent multiplicity* has been defined previously in the context of finite filtrations (see [24]). The definition of  $\mu_p^{s,t}(\mathcal{F})$  given in Eqn. (2.1) generalizes that given in *loc.cit.* in the case of finite filtrations, who defined it using Eqn. (3.5) in Proposition 3.4 stated below. Our definition gives a geometric meaning to this number as a dimension of a certain vector space (a subquotient of  $H_p(X_s)$ ), and we prove that it agrees with that given in *loc.cit.* in Proposition 3.4. Also, it is important to note for what follows that our definition of a barcode applies uniformly to all filtrations with index coming from an ordered set, and we make no additional assumption on the indexing set.

*Remark 5* (Continuous vs finite filtrations). In most applications the filtration  $\mathcal{F}$  is assumed to be finite (i.e. the ordered set  $T$  is finite). Since we are considering filtration of semi-algebraic sets by the sub-level sets of a polynomial function, our filtration is indexed by  $\mathbb{R}$  and is an example of a continuous (infinite) filtration. Nevertheless, we will reduce to the finite filtration case by proving that the barcode of the given filtration is equal to that of a finite filtration. A general theory encompassing both finite and infinite filtrations using a categorical view-point has been developed (see [13, 18]). We avoid using the categorical definitions and the module-theoretic language used in [18]. We will prove directly the equality of the barcodes of the infinite and the corresponding finite filtration (cf. Proposition 3.3) that is important in designing our algorithm, starting from the definition of persistent multiplicities given above.

We now give a concrete example of a barcode associated to a (infinite) filtration.

**Example 1.** Let  $S$  be the two-dimensional torus (topologically  $\mathbf{S}^1 \times \mathbf{S}^1$ ) embedded in  $\mathbb{R}^3$ , and  $\mathcal{F}$  be the filtration of the torus by the sub-level sets of the height function (depicted in Figure 1(a)). We denote by  $S_{\leq t}$  the subset of the torus having “height”  $\leq t$ .

We consider homology in dimensions 0, 1 and 2.

Informally, one observes that a 0-dimensional homology class is born at time  $t_0$  which never dies. There are two 1-dimensional homology classes, the horizontal loop born at time  $t_2$  and the vertical loop born at time  $t_4$ , which also never die. Lastly, there is a 2-dimensional homology class born at time  $t_5$  which never dies. Since there are no homology classes of the same dimension being born and dying at the same time, multiplicities in all the cases are 1.

More formally, following Definitions 2.3, 2.4 and 2.5, we obtain:

(Case  $p = 0$ ) If  $t_0 \leq t < \infty$  then (using Definition 2.3)

$$M_0^{t_0,t}(\mathcal{F}) = 0,$$

and hence (using Definitions 2.4 and 2.5)

$$P_0^{t_0,t}(\mathcal{F}) = 0, \text{ and } \mu_0^{t_0,t}(\mathcal{F}) = 0.$$



On the other hand,

$$P_0^{t_0, \infty}(\mathcal{F}) = H_0(S_{\leq t_0}),$$

implying

$$\mu_0^{t_0, \infty}(\mathcal{F}) = 1.$$

(Case p = 1) For  $t_2 \leq t < \infty$ ,

$$M_1^{t_2, t}(\mathcal{F}) = 0,$$

and hence

$$P_1^{t_2, t}(\mathcal{F}) = 0, \text{ and } \mu_1^{t_2, t}(\mathcal{F}) = 0.$$

Moreover,

$$P_1^{t_2, \infty}(\mathcal{F}) = H_1(S_{\leq t_2}),$$

and therefore,

$$\mu_1^{t_2, \infty}(\mathcal{F}) = 1.$$

For  $t_4 \leq t < \infty$ ,

$$M_1^{t_4, t}(\mathcal{F}) = N_1^{t_4, t}(\mathcal{F}) = H_1(S_{< t_4}),$$

and hence

$$P_1^{t_4, t}(\mathcal{F}) = 0, \text{ and } \mu_1^{t_4, t}(\mathcal{F}) = 0.$$

Moreover,

$$P_1^{t_4, \infty}(\mathcal{F}) = H_1(S_{\leq t_4})/H_1(S_{< t_4}),$$

and therefore

$$\mu_1^{t_4, \infty}(\mathcal{F}) = 1.$$

(Case p = 2) For  $t_5 \leq t < \infty$ ,

$$M_2^{t_5, t}(\mathcal{F}) = 0,$$

and hence

$$P_2^{t_5, t}(\mathcal{F}) = 0, \text{ and } \mu_2^{t_5, t}(\mathcal{F}) = 0.$$

Moreover,

$$P_2^{t_5, \infty}(\mathcal{F}) = H_2(S),$$

and therefore

$$\mu_2^{t_5, \infty}(\mathcal{F}) = 1.$$

Therefore the barcodes are as follows (using Eqn. (2.2)).

$$\begin{aligned} \mathcal{B}_0(\mathcal{F}) &= \{(t_0, +\infty, 1)\}, \\ \mathcal{B}_1(\mathcal{F}) &= \{(t_2, +\infty, 1), (t_4, +\infty, 1)\}, \\ \mathcal{B}_2(\mathcal{F}) &= \{(t_5, +\infty, 1)\}. \end{aligned}$$

Figure 1(b) illustrates the corresponding bars. Notice that even though the filtration  $\mathcal{F}$  is an infinite filtration indexed by  $\mathbb{R}$ , the barcodes,  $\mathcal{B}_p(\mathcal{F})$ , are finite.

The main type of filtration that we consider in this paper is filtration of semi-algebraic sets by the sub-level sets of continuous semi-algebraic functions – which we define below.

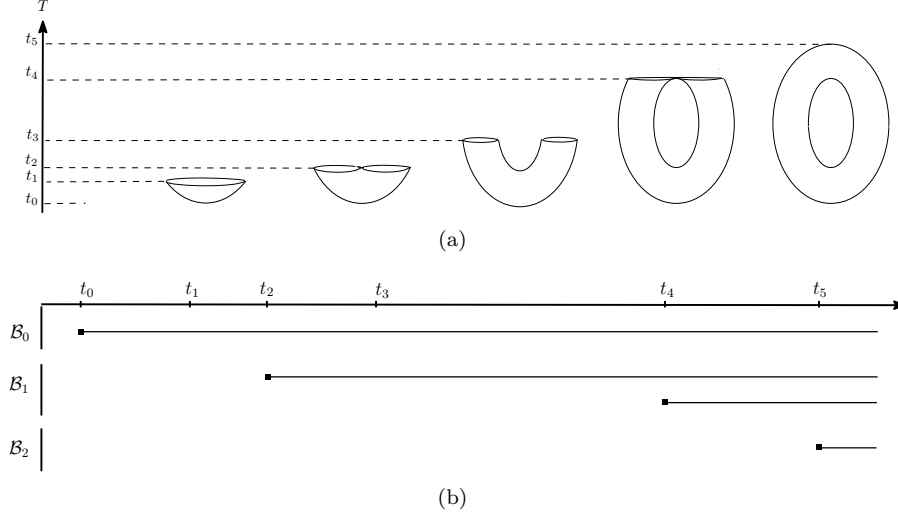


FIGURE 1. (a) Torus filtered by the sub-level sets of the height function, (b) corresponding barcodes for homology classes of dimension 0, 1 and 2.

### 2.1.1. $\mathcal{P}$ -formulas and $\mathcal{P}$ -semi-algebraic sets.

**Notation 3** (Realizations,  $\mathcal{P}$ -,  $\mathcal{P}$ -closed semi-algebraic sets). For any finite set of polynomials  $\mathcal{P} \subset \mathbb{R}[X_1, \dots, X_k]$ , we call any quantifier-free first order formula  $\phi$  with atoms,  $P = 0, P < 0, P > 0, P \in \mathcal{P}$ , to be a  $\mathcal{P}$ -formula. Given any semi-algebraic subset  $Z \subset \mathbb{R}^k$ , we call the realization of  $\phi$  in  $Z$ , namely the semi-algebraic set

$$\mathcal{R}(\phi, Z) := \{\mathbf{x} \in Z \mid \phi(\mathbf{x})\}$$

a  $\mathcal{P}$ -semi-algebraic subset of  $Z$ .

We say that a quantifier-free formula  $\phi$  is *closed* if it is a formula in disjunctive normal form with no negations, and with atoms of the form  $P \geq 0, P \leq 0$  (resp.  $P > 0, P < 0$ ), where  $P \in \mathcal{D}[X_1, \dots, X_k]$ . If the set of polynomials appearing in a closed (resp. open) formula is contained in a finite set  $\mathcal{P}$ , we will call such a formula a  $\mathcal{P}$ -closed formula, and we call the realization,  $\mathcal{R}(\phi)$ , a  $\mathcal{P}$ -closed semi-algebraic set.

**2.2. Semi-algebraic filtrations.** We consider the algorithmic problem of computing the dimensions of persistent homology groups and barcodes of the filtration induced on a given semi-algebraic set by a polynomial function.

**Definition 2.6.** Let  $S \subset \mathbb{R}^k$  be a semi-algebraic set and  $P : S \rightarrow \mathbb{R}$  a continuous semi-algebraic map.

For  $t \in \mathbb{R} \cup \{\pm\infty\}$ , let

$$S_{P \leq t} = \{x \in S \mid P(x) \leq t\}.$$

Then,  $(S_{P \leq t})_{t \in \mathbb{R} \cup \{\pm\infty\}}$  is a filtration of the semi-algebraic set  $S$  indexed by  $\mathbb{R} \cup \{\pm\infty\}$ , and we will denote this filtration by  $\mathcal{F}(S, P)$ .

**Notation 4.** For  $p \geq 0$ , we will denote

$$\mathcal{B}_p(S, P) = \mathcal{B}_p(\mathcal{F}(S, P)).$$

*Remark 6.* In the definition of  $\mathcal{B}_p(\mathcal{F}(S, P))$  we need to specify the homology theory we are using. For a semi-algebraic set  $X$  defined over an arbitrary real closed field  $\mathbb{R}$  we take homology groups  $H_*(X) = H_*(X, \mathbb{Q})$  as defined in [21, (3.6), page 141]. It agrees with singular homology in case  $\mathbb{R} = \mathbb{R}$ .

*Remark 7.* Note that many filtrations commonly used in computational topology are examples of filtrations of semi-algebraic sets by polynomial functions as defined above. One example of this is the well-known *Čech-complex* [22] which can be described as follows.

Let  $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\}$  be a finite set of points in  $\mathbb{R}^k$ , and let  $S \subset \mathbb{R}^{k+1}$ , be the semi-algebraic set defined by the formula

$$\phi(X_1, \dots, X_k, T) := \bigvee_{i=1}^n \left( |\mathbf{X} - \mathbf{x}^{(i)}|^2 - T \leq 0 \right),$$

where  $\mathbf{X} = (X_1, \dots, X_k)$ . Let  $P = T$ . Then, the filtration  $\mathcal{F}(S, P)$  is homeomorphic to the filtration obtained by taking unions of balls of growing radius centered at  $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\}$ . This latter filtration plays a very important role in applications (for example, in analyzing the topological structure of point-cloud data).

*Remark 8.* Note also that the barcode of a polynomial function restricted to a semi-algebraic set  $S$  gives important topological information about the function  $P$  on  $S$ . It allows one to define a  $p$ -dimensional distance between two such polynomial functions restricted to  $S$ , by defining a notion of distance between two barcodes. Various distances have been proposed but the most commonly used one is the so called “bottle-neck distance” [24]. An algorithm with singly exponential complexity for computing the barcode of a polynomial also gives an algorithm with singly exponential complexity for computing such distances as well. To our knowledge the algorithmic problem of computing barcodes of polynomial functions on semi-algebraic sets have not been considered prior to our work.

We prove the following theorem.

**Theorem 1.** *There exists an algorithm that takes as input:*

1. *a finite set of polynomials,  $\mathcal{P} \subset \mathbb{D}[X_1, \dots, X_k]$ ;*
2. *a  $\mathcal{P}$ -closed formula  $\phi$  such that  $\mathcal{R}(\phi)$  is bounded;*
3. *a polynomial  $P \in \mathbb{D}[X_1, \dots, X_k]$ ;*
4.  *$\ell \geq 0$ ;*

*and computes  $\mathcal{B}_p(\mathcal{R}(\phi), P)$ , for  $0 \leq p \leq \ell$ . The complexity of the algorithm is bounded by  $(sd)^{k^{O(\ell)}}$ , where  $s = \text{card}(\mathcal{P})$ , and  $d$  is the maximum amongst the degrees of  $P$  and the polynomials in  $\mathcal{P}$ .*

2.2.1. *Barcodes of non-proper maps.* Notice that in Theorem 1 we only consider semi-algebraic sets  $S$  which are closed and bounded. In particular, this implies that any continuous semi-algebraic function on  $S$  is a proper map  $S \rightarrow \mathbb{R}$  (i.e. the inverse image of a closed and bounded semi-algebraic set is closed and bounded).

One reason to assume the properness is that for non-proper semi-algebraic maps  $P : S \rightarrow \mathbb{R}$ , the barcode  $\mathcal{B}_p(S, P)$  may not reflect the topology of  $S$  as illustrated in the following example (see also [33, Examples 15.11 and 15.14]).

**Example 2.** Let  $S \subset \mathbb{R}^2$  be the (unbounded) semi-algebraic set defined by the formula

$$\phi := (0 < X_1 < 1) \wedge (X_1(X_1 - 1)X_2 - 1 = 0)$$

(depicted in Figure 2), and let  $P = X_1$ . Consider the semi-algebraic filtration  $\mathcal{F}(S, P)$ . Note that  $P$  restricted to  $S$  is not a proper semi-algebraic map ( $P^{-1}([0, 1])$  is not bounded).

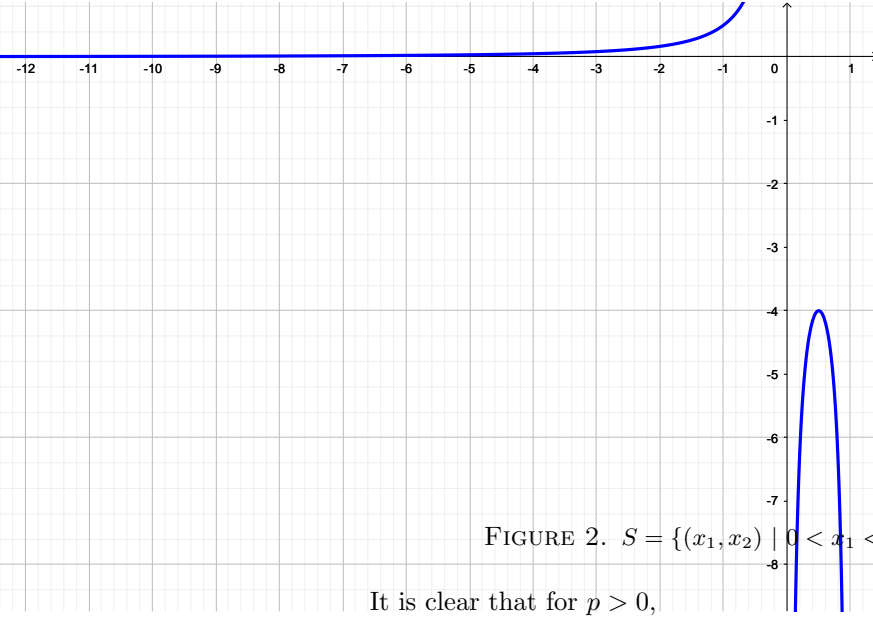


FIGURE 2.  $S = \{(x_1, x_2) \mid 0 < x_1 < 1, x_1(x_1 - 1)x_2 - 1 = 0\}$

It is clear that for  $p > 0$ ,

$$\mathcal{B}_p(S, P) = \emptyset.$$

We claim that even for  $p = 0$  (contrary to the expectation)

$$\mathcal{B}_p(S, P) = \emptyset.$$

To see this observe that for all  $s \leq 0$ ,  $t \geq s$ , we have that

$$\begin{aligned} M_0^{s,t}(\mathcal{F}(S, P)) &= \bigcup_{s' < s} (i_p^{s,t})^{-1}(\mathcal{H}_0^{s',t}(\mathcal{S})) \\ &= 0, \\ \mathcal{H}_0(S_{P \leq s}) &= 0, \end{aligned}$$

since  $S_{P \leq s} = \emptyset$  for  $s \leq 0$ . This shows that

$$(2.3) \quad \mu_0^{s,t}(\mathcal{F}(S, P)) = 0, s \leq 0, t \geq s.$$

For  $s > 0$ , and  $t \geq s$ , it follows from Definition 2.5 that

$$M_0^{s,t}(\mathcal{F}(S, P)) = N_0^{s,t}(\mathcal{F}(S, P)) = \mathcal{H}_0(S_{P \leq s}),$$

proving that

$$(2.4) \quad \mu_0^{s,t}(\mathcal{F}(S, P)) = 0, s > 0, t \geq s.$$

Together Eqns. (2.3) and (2.4) imply that

$$\mathcal{B}_0(S, P) = \emptyset.$$

In order to have a more reasonable definition of barcodes (and allow “bars” which have open endpoints) we propose the following definition. We use two notions from real algebraic geometry – that of the real spectrum and the real closed extension of  $\mathbb{R}$  by the field of Puiseux series.

Let  $S \subset \mathbb{R}^k$  be an arbitrary semi-algebraic set and  $P : S \rightarrow \mathbb{R}$  a continuous semi-algebraic function. We define a new filtration  $\tilde{\mathcal{F}}(S, P)$  as follows.

The indexing set of the new filtration will be the set

$$\tilde{\mathbb{R}} = \{-\infty, +\infty\} \cup \bigcup_{x \in \mathbb{R}} \{x_-, x, x_+\},$$

on which a total order is specified by

$$-\infty < x_- < x < x_+ < y_- < y < y_+ < \infty,$$

for all  $x < y$  in  $\mathbb{R}$ . (The ordered set  $\tilde{\mathbb{R}}$  is the *real spectrum* of the ring  $\mathbb{R}[X]$  – see for example [12, page 134]).

We now define the filtration  $\tilde{\mathcal{F}}(S, P)$ .

**Definition 2.7** (Filtration for semi-algebraic maps not necessarily proper). For  $\tilde{t} \in \tilde{\mathbb{R}}$  define

$$\begin{aligned} \tilde{S}_{\tilde{t}} &= \text{ext}(S, \mathbb{R}\langle \varepsilon \rangle)_{P=-1/\varepsilon}, \text{ if } \tilde{t} = -\infty, \\ &= \text{ext}(S, \mathbb{R}\langle \varepsilon \rangle)_{P \leq t-\varepsilon}, \text{ if } \tilde{t} = t_-, t \in \mathbb{R}, \\ &= \text{ext}(S, \mathbb{R}\langle \varepsilon \rangle)_{P \leq t}, \text{ if } \tilde{t} = t \in \mathbb{R}, \\ &= \text{ext}(S, \mathbb{R}\langle \varepsilon \rangle)_{P \leq t+\varepsilon}, \text{ if } \tilde{t} = t_+, t \in \mathbb{R}, \\ &= \text{ext}(S, \mathbb{R}\langle \varepsilon \rangle)_{P=1/\varepsilon}, \text{ if } \tilde{t} = +\infty. \end{aligned}$$

(where  $\mathbb{R}\langle \varepsilon \rangle$  is the field of algebraic Puiseux series in  $\varepsilon$ , and  $\text{ext}(\cdot, \mathbb{R}\langle \varepsilon \rangle)$  denotes the extension of a semi-algebraic subset of  $\mathbb{R}^k$  to  $\mathbb{R}\langle \varepsilon \rangle^k$  – see Notation 5 and Notation 7).

**Definition 2.8** (Barcode for filtration induced by a semi-algebraic map not necessarily proper). For  $S \subset \mathbb{R}^k$  an arbitrary semi-algebraic set and  $P : S \rightarrow \mathbb{R}$  a continuous semi-algebraic function, we define

$$\tilde{\mathcal{B}}_p(S, P) = \mathcal{B}_p(\tilde{\mathcal{F}}(S, P)).$$

It is easy now to verify that for the pair  $S, P$  in Example 2

$$\tilde{\mathcal{B}}_0(S, P) = \{(0_+, +\infty, 1)\}.$$

Note that

$$\tilde{\mathcal{B}}_p(S, P) \subset \tilde{\mathbb{R}} \times \tilde{\mathbb{R}} \times \mathbb{Z}_{>0}.$$

Using Hardt triviality theorem, one can deduce that  $\tilde{\mathcal{B}}_p(S, P)$  is a finite set. We will formally prove this statement later for proper semi-algebraic maps (see Proposition 3.1).

The barcode for a proper semi-algebraic map takes its value in  $\mathbb{R} \times \mathbb{R} \times \mathbb{Z}_{>0}$  which is properly contained in the  $\tilde{\mathbb{R}} \times \tilde{\mathbb{R}} \times \mathbb{Z}_{>0}$ . It is not difficult to prove that in case  $P : S \rightarrow \mathbb{R}$  is a proper semi-algebraic map, the new definition of barcode agrees with the previous one.

We record the above mentioned facts in the following proposition for future reference and omit the proofs. We will not use it in this paper since we restrict ourselves to the proper case.

**Proposition 2.1.** *For any continuous semi-algebraic map  $P : S \rightarrow \mathbb{R}$  and for all  $p \geq 0$ ,  $\tilde{\mathcal{B}}(S, P)$  is a finite set. Moreover, if  $P$  is a proper semi-algebraic map, then for all  $p \geq 0$ ,*

$$\tilde{\mathcal{B}}_p(S, P) = \mathcal{B}_p(S, P).$$

*Proof.* Omitted. □

## 3. CONTINUOUS TO FINITE FILTRATION

In this section we describe how to efficiently reduce the problem of computing the barcode of a continuous semi-algebraic filtration to that of a finite filtration of semi-algebraic sets. The mathematical results are encapsulated in Propositions 3.1 and 3.3 stated and proved in Section 3.1. Then in Section 3.2 we prove a formula used to compute the barcode of a finite filtration (Proposition 3.4). This formula is not new (see [24, page 152][24]), however, it is important to deduce that from our new definition of barcodes.

Recall that we are interested in the persistent homology of filtrations of semi-algebraic sets by the sub-level sets of a polynomial. Recall also (cf. Definition 2.6) that for a closed and bounded semi-algebraic set  $S \subset \mathbb{R}^k$ ,  $P \in \mathbb{R}[X_1, \dots, X_k]$ , and  $t \in \mathbb{R} \cup \{\pm\infty\}$ , we denote the filtration

$$(S_{P \leq t} = \{x \in S \mid P(x) \leq t\})_{t \in \mathbb{R} \cup \{\pm\infty\}}$$

by  $\mathcal{F}(S, P)$ .

Our first observation is that, even though the indexing set  $\mathbb{R} \cup \{\pm\infty\}$  is infinite, for each  $p \geq 0$ , the barcode  $\mathcal{B}_p(\mathcal{F}(S, P))$  is a finite set (cf. Example 1).

**Proposition 3.1.** *For each  $p \geq 0$ , the cardinality of  $\mathcal{B}_p(\mathcal{F}(S, P))$  is finite.*

**3.1. Reduction to the case of a finite filtration.** We will now prove a result (cf. Proposition 3.3 below) from which Proposition 3.1 will follow. Our strategy is to identify a finite set of values  $\{s_0, \dots, s_M\} \subset \mathbb{R}$ , such that the semi-algebraic homotopy type of the increasing family  $S_{P \leq t}$  (as  $t$  goes from  $-\infty$  to  $\infty$ ), can change only when  $t$  crosses one of the  $s_i$ 's. This would imply that the barcode,  $\mathcal{B}_p(\mathcal{F}(S, P))$ , of the infinite filtration  $\mathcal{F}(S, P)$ , is *equal* to the barcode of the *finite filtration*  $\emptyset \subset S_{P \leq s_0} \subset \dots \subset S_{P \leq s_M} \subset S$  (cf. Proposition 3.3 below). In addition, we will obtain a bound on the number  $M$  in terms of the number of polynomials appearing in the definition of  $S$  and their degrees, as well as the degree of the polynomial  $P$ . The technique used in the proofs of these results are adaptations of the technique used in the proof of the main result (Theorem 2.1) in [10], which gives a singly exponential bound on the number of distinct homotopy types amongst the fibers of a semi-algebraic map in [10]. We need a slightly different statement than that of Theorem 2.1 in [10]. However, our situation is simpler since we only need the result for maps to  $\mathbb{R}$  (rather than to  $\mathbb{R}^n$  as is the case in [10, Theorem 2.1]).

**3.1.1. Real closed extensions and Puiseux series.** We will need some properties of Puiseux series with coefficients in a real closed field. We refer the reader to [9] for further details.

**Notation 5.** For  $\mathbb{R}$  a real closed field we denote by  $\mathbb{R}\langle\varepsilon\rangle$  the real closed field of algebraic Puiseux series in  $\varepsilon$  with coefficients in  $\mathbb{R}$ . We use the notation  $\mathbb{R}\langle\varepsilon_1, \dots, \varepsilon_m\rangle$  to denote the real closed field  $\mathbb{R}\langle\varepsilon_1\rangle\langle\varepsilon_2\rangle\cdots\langle\varepsilon_m\rangle$ . Note that in the unique ordering of the field  $\mathbb{R}\langle\varepsilon_1, \dots, \varepsilon_m\rangle$ ,  $0 < \varepsilon_m \ll \varepsilon_{m-1} \ll \dots \ll \varepsilon_1 \ll 1$ .

**Notation 6.** For elements  $x \in \mathbb{R}\langle\varepsilon\rangle$  which are bounded over  $\mathbb{R}$  we denote by  $\lim_\varepsilon x$  to be the image in  $\mathbb{R}$  under the usual map that sets  $\varepsilon$  to 0 in the Puiseux series  $x$ .

**Notation 7.** If  $\mathbb{R}'$  is a real closed extension of a real closed field  $\mathbb{R}$ , and  $S \subset \mathbb{R}^k$  is a semi-algebraic set defined by a first-order formula with coefficients in  $\mathbb{R}$ , then we will denote by  $\text{ext}(S, \mathbb{R}') \subset \mathbb{R}'^k$  the semi-algebraic subset of  $\mathbb{R}'^k$  defined by the

same formula. It is well known that  $\text{ext}(S, R')$  does not depend on the choice of the formula defining  $S$  [9, Proposition 2.87].

**Notation 8.** Suppose  $R$  is a real closed field, and let  $X \subset R^k$  be a closed and bounded semi-algebraic subset, and  $X^+ \subset R\langle \varepsilon \rangle^k$  be a semi-algebraic subset bounded over  $R$ . Let for  $t \in R, t > 0$ ,  $\tilde{X}_t^+ \subset R^k$  denote the semi-algebraic subset obtained by replacing  $\varepsilon$  in the formula defining  $X^+$  by  $t$ , and it is clear that for  $0 < t \ll 1$ ,  $\tilde{X}_t^+$  does not depend on the formula chosen. We say that  $X^+$  is *monotonically decreasing to  $X$* , and denote  $X^+ \searrow X$  if the following conditions are satisfied.

- (a) for all  $0 < t < t' \ll 1$ ,  $\tilde{X}_t^+ \subset \tilde{X}_{t'}^+$ ;
- (b)

$$\bigcap_{t>0} \tilde{X}_t^+ = X;$$

or equivalently  $\lim_{\varepsilon} X^+ = X$ .

More generally, if  $X \subset R^k$  be a closed and bounded semi-algebraic subset, and  $X^+ \subset R\langle \varepsilon_1, \dots, \varepsilon_m \rangle^k$  a semi-algebraic subset bounded over  $R$ , we will say  $X^+ \searrow X$  if and only if

$$X_{m+1}^+ = X^+ \searrow X_m^+, X_m^+ \searrow X_{m-1}^+, \dots, X_2^+ \searrow X_1^+ = X,$$

where for  $i = 1, \dots, m$ ,  $X_i^+ = \lim_{\varepsilon_i} X_{i+1}^+$ .

The following lemma will be useful later.

**Lemma 3.1.** *Let  $X \subset R^k$  be a closed and bounded semi-algebraic subset, and  $X^+ \subset R\langle \bar{\varepsilon}_1, \dots, \bar{\varepsilon}_m \rangle^k$  a semi-algebraic subset bounded over  $R$ , such that  $X^+ \searrow X$ . Then,  $\text{ext}(X, R\langle \bar{\varepsilon}_1, \dots, \bar{\varepsilon}_m \rangle)$  is semi-algebraic deformation retract of  $X^+$ .*

*Proof.* See proof of Lemma 16.17 in [9]. □

**3.1.2. Outline of the reduction.** Before delving into the detail we first give an outline of the main idea behind the reduction to the finite filtration case. The key mathematical result that we need is the following. Given a semi-algebraic subset  $X \subset R^{k+1}$ , obtain a semi-algebraic partition of  $R \cup \{\pm\infty\}$  into points  $-\infty = s_{-1} < s_0 < s_1 < \dots < s_M < s_{M+1} = \infty$ , and open intervals  $(s_i, s_{i+1})$ ,  $-1 \leq i \leq M$ , such that the homotopy type of  $X_t = X \cap \pi_{k+1}^{-1}(s_i, s_{i+1})$  stays constant over each open interval  $(s_i, s_{i+1})$  (here  $\pi_{k+1}$  denotes the projection on the last coordinate). In our application the fibers  $X_t$  will be a non-decreasing in  $t$  (in fact,  $X_t$  will be equal to  $S_{P \leq t}$ ) but we do not need this property to hold for obtaining the partition mentioned above.

The following example is illustrative.

Suppose that  $X \subset R^2$  is a singular curve shown in blue in Figure 3. We define a semi-algebraic tubular neighborhood  $X^*(\varepsilon)$  of  $X$  using an infinitesimal  $\varepsilon$  (shown in red), whose boundary has good algebraic properties – namely, in this case a finite number of critical values  $t_0 < t_1 < \dots < t_5$  for the projection map onto the chosen coordinate  $X_{k+1}$  which is shown as  $X_1$  in the figure. The  $t_i$ 's give a partition of  $R\langle \varepsilon \rangle$  rather than that of  $R$ , and over each interval  $(t_i, t_{i+1})$  the semi-algebraic homeomorphism type of  $X^*(\varepsilon)$  (but not necessarily the semi-algebraic homotopy type of  $\text{ext}(X, R\langle \varepsilon \rangle)$ ) stay constant. Clearly this partition does not have the homotopy invariance property with respect to the set  $\text{ext}(X, R\langle \varepsilon \rangle)$ . However,



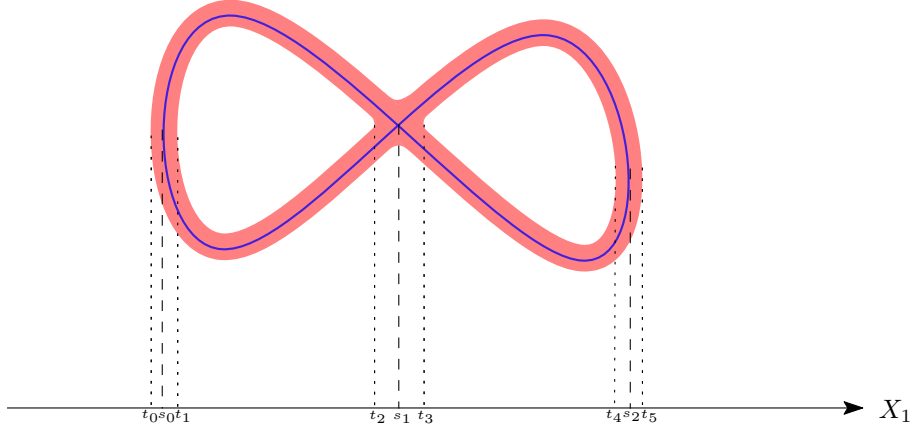


FIGURE 3. Homotopy types of fibers

the intervals  $(t_1, t_2) \cap R = (s_0, s_1)$  and  $(t_3, t_4) \cap R = (s_1, s_2)$  does have the required property with respect to  $X$ , and the points  $s_0, s_1, s_2$  gives us the required partition.

In the general case the definition of the tube  $X^*(\varepsilon)$  is more involved and uses more than one infinitesimal (cf. Notation 10). The set of points corresponding to the  $t_i$ 's in the above example is defined precisely in Proposition 3.2 where the important property of the partition of  $R(\varepsilon)$  they induce is also proved. The passage from the  $t_i$ 's to the  $s_i$ 's and the important property satisfied by the  $s_i$ 's is described in Lemma 3.5. The finite set of values  $\{s_0, \dots, s_M\} \subset R$  is then used to define a finite filtration of the given semi-algebraic set, and the fact that this finite filtration has the same barcode as the infinite filtration we started with is proved in Proposition 3.3. Proposition 3.3 immediately implies Proposition 3.1.

There are several further technicalities involved in converting the above construction into an efficient algorithm. These are explained in Section 4. The complexity of the whole procedure is bounded singly exponentially.

3.1.3. *Proof of Proposition 3.1.* We begin by fixing some notation.

**Notation 9.** For  $\mathcal{Q} \subset R[X_1, \dots, X_k]$  we will denote by

$$Z(\mathcal{Q}, R^k) = \{x \in R^k \mid \bigwedge_{Q \in \mathcal{Q}} Q(x) = 0\}.$$

For  $Q \in R[X_1, \dots, X_k]$ , we will denote by  $Z(Q, R^k) = \{x \in R^k \mid Q(x) = 0\}$ .

**Definition 3.1.** Let  $\mathcal{Q}$  be a finite subset of  $R[X_1, \dots, X_k]$ . A *sign condition* on  $\mathcal{Q}$  is an element of  $\{0, 1, -1\}^{\mathcal{Q}}$ . We say that  $\mathcal{Q}$  *realizes* the sign condition  $\sigma$  at  $x \in R^k$  if

$$\bigwedge_{Q \in \mathcal{Q}} \text{sign}(Q(x)) = \sigma(Q).$$

The *realization* of the sign condition  $\sigma$  is

$$\mathcal{R}(\sigma) = \{x \in R^k \mid \bigwedge_{Q \in \mathcal{Q}} \text{sign}(Q(x)) = \sigma(Q)\}.$$

The sign condition  $\sigma$  is *realizable* if  $\mathcal{R}(\sigma)$  is non-empty. We denote by  $\text{Sign}(\mathcal{Q})$  the set of realizable sign conditions of  $\mathcal{Q}$ .

Let  $R \in \mathbb{R}$  with  $R > 0$ , and let

$$\mathcal{P} = \{P_0, P_1, \dots, P_s\} \subset \mathbb{R}[X_1, \dots, X_k],$$

with  $P_0 = X_1^2 + \dots + X_k^2 - R$ . Let  $P \in \mathbb{R}[X_1, \dots, X_k]$ , and also let  $\phi$  be a closed  $(\mathcal{P} - \{P_0\})$ -formula, and  $\tilde{\phi}$  be  $\phi \wedge (P_0 \leq 0) \wedge (P - Y \leq 0)$ , where  $Y$  is a new variable. So  $\phi$  is a  $(\mathcal{P} \cup \{P - Y\})$ -closed formula. Let  $P_{s+1} = P - Y$ .

**Notation 10.** For  $\bar{\varepsilon} = (\varepsilon_0, \dots, \varepsilon_{s+1})$ , we denote by  $\phi^*(\bar{\varepsilon})$ , the  $\mathcal{P}^*(\bar{\varepsilon})$ -closed formula obtained by replacing each occurrence of  $P_i \geq 0$  in  $\phi$  by  $P_i + \varepsilon_i \geq 0$  (resp.  $P_i \leq 0$  in  $\tilde{\phi}$  by  $P_i - \varepsilon_i \leq 0$ ) for  $0 \leq i \leq s+1$ , where

$$\mathcal{P}^*(\bar{\varepsilon}) = \bigcup_{0 \leq i \leq s+1} \{P_i + \varepsilon_i, P_i - \varepsilon_i\}.$$

Observe that

$$S^*(\bar{\varepsilon}) := \mathcal{R}(\phi^*(\bar{\varepsilon})) \subset \mathbb{R}\langle \bar{\varepsilon} \rangle^{k+1}$$

is a  $\mathcal{P}^*(\bar{\varepsilon})$ -closed semi-algebraic set, and we define  $\Sigma_\phi \subset \{-1, 0, 1\}^{\mathcal{P}^*(\bar{\varepsilon})}$  by

$$(3.1) \quad S^*(\bar{\varepsilon}) = \bigcup_{\sigma \in \Sigma_\phi, \mathcal{R}(\sigma) \neq \emptyset} \mathcal{R}(\sigma).$$

**Lemma 3.2.** *For each  $\mathcal{Q} \subset \mathcal{P}^*(\bar{\varepsilon})$ ,  $Z(\mathcal{Q}, \mathbb{R}\langle \bar{\varepsilon} \rangle^{k+1})$  is either empty or is a non-singular  $(k+1 - \text{card}(\mathcal{Q}))$ -dimensional real variety such that at every point  $(x_1, \dots, x_k, y) \in Z(\mathcal{Q}, \mathbb{R}\langle \bar{\varepsilon} \rangle^{k+1})$ , the  $(\text{card}(\mathcal{Q}) \times (k+1))$ -Jacobi matrix,*

$$\left( \frac{\partial P}{\partial X_i}, \frac{\partial P}{\partial Y} \right)_{P \in \mathcal{Q}, 1 \leq i \leq k}$$

has the maximal rank  $\text{card}(\mathcal{Q})$ .

*Proof.* See [10]. □

Now let  $\pi_{k+1} : \mathbb{R}\langle \bar{\varepsilon} \rangle^{k+1} \rightarrow \mathbb{R}\langle \bar{\varepsilon} \rangle$  denote the projection to the last (i.e. the  $Y$ ) coordinate, and  $\pi_{[1,k]} : \mathbb{R}\langle \bar{\varepsilon} \rangle^{k+1} \rightarrow \mathbb{R}\langle \bar{\varepsilon} \rangle^k$  denote the projection to the first (i.e.  $(X_1, \dots, X_k)$ )  $k$  coordinates.

For any semi-algebraic subset  $S \subset \mathbb{R}\langle \bar{\varepsilon} \rangle^{k+1}$ , and  $T \subset \mathbb{R}\langle \bar{\varepsilon} \rangle$ , we denote by  $S_T = \pi_{[1,k]}(S \cap \pi_{k+1}^{-1}(T))$ . For  $t \in \mathbb{R}\langle \bar{\varepsilon} \rangle$ , we will denote by  $S_{\leq t} = S_{(-\infty, t]}$ , and  $S_t = S_{\{t\}}$ .

**Notation 11** (Critical points and critical values). For  $\mathcal{Q} \subset \mathcal{P}^*(\bar{\varepsilon})$ , we denote by  $\text{Crit}(\mathcal{Q})$  the subset of  $Z(\mathcal{Q}, \mathbb{R}\langle \bar{\varepsilon} \rangle^{k+1})$  at which the the Jacobian matrix,

$$\left( \frac{\partial P}{\partial X_i} \right)_{P \in \mathcal{Q}, 1 \leq i \leq k}$$

is not of the maximal possible rank. We denote  $\text{crit}(\mathcal{Q}) = \pi(\text{Crit}(\mathcal{Q}))$ .

**Lemma 3.3.** *The set*

$$\bigcup_{\mathcal{Q} \subset \mathcal{P}^*(\bar{\varepsilon})} \text{crit}(\mathcal{Q})$$

*is finite.*

*Proof.* Follows from Lemma 3.2 and the semi-algebraic Sard's lemma (see for example [9, Theorem 5.56]). □

**Lemma 3.4.** *The partitions*

$$\begin{aligned} \mathbf{R}^{k+1} &= \bigcup_{\sigma \in \text{Sign}(\mathcal{P}^*(\bar{\varepsilon}))} \mathcal{R}(\sigma), \\ S^*(\bar{\varepsilon}) &= \bigcup_{\sigma \in \Sigma_\phi} \mathcal{R}(\sigma), \end{aligned}$$

are compatible Whitney stratifications of  $\mathbf{R}^{k+1}$  and  $S^*(\bar{\varepsilon})$  respectively.

*Proof.* Follows directly from the definition of Whitney stratification (see [28, 19]), and Lemma 3.2.  $\square$

We are now in a position to prove the key mathematical result that allows us to reduce the filtration of a semi-algebraic set by the sub-level sets of a polynomial to the case of a finite filtration.

**Proposition 3.2.** *Suppose*

$$\bigcup_{\mathcal{Q} \subset \mathcal{P}^*(\bar{\varepsilon})} \text{crit}(\mathcal{Q}) = \{t_0, \dots, t_N\},$$

with  $t_0 < t_1 < \dots < t_N$  (cf. Lemma 3.3). Then for  $0 \leq i < N$ ,  $a, b \in \mathbf{R}$  such that  $(a, b) \subset (t_i, t_{i+1}) \cap \mathbf{R}$ , and for any  $c \in (a, b)$ , the inclusion

$$\mathcal{R}(\phi(\cdot, a)) \hookrightarrow \mathcal{R}(\phi(\cdot, c))$$

is a semi-algebraic homotopy equivalence,

*Proof.* The proof is an adaptation of a proof of a similar result in [10] (Lemma 3.8), though our situation is much simpler. It follows from Lemma 3.4 that the semi-algebraic set

$$\widehat{S^*(\bar{\varepsilon})} := S^*(\bar{\varepsilon}) \setminus \pi_{k+1}^{-1}(\{t_0, \dots, t_N\})$$

is a Whitney-stratified set. Moreover,  $\pi_{k+1}|_{\widehat{S^*(\bar{\varepsilon})}}$  is a proper stratified submersion. By Thom's first isotopy lemma (in the semi-algebraic version, over real closed fields [19]) the map  $\pi_{k+1}|_{\widehat{S^*(\bar{\varepsilon})}}$  is a locally trivial fibration.

Now let  $0 \leq i < N$ . It follows that for  $a', b' \in \mathbf{R} \langle \bar{\varepsilon} \rangle$  with  $t_i < a' \leq b' < t_{i+1}$ , that there exists a semi-algebraic homeomorphism

$$\theta_{a', b'} : S^*(\bar{\varepsilon})_{[a', b']} \rightarrow S^*(\bar{\varepsilon})_{a'} \times [a', b']$$

such that the following diagram commutes.

$$\begin{array}{ccc} S^*(\bar{\varepsilon})_{[a', b']} & \xrightarrow{\theta_{a', b'}} & S^*(\bar{\varepsilon})_{a'} \times [a', b'] \\ & \searrow \pi_{k+1} \quad \swarrow \pi_{k+1} & \\ & \mathbf{R} \langle \bar{\varepsilon} \rangle & \end{array}$$

Let

$$r_{a', b'} : S^*(\bar{\varepsilon})_{b'} \times [a', b'] \rightarrow S^*(\bar{\varepsilon})_{a'},$$

be the map defined by

$$\begin{aligned} r_{a', b'}(x, t) &= \pi_{[1, k]} \circ \theta_{a', b'}(x, t) \text{ if } t \leq P(x), \\ &= x, \text{ else.} \end{aligned}$$

Notice,  $r_{a',b'}$  is a semi-algebraic continuous map, and moreover for  $x \in S^*(\bar{\varepsilon})_{a'}$ ,  $r_{a',b'}(x, a') = x$ . Thus,  $r_{a',b'}$  is a semi-algebraic deformation retraction of  $S^*(\bar{\varepsilon})_{b'}$  to  $S^*(\bar{\varepsilon})_{a'}$ .

This implies that the inclusion

$$(3.2) \quad S^*(\bar{\varepsilon})_{a'} \hookrightarrow S^*(\bar{\varepsilon})_{b'}$$

is a semi-algebraic homotopy equivalence.

Now suppose that  $a, b \in \mathbb{R}$  with  $t_i < a \leq b < t_{i+1}$ .  $S^*(\bar{\varepsilon})_a$  and  $S^*(\bar{\varepsilon})_b$  are closed and bounded over  $\mathbb{R}$ , and that  $S^*(\bar{\varepsilon})_a \searrow \mathcal{R}(\phi(\cdot, a))$ ,  $S^*(\bar{\varepsilon})_b \searrow \mathcal{R}(\phi(\cdot, b))$ .

Then, it follows from Lemma 3.1 that the inclusions,

$$(3.3) \quad \text{ext}(\mathcal{R}(\phi(\cdot, a)), \mathbb{R}(\bar{\varepsilon})) \hookrightarrow S^*(\bar{\varepsilon})_a,$$

and

$$(3.4) \quad \text{ext}(\mathcal{R}(\phi(\cdot, b)), \mathbb{R}(\bar{\varepsilon})) \hookrightarrow S^*(\bar{\varepsilon})_b,$$

are semi-algebraic homotopy equivalences.

Thus, we have the following commutative diagram of inclusions

$$\begin{array}{ccc} S^*(\bar{\varepsilon})_a & \hookrightarrow & S^*(\bar{\varepsilon})_b \\ \uparrow & & \uparrow \\ \text{ext}(\mathcal{R}(\phi(\cdot, a)), \mathbb{R}(\bar{\varepsilon})) & \hookrightarrow & \text{ext}(\mathcal{R}(\phi(\cdot, b)), \mathbb{R}(\bar{\varepsilon})) \end{array}$$

in which all arrows other than the bottom inclusion are semi-algebraic homotopy equivalences, and hence so is the bottom arrow. This implies that the inclusion  $\mathcal{R}(\phi(\cdot, a)) \hookrightarrow \mathcal{R}(\phi(\cdot, b))$  is a semi-algebraic homotopy equivalence by an application of the Tarski-Seidenberg transfer principle (see for example [9, Chapter 2]).

Now assume that  $a = t_i$ . Using Lemma 3.1 we have that for all small enough  $\varepsilon > 0$ , the inclusion  $\mathcal{R}(\phi(\cdot, a)) \hookrightarrow \mathcal{R}(\phi(\cdot, a + \varepsilon))$  is a semi-algebraic homotopy equivalence. Moreover, from what has been already shown, the inclusion  $\mathcal{R}(\phi(\cdot, a + \varepsilon)) \hookrightarrow \mathcal{R}(\phi(\cdot, c))$  is a semi-algebraic homotopy equivalence. It now follows that  $\mathcal{R}(\phi(\cdot, a)) \hookrightarrow \mathcal{R}(\phi(\cdot, c))$  is a semi-algebraic homotopy equivalence. This completes the proof.  $\square$

**Lemma 3.5.** *Let  $\mathcal{G} \subset \mathbb{R}[\bar{\varepsilon}][T]$  be a finite set of non-zero polynomials and*

$$\{t_0, \dots, t_N\} \subset \bigcup_{G \in \mathcal{G}} \mathbb{Z}(G, \mathbb{R}(\bar{\varepsilon}))$$

*with  $t_0 < \dots < t_N$ . For  $G \in \mathcal{G}$ , let  $G = \sum_{\alpha} m_{G,\alpha} G_{\alpha}$ , with  $G_{\alpha} \in \mathbb{R}[T]$ ,  $m_{G,\alpha} \in \mathbb{R}[\bar{\varepsilon}]$ , and let  $M(G) = \{\alpha \mid m_{G,\alpha} \neq 0\}$ . Let  $\mathcal{H} = \bigcup_{G \in \mathcal{G}, \alpha \in M(G)} \{G_{\alpha}\}$ , and let*

$$\{s_0, \dots, s_M\} = \bigcup_{H \in \mathcal{H}} \mathbb{Z}(H, \mathbb{R})$$

*with  $s_0 < s_1 < \dots < s_M$ . Then, for each  $i, 0 \leq i < M$ , there exists  $j, 0 \leq j < N$ , such that  $(s_i, s_{i+1}) \subset \mathbb{R}$  is contained in  $(t_j, t_{j+1}) \cap \mathbb{R}$ .*

*Proof.* Notice that it follows from the definition of the set  $\{s_0, \dots, s_M\}$  that for any  $i, 0 \leq i < M$ , the sign condition (cf. Definition 3.1) realized by  $\mathcal{H}$  at  $t$  stays fixed for all  $t \in \mathbb{R}$ , such that  $t \in (s_i, s_{i+1})$ .

Since for any  $t \in \mathbb{R}$ , the sign condition realized by  $\mathcal{H}$  at  $t$  determines the sign condition of  $\mathcal{G}$  realized at  $t$ , it follows that the sign condition (cf. Definition 3.1) realized by  $\mathcal{G}$  at  $t$  also stays fixed for all  $t \in \mathbb{R}$ , such that  $t \in (s_i, s_{i+1})$ .

Suppose that  $t' \in \text{ext}((s_i, s_{i+1}), \mathbb{R}(\bar{\varepsilon}))$  such that  $G(t') = 0$  for some  $G \in \mathcal{G}$ . We claim that this implies that  $\lim_{\bar{\varepsilon}} t' \in \{s_i, s_{i+1}\}$ . Suppose not. Then,  $\lim_{\bar{\varepsilon}} t' \in (s_i, s_{i+1})$ , which contradicts the fact that the sign condition (cf. Definition 3.1) realized by  $\mathcal{G}$  at  $t$  stays fixed for all  $t \in \mathbb{R}$ , such that  $t \in (s_i, s_{i+1})$ , since  $G$  is a non-zero polynomial.

The lemma now follows from the hypothesis that  $\{t_0, \dots, t_N\} \subset \bigcup_{G \in \mathcal{G}} \mathbb{Z}(G, \mathbb{R}(\bar{\varepsilon}))$ .  $\square$

Let  $S = \mathcal{R}(\Phi)$  and  $P, t_0, \dots, t_N$  as in Proposition 3.2, and let  $\mathcal{G}, \mathcal{H}$ , and  $s_0 < \dots < s_M$  as in Lemma 3.5. Let  $s_{-1} = -\infty, s_{M+1} = \infty$ . Let  $\mathcal{F}$  denote the finite filtration of semi-algebraic sets, indexed by the finite ordered set  $T = \{s_i \mid -1 \leq i \leq M+1\}$ , with the element of  $\mathcal{F}$  indexed by  $s_i$  equal to  $S_{P \leq s_i}$ . We have the following proposition.

**Proposition 3.3.** *For each  $p \geq 0$ ,*

$$\mathcal{B}_p(S, P) = \mathcal{B}_p(\mathcal{F}).$$

*Proof.* It follows from Proposition 3.2 and Lemma 3.5 that for each  $i, -1 \leq i \leq M$  and  $s \in (s_i, s_{i+1})$ , the inclusion  $S_{P \leq s_i} \hookrightarrow S_{P \leq s}$  is a semi-algebraic homotopy equivalence.

The proposition will now follow from the following two claims.

**Claim 3.1.** *Suppose that  $s, t \in [s_{-1}, s_{M+1}], s \leq t$ . Then,  $\mu_p^{s,t}(\mathcal{F}(S, P)) \neq 0 \Rightarrow s, t \in \{s_{-1}, \dots, s_{M+1}\}$ .*

*Proof.* We consider the following two cases.

1.  $s \notin \{s_{-1}, \dots, s_{M+1}\}$ : Without loss of generality we can assume that  $s \in (s_i, s_{i+1})$  for some  $i, -1 \leq i \leq M$ . Now the inclusion  $S_{P \leq s'} \hookrightarrow S_{P \leq s}$ , is a semi-algebraic homotopy equivalence for all  $s' \in [s_i, s)$ , and hence  $i_p^{s', s}$  is an isomorphism for all  $s' \in [s_i, s)$ .

It follows that for all  $s' \in [s_i, s)$ ,

$$H_p^{s', t}(\mathcal{F}(S, P)) = \text{Im}(i_p^{s', t}) = \text{Im}(i_p^{s, t} \circ i_p^{s', s}) = \text{Im}(i_p^{s, t}) = H_p^{s, t}(\mathcal{F}(S, P)),$$

which implies that

$$(i_p^{s, t})^{-1}(H_p^{s', t}(\mathcal{F}(S, P))) = (i_p^{s, t})^{-1}(H_p^{s, t}(\mathcal{F}(S, P))) = H_p(S_{P \leq s}).$$

Noting that

$$\bigcup_{s' < s} (i_p^{s, t})^{-1}(H_p^{s', t}(\mathcal{F}(S, P))) = \bigcup_{s' \in [s_i, s)} (i_p^{s, t})^{-1}(H_p^{s', t}(\mathcal{F}(S, P))),$$

it now follows that

$$\begin{aligned}
M_p^{s,t}(\mathcal{F}(S, P)) &= \bigcup_{s' < s} (i_p^{s,t})^{-1}(\mathcal{H}_p^{s',t}(\mathcal{F}(S, P))) \\
&= \mathcal{H}_p(S_{P \leq s}), \\
N_p^{s,t}(\mathcal{F}(S, P)) &= \bigcup_{s' < s \leq t' < t} (i_p^{s,t'})^{-1}(\mathcal{H}_p^{s',t'}(\mathcal{F}(S, P))) \\
&= \bigcup_{s \leq t' < t} (i_p^{s,t'})^{-1}(\mathcal{H}_p^{s,t'}(\mathcal{F}(S, P))) \\
&= \mathcal{H}_p(S_{P \leq s}).
\end{aligned}$$

We have two sub-cases to consider.

(a) If  $t < s_{M+1}$ :

$$P_p^{s,t}(\mathcal{F}(S, P)) = M_p^{s,t}(\mathcal{F}(S, P)) / N_p^{s,t}(\mathcal{F}(S, P)) = 0.$$

(b) If  $t = s_{M+1} = \infty$ :

$$P_p^{s,\infty}(\mathcal{F}(S, P)) = \mathcal{H}_p(S_{P \leq s}) / \bigcup_{s \leq t} M_p^{s,t}(\mathcal{F}(S, P)) = 0.$$

since

$$\bigcup_{s \leq t} M_p^{s,t}(\mathcal{F}(S, P)) = \bigcup_{s \leq t} \mathcal{H}_p(S_{P \leq s}) = \mathcal{H}_p(S_{P \leq s}).$$

2.  $t \notin \{s_{-1}, \dots, s_{M+1}\}$ : Without loss of generality we can assume that  $t \in (s_i, s_{i+1})$  for some  $i, -1 \leq i \leq M$ . The inclusion  $S_{P \leq t'} \hookrightarrow S_{P \leq t}$ , is a semi-algebraic homotopy equivalence for all  $t' \in [s_i, t)$ , and hence  $i_p^{t',t}$  is an isomorphism for all  $t' \in [s_i, t)$ . This implies that for all  $t' \in [s_i, t)$ , and  $s' < t'$ ,  $\text{Im}(i_p^{s',t'})$  can be identified with  $\text{Im}(i_p^{s',t})$  using the isomorphism  $i_p^{t',t}$ . Furthermore, it is easy to verify that for every fixed  $s' < s$  and  $s \leq t' \leq t''$ ,

$$(i_p^{s,t'})^{-1}(\mathcal{H}_p^{s',t'}(\mathcal{F}(S, P))) \subset (i_p^{s,t'})^{-1}(\mathcal{H}_p^{s',t''}(\mathcal{F}(S, P))),$$

and hence for each fixed  $s' < s$ ,

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

It follows that for  $t \in (s_i, s_{i+1})$

$$\begin{aligned}
N_p^{s,t}(\mathcal{F}(S, P)) &= \bigcup_{s' < s \leq t' < t} (i_p^{s,t'})^{-1}(\mathcal{H}_p^{s',t'}(\mathcal{F}(S, P))) \\
&= \bigcup_{s' < s} (i_p^{s,t})^{-1}(\mathcal{H}_p^{s',t}(\mathcal{F}(S, P))) \\
&= M_p^{s,t}(\mathcal{F}(S, P)).
\end{aligned}$$

We have

$$P_p^{s,t}(\mathcal{F}(S, P)) = M_p^{s,t}(\mathcal{F}(S, P)) / N_p^{s,t}(\mathcal{F}(S, P)) = 0.$$

This completes the proof.  $\square$

**Claim 3.2.** For each  $i, j, -1 \leq i \leq j \leq M+1$ ,  $\mu_p^{s_i, s_j}(\mathcal{F}(S, P)) = \mu_p^{s_i, s_j}(\mathcal{F})$ .

*Proof.* It suffices to prove that

$$\begin{aligned} M_p^{s_i, s_j}(\mathcal{F}(S, P)) &= M_p^{s_i, s_j}(\mathcal{F}), \\ N_p^{s_i, s_j}(\mathcal{F}(S, P)) &= N_p^{s_i, s_j}(\mathcal{F}). \end{aligned}$$

To prove the first equality we use the fact that  $s' \in [s_{i-1}, s_i)$ , the inclusion  $S_{P \leq s_{i-1}} \hookrightarrow S_{P \leq s'}$  is a semi-algebraic homotopy equivalence.

Hence,

$$\begin{aligned} M_p^{s_i, s_j}(\mathcal{F}(S, P)) &= \bigcup_{s' < s_i} (i_p^{s_i, s_j})^{-1}(\mathcal{H}_p^{s', s_j}(\mathcal{F}(S, P))) \\ &= (i_p^{s_i, s_j})^{-1}(\mathcal{H}_p^{s_{i-1}, s_j}(\mathcal{F}(S, P))) \\ &= M_p^{s_i, s_j}(\mathcal{F}). \end{aligned}$$

Using additionally the fact that  $t' \in [s_{j-1}, s_j)$ , the inclusion  $S_{P \leq s_{j-1}} \hookrightarrow S_{P \leq t'}$  is a semi-algebraic homotopy equivalence, we have:

$$\begin{aligned} N_p^{s_i, s_j}(\mathcal{F}(S, P)) &= \bigcup_{s' < s_i \leq t' < s_j} (i_p^{s_i, t'})^{-1}(\mathcal{H}_p^{s', t'}(\mathcal{F}(S, P))) \\ &= (i_p^{s_i, s_{j-1}})^{-1}(\mathcal{H}_p^{s_{i-1}, s_{j-1}}(\mathcal{F}(S, P))) \\ &= N_p^{s_i, s_j}(\mathcal{F}). \end{aligned}$$

□

This concludes the proof of Proposition 3.3. □

*Proof of Proposition 3.1.* Follows immediately from Proposition 3.3. □

**3.2. Persistent multiplicities for finite filtration.** In this section, we prove a formula for the persistent multiplicities associated to a finite filtration  $\mathcal{F}$ , which we later use in Algorithm 4 to obtain the barcodes of a finite filtration. We deduce the formula from our definition of persistent multiplicity (cf. Eqn. (2.1) in Definition 2.5).<sup>2</sup>

**Proposition 3.4.** *Let  $\mathcal{F}$  denote a finite filtration, given by  $X_0 \subset \cdots \subset X_M = X_{M+1} = \cdots = X$ , such that rank of  $\mathcal{H}_p(X_j)$  is finite for each  $p \geq 0$ . Then for  $0 < j < k$ ,*

$$(3.5) \quad \mu_p^{j,k}(\mathcal{F}) = \begin{cases} (b_p^{j,k-1}(\mathcal{F}) - b_p^{j,k}(\mathcal{F})) - (b_p^{j-1,k-1}(\mathcal{F}) - b_p^{j-1,k}(\mathcal{F})), & k < \infty, \\ b_p^{j,k}(\mathcal{F}) - b_p^{j-1,k}(\mathcal{F}), & k = \infty. \end{cases}$$

*Proof.* We first prove the case where  $k$  is finite. By Definition 2.4,

$$\begin{aligned} \mu_p^{j,k}(\mathcal{F}) &= \dim P_p^{j,k}(\mathcal{F}) \\ &= \dim M_p^{j,k}(\mathcal{F}) - \dim N_p^{j,k}(\mathcal{F}). \end{aligned}$$

<sup>2</sup>This formula already appears in [24, page 152], but what is meant by “independent  $p$ -dimensional classes that are born at  $K_i$ , and die entering  $K_j$ ” *loc. cit.* is not totally transparent. See also Remark 1.



Since  $\mathcal{F}$  is finite, we have

$$\begin{aligned} M_p^{j,k}(\mathcal{F}) &= (i_p^{j,k})^{-1}(\mathbb{H}_p^{j-1,k}(\mathcal{F})), \\ N_p^{j,k}(\mathcal{F}) &= (i_p^{j,k-1})^{-1}(\mathbb{H}_p^{j-1,k-1}(\mathcal{F})). \end{aligned}$$

Note that  $(i_p^{j,k})^{-1}(\mathbb{H}_p^{j-1,k}(\mathcal{F}))$  is a subspace of  $\mathbb{H}_p(X_j)$ , and hence the linear map  $i_p^{j,k} : \mathbb{H}_p(X_j) \rightarrow \mathbb{H}_p(X_k)$  factors through a surjection  $f : \mathbb{H}_p(X_j) \rightarrow \mathbb{H}_p^{j,k}(\mathcal{F})$  followed by an injection  $\mathbb{H}_p^{j,k}(\mathcal{F}) \hookrightarrow \mathbb{H}_p(X_k)$  as shown in the following diagram.

$$\begin{array}{ccc} \mathbb{H}_p(X_j) & \xrightarrow{i_p^{j,k}} & \mathbb{H}_p(X_k) \\ & \searrow f & \nearrow \\ & \mathbb{H}_p^{j,k}(\mathcal{F}) & \end{array} .$$

Now  $\mathbb{H}_p^{j-1,k}(\mathcal{F})$  is a subspace of  $\mathbb{H}_p^{j,k}(\mathcal{F})$ , and let

$$m : \mathbb{H}_p^{j,k}(\mathcal{F}) \rightarrow \mathbb{H}_p^{j,k}(\mathcal{F}) / \mathbb{H}_p^{j-1,k}(\mathcal{F})$$

be the canonical surjection. Let  $g = m \circ f$ . Since  $f$  and  $m$  are both surjective, so is  $g$ .

$$\begin{array}{ccccc} \mathbb{H}_p(X_j) & \xrightarrow{f} & \mathbb{H}_p^{j,k}(\mathcal{F}) & \xrightarrow{m} & \mathbb{H}_p^{j,k}(\mathcal{F}) / \mathbb{H}_p^{j-1,k}(\mathcal{F}) \\ & & \searrow & \nearrow & \\ & & g = m \circ f & & \end{array}$$

Now notice that

$$\begin{aligned} M_p^{j,k}(\mathcal{F}) &= (i_p^{j,k})^{-1}(\mathbb{H}_p^{j-1,k}(\mathcal{F})) \\ &= f^{-1}(\mathbb{H}_p^{j-1,k}(\mathcal{F})) \\ &= \ker(g). \end{aligned}$$

Since  $g$  is surjective,

$$\text{rank}(g) = \dim \mathbb{H}_p^{j-1,k}(\mathcal{F}) - \dim \mathbb{H}_p^{j,k}(\mathcal{F}),$$

and using the rank-nullity theorem we obtain

$$(3.6) \quad \dim M_p^{j,k}(\mathcal{F}) = b_p(X_j) - (b_p^{j,k}(\mathcal{F}) - b_p^{j-1,k}(\mathcal{F})).$$

Using a similar argument we obtain

$$(3.7) \quad \dim N_p^{j,k}(\mathcal{F}) = b_p(X_j) - (b_p^{j,k-1}(\mathcal{F}) - b_p^{j-1,k-1}(\mathcal{F})).$$

Finally,

$$\begin{aligned} \mu_p^{j,k}(\mathcal{F}) &= \dim M_p^{j,k}(\mathcal{F}) - \dim N_p^{j,k}(\mathcal{F}) \\ &= b_p^{j-1,k}(\mathcal{F}) - b_p^{j,k}(\mathcal{F}) + (b_p^{j,k-1}(\mathcal{F}) - b_p^{j-1,k-1}(\mathcal{F})) \\ &= (b_p^{j,k-1}(\mathcal{F}) - b_p^{j,k}(\mathcal{F})) - (b_p^{j-1,k-1}(\mathcal{F}) - b_p^{j-1,k}(\mathcal{F})). \end{aligned}$$

If  $k = \infty$ , then by Definition 2.4,

$$\begin{aligned} \mu_p^{j,k}(\mathcal{F}) &= \dim P_p^{j,k}(\mathcal{F}) \\ &= \dim \mathbb{H}_p(X_j) - \dim \bigcup_{j \leq t} M_p^{j,t}(\mathcal{F}). \end{aligned}$$

Since  $M_p^{s,t}(\mathcal{F}) \subset M_p^{s,t'}(\mathcal{F})$  for  $t \leq t'$ , we have

$$M_p^{j,t}(\mathcal{F}) \subset M_p^{j,t+1}(\mathcal{F}) \subset \dots \subset M_p^{j,M}(\mathcal{F}) = M_p^{j,M+1}(\mathcal{F}) = \dots = M_p^{j,\infty}(\mathcal{F})$$

$$\bigcup_t M_p^{j,t}(\mathcal{F}) = M_p^{j,M}(\mathcal{F})$$

Therefore,

$$\begin{aligned} \mu_p^{j,k}(\mathcal{F}) &= \dim H_p(K_j) - \dim M_p^{j,M}(\mathcal{F}) \\ &= b_p(X_j) - (b_p(X_j) - (b_p^{j,M}(\mathcal{F}) - b_p^{j-1,M}(\mathcal{F}))) \\ &= b_p^{j,M}(\mathcal{F}) - b_p^{j-1,M}(\mathcal{F}) \end{aligned}$$

□

#### 4. ALGORITHMS AND PROOF OF THEOREM 1

In this section we describe our algorithmic results leading to the proof of Theorem 1. We begin by stating some preliminary mathematical results in Section 4.1 that we will need for our algorithms. We describe two technical algorithms that we will need in Section 4.2. In Section 4.4 we describe Algorithm 3 for reducing the given continuous filtration to a finite one. The proof of correctness of this algorithm relies on Proposition 3.3 proved earlier. Finally, in Section 4.4 we describe our algorithm for computing the barcode of a semi-algebraic filtration (algorithm 5), prove its correctness and analyze its complexity, thereby proving Theorem 1.

##### 4.1. Preliminaries.

**Notation 12** (Derivatives). Let  $P$  be a univariate polynomial of degree  $p$  in  $\mathbb{R}[X]$ . We will denote by  $\text{Der}(P)$  the tuple  $(P, P', \dots, P^{(p)})$  of derivatives of  $P$ .

The significance of  $\text{Der}(P)$  is encapsulated in the following lemma which underlies our representations of elements of  $\mathbb{R}$  which are algebraic over  $\mathbb{D}$  (cf. Definition 4.1).

**Proposition 4.1** (Thom's Lemma). *Let  $f \in \mathbb{R}[X]$  be a univariate polynomial, and, let  $\sigma$  be a sign condition on  $\text{Der}(f)$ . Then  $\mathcal{R}(\sigma)$  is either empty, a point, or an open interval.*

*Proof.* See [9, Proposition 2.27]. □

Proposition 4.1 allows us to specify elements of  $\mathbb{R}$  which are algebraic over  $\mathbb{D}$  by means of a pair  $(f, \sigma)$  where  $f \in \mathbb{D}[X]$  and  $\sigma \in \{0, 1, -1\}^{\text{Der}(f)}$ .

**Definition 4.1.** We say that  $x \in \mathbb{R}$  is *associated to the pair*  $(f, \sigma)$ , if  $\sigma(f) = 0$  and if  $\text{Der}(f)$  realizes the sign condition  $\sigma$  at  $x$ . We call the pair  $(f, \sigma)$  to be a *Thom encoding* specifying  $x$ .

We will also use the notion of a weak sign condition (cf. Definition 3.1).

**Definition 4.2.** A *weak sign condition* is an element of

$$\{\{0\}, \{0, 1\}, \{0, -1\}\}.$$

We say

$$\begin{cases} \mathbf{sign}(x) \in \{0\} & \text{if and only if } x = 0, \\ \mathbf{sign}(x) \in \{0, 1\} & \text{if and only if } x \geq 0, \\ \mathbf{sign}(x) \in \{0, -1\} & \text{if and only if } x \leq 0. \end{cases}$$

A *weak sign condition* on  $\mathcal{Q}$  is an element of  $\{\{0\}, \{0, 1\}, \{0, -1\}\}^{\mathcal{Q}}$ . If  $\sigma \in \{\{0, 1, -1\}\}^{\mathcal{Q}}$ , its *relaxation*  $\bar{\sigma}$  is the weak sign condition on  $\mathcal{Q}$  defined by  $\bar{\sigma}(Q) = \overline{\sigma(Q)}$ . The *realization of the weak sign condition*  $\tau$  is

$$\mathcal{R}(\tau) = \{x \in \mathbb{R}^k \mid \bigwedge_{Q \in \mathcal{Q}} \mathbf{sign}(Q(x)) \in \tau(Q)\}.$$

**Definition 4.3.** We say that a set of polynomials  $\mathcal{F} \subset \mathbb{R}[X]$  is *closed under differentiation* if  $0 \notin \mathcal{F}$  and if for each  $f \in \mathcal{F}$  then  $f' \in \mathcal{F}$  or  $f' = 0$ .

**Lemma 4.1.** ([9, Lemma 5.33]) *Let  $\mathcal{F} \subset \mathbb{R}[X]$  be a finite set of polynomials closed under differentiation and let  $\sigma$  be a sign condition on the set  $\mathcal{F}$ . Then*

- (a)  $\mathcal{R}(\sigma)$  is either empty, a point, or an open interval.
- (b) If  $\mathcal{R}(\sigma)$  is empty, then  $\mathcal{R}(\bar{\sigma})$  is either empty or a point.
- (c) If  $\mathcal{R}(\sigma)$  is a point, then  $\mathcal{R}(\bar{\sigma})$  is the same point.
- (d) If  $\mathcal{R}(\sigma)$  is an open interval then  $\mathcal{R}(\bar{\sigma})$  is the corresponding closed interval.

*Remark 9.* In what follows we will allow ourselves to use for  $P \in \mathbb{R}[X_1, \dots, X_k]$ ,  $\mathbf{sign}(P) = 0$  (resp.  $\mathbf{sign}(P) = 1$ ,  $\mathbf{sign}(P) = -1$ ) in place of the atoms  $P = 0$  (resp.  $P > 0$ ,  $P < 0$ ) in formulas. Similarly, we might write  $\mathbf{sign}(P) \in \bar{\sigma}$ , where  $\bar{\sigma}$  is a weak sign condition in place of the corresponding weak inequality  $P \geq 0$  or  $P \leq 0$ . It should be clear that this abuse of notation is harmless.

In addition to the mathematical preliminaries described above, we also need two technical algorithmic results that we describe in the next section

**4.2. Some preliminary algorithms.** For technical reasons that will become clear when we describe Algorithm 3, we will need to convert efficiently a given quantifier-free formula defining a closed semi-algebraic set, into a closed formula defining the same semi-algebraic set. This is a non-trivial problem, since the standard quantifier-elimination algorithms in algorithmic semi-algebraic geometry does not guarantee that the output will be a closed formula even if it is known in advance that the semi-algebraic set that the formula is describing is closed. Luckily we only need to deal with formulas in one variable, where the problem is somewhat simpler. Note that even in this case, it is not possible to obtain the description of the given closed semi-algebraic set as a closed formula by merely weakening the inequalities in the original formula.

For example, consider the formula  $\phi(X) := (X^2(X - 1) > 0) \wedge ((X \geq 2) \vee (X \leq 0))$ . Then,  $\mathcal{R}(\phi) = [2, \infty)$  is a closed semi-algebraic set, but the formula obtained by weakening the inequality  $X^2(X - 1) > 0$ , namely

$$\tilde{\phi} := (X^2(X - 1) \geq 0) \wedge ((X \geq 2) \vee (X \leq 0)),$$

has as its realization the set  $\{0\} \cup [2, \infty)$  which is strictly bigger than  $\mathcal{R}(\phi)$ .

Nevertheless, using Lemma 4.1 we have the following algorithm to achieve the above mentioned task efficiently.

---

**Algorithm 1** (Make closed)

---

**Input:**

A quantifier-free formula  $\theta(Y)$  with coefficients in  $D$ , in one free variable  $Y$ , such that  $\mathcal{R}(\theta)$  is closed.

**Output:**

A closed formula  $\psi(Y)$  equivalent to  $\theta(Y)$ .

**Procedure:**

- 1: Let  $\theta(Y) = \bigvee_{1 \leq i \leq M} \bigwedge_{1 \leq j \leq N_i} (\text{sign}(F_{i,j}) = \sigma_{i,j})$ .
- 2: **for** each  $(i, j)$  such that  $\sigma_{i,j} \neq 0$  **do**
- 3:   Call Algorithm 13.1 (Computing realizable sign conditions) in [9] with input  $\text{Der}(F_{i,j})$ , and obtain the set  $\Sigma_{i,j}$  of realizable sign conditions of  $\text{Der}(F_{i,j})$ .
- 4:    $\Sigma'_{i,j} \leftarrow \{\sigma \in \Sigma_F \mid \sigma(F_{i,j}) = \sigma_{i,j}\}$ .
- 5:    $\bar{\Sigma}_{i,j} \leftarrow \{\bar{\sigma} \mid \sigma \in \Sigma'_{i,j}\}$ .
- 6: **end for**
- 7: **return** the formula

$$\psi(Y) = \bigvee_{1 \leq i \leq M} \left( \bigwedge_{\sigma_{i,j}=0} (\text{sign}(F_{i,j}) = 0) \wedge \bigwedge_{\sigma_{i,j} \neq 0} \bigvee_{\bar{\sigma} \in \bar{\Sigma}_{i,j}} (\text{sign}(F_{i,j}) \in \bar{\sigma}) \right).$$

**Complexity:** The complexity of the algorithm is bounded by  $(sd)^{O(1)}$  where  $s$  is the number of polynomials appearing  $\theta$  and  $d$  a bound on their degrees.

---

*Proof of correctness.* The correctness of the algorithm follows from the correctness of Algorithm 13.1 (Computing realizable sign conditions) in [9], and Lemma 4.1.  $\square$

*Complexity analysis.* The complexity bound follows from the complexity of Algorithm 13.1 (Computing realizable sign conditions) in [9].  $\square$

We will also need an algorithm that takes as input a finite set of polynomials  $\mathcal{G}$  in one variable with coefficients in  $D[\bar{\varepsilon}]$ , and outputs a set of Thom encodings whose set of associated points  $\{s_0, \dots, s_M\}$  satisfy the property stated in Lemma 3.5.

---

**Algorithm 2** (Removal of infinitesimals)

---

**Input:**

A finite set  $\mathcal{G} \subset D[\bar{\varepsilon}][T]$  such that each  $P \in \mathcal{G}$  depends on at most  $k+1$  of the  $\varepsilon_i$ 's.

**Output:**

A finite set of Thom encodings  $\mathcal{F} = \{(f_i, \sigma_i) \mid 0 \leq i \leq N\}$ , with  $f_i \in D[T]$  with associated points  $s_0 < \dots < s_M$ , such that letting  $s_{-1} = -\infty, s_{M+1} = \infty$ , for each  $i, 0 \leq i < M$ , there exists  $j, 0 \leq j < N$ , such that  $(s_i, s_{i+1}) \subset \mathbb{R}$  is contained in  $(t_j, t_{j+1}) \cap \mathbb{R}$ , where  $\{t_0, \dots, t_N\} = \bigcup_{G \in \mathcal{G}} Z(G, \mathbb{R}(\bar{\varepsilon}))$ , with  $t_0 < \dots < t_N$ .

---

**Procedure:**

- 1: **for**  $G \in \mathcal{G}$  **do**
- 2:      $0 \leq i_0 < \dots < i_h \leq s+1$  be such that  $G \in D[\varepsilon_{i_0}, \dots, \varepsilon_{i_h}][T]$ .
- 3:     Write  $G = \sum_{\alpha} m_{G,\alpha}(\varepsilon_{i_0}, \dots, \varepsilon_{i_h})G_{\alpha}$ , with  $G_{\alpha} \in D[T], m_{G,\alpha} \in D[\varepsilon_{i_0}, \dots, \varepsilon_{i_h}]$ .
- 4:     Let  $M(G) = \{\alpha \mid m_{G,\alpha} \neq 0\}$ .
- 5: **end for**
- 6: Let  $\mathcal{H} = \bigcup_{G \in \mathcal{G}, \alpha \in M(G)} \{G_{\alpha}\}$ .
- 7: Use Algorithm 10.17 from [9] with  $\mathcal{H}$  as input to obtain an ordered list of Thom encodings  $\mathcal{F}$ .
- 8: **return**  $\mathcal{F}$ .

**Complexity:** The complexity of the algorithm is bounded by  $sD^{O(k)}$ , where  $s = \text{card}(\mathcal{G})$  and  $D$  is a bound on the degrees of the polynomials in  $\mathcal{G}$  in  $\varepsilon$  and in  $T$ .

*Proof of correctness.* The correctness of the algorithm follows from Lemma 3.5 and the correctness of Algorithm 10.17 from [9].  $\square$

*Complexity analysis.* The complexity bound follows from the complexity bound of Algorithm 10.17 from [9].  $\square$

**4.3. Algorithm for computing simplicial replacement.** We recall the following definition from [7].

**Notation 13** (Diagram of various unions of a finite number of subspaces). Let  $J$  be a finite set,  $A$  a topological space, and  $\mathcal{A} = (A_j)_{j \in J}$  a tuple of subspaces of  $A$  indexed by  $J$ .

For any subset  $J' \subset J$ , we denote

$$\begin{aligned} \mathcal{A}^{J'} &= \bigcup_{j' \in J'} A_{j'}, \\ \mathcal{A}_{J'} &= \bigcap_{j' \in J'} A_{j'}, \end{aligned}$$

We consider  $2^J$  as a category whose objects are elements of  $2^J$ , and whose only morphisms are given by:

$$\begin{aligned} 2^J(J', J'') &= \emptyset \text{ if } J' \not\subset J'', \\ 2^J(J', J'') &= \{\iota_{J', J''}\} \text{ if } J' \subset J''. \end{aligned}$$

We denote by  $\mathbf{Simp}^J(\mathcal{A}) : 2^J \rightarrow \mathbf{Top}$  the functor (or the diagram) defined by

$$\mathbf{Simp}^J(\mathcal{A})(J') = \mathcal{A}^{J'}, J' \in 2^J,$$

and  $\mathbf{Simp}^J(\mathcal{A})(\iota_{J', J''})$  is the inclusion map  $\mathcal{A}^{J'} \hookrightarrow \mathcal{A}^{J''}$ .

We will use an algorithm whose existence is proved in [7, Theorem 1], and which we will refer to as *Algorithm for computing simplicial replacement*, that given a tuple of closed-formulas  $\Phi = (\phi_0, \dots, \phi_N)$ ,  $R > 0$ , and  $\ell \geq 0$ , produces as output a simplicial complex  $K$  and subcomplexes  $K_i$ ,  $0 \leq i \leq N$  of  $K$ , such that the diagram

$$\mathbf{Simp}^{[N]} \left( (\mathcal{R}(\phi_i, \overline{B_k(0, R)}))_{i \in [N]} \right)$$

is homologically  $\ell$ -equivalent ([7, Section 2.1.1]) to the diagram

$$\mathbf{Simp}^{[N]}((|K_i|)_{i \in [N]})$$

(where  $|K_i| \subset |K|$  is the geometric realization of  $K_i$  and  $[N] = \{0, \dots, N\}$ ).

We refer the reader to [7] for the details.

The complexity of this algorithm, as well as the size of the output simplicial complex  $\Delta$ , are bounded by

$$(Nsd)^{k^{O(m)}},$$

where  $s = \text{card}(\mathcal{P})$ , and  $d = \max_{P \in \mathcal{P}} \deg(P)$ .

**4.4. Algorithm for reducing to a finite filtration.** We are now in a position to describe our algorithm for reducing the problem of computing the barcode of a filtration of a semi-algebraic set  $S$  by the sub-level sets of a polynomial  $P$ , to the problem of computing the barcode of a finite filtration.

Algorithm 3 computes a finite subset of  $\mathbb{R}$ , as Thom encodings (cf. Definition 4.1), such that it includes the values of  $P$  at which the homotopy type of the sub-level sets of  $S$  changes. The algorithm has singly exponentially bounded complexity.

---

**Algorithm 3** (Reducing to a finite filtration)

---

**Input:**

- (a)  $\ell \in \mathbb{Z}_{\geq 0}$ .
- (b)  $R \in \mathbb{D}, R > 0$ .
- (c) A finite set  $\mathcal{P} = \{P_1, \dots, P_s\} \subset \mathbb{D}[X_1, \dots, X_k]$ .
- (d) A  $\mathcal{P}$ -closed formula  $\phi$ .
- (e) A polynomial  $P \in \mathbb{R}[X_1, \dots, X_k]$ .

**Output:**

- (a) A finite set of Thom encodings  $\mathcal{F} = \{(f_i, \sigma_i) \mid 0 \leq i \leq N\}$ , with  $f_i \in \mathbb{D}[T]$  with associated points  $t_0 < \dots < t_N$ , such that for  $t \in \mathbb{R}$ , denoting by  $S_t = \mathcal{R}(\phi) \cap \overline{B_k(0, R)} \cap \{x \mid P(x) \leq t\}$ , for each  $i, 0 \leq i \leq N-1$ , and all  $t \in [t_i, t_{i+1})$  the inclusion maps  $S_{t_i} \hookrightarrow S_t$  are homological equivalences.
- (b) A filtration of finite simplicial complexes

$$K_0 \subset K_1 \subset \dots \subset K_N$$

such that  $\mathbf{Simp}^{[N]}(S_{t_0}, \dots, S_{t_N})$  is homologically  $\ell$ -equivalent to  $\mathbf{Simp}^{[N]}(|K_0|, \dots, |K_N|)$ .

**Procedure:**

- 1:  $P_0 \leftarrow \sum_{i=1}^k X_i^2 - R$ .
- 2:  $P_{s+1} \leftarrow P - Y$ .
- 3:

$$\mathcal{P}^*(\bar{\varepsilon}) \leftarrow \bigcup_{0 \leq i \leq s+1} \{P_i + \varepsilon_i, P_i - \varepsilon_i\}.$$

- 4: Denote by  $\phi^*(\bar{\varepsilon})$ , the  $\mathcal{P}^*(\bar{\varepsilon})$ -closed formula obtained by replacing each occurrence of  $P_i \geq 0$  in  $\phi$  by  $P_i + \varepsilon_i \geq 0$  (resp.  $P_i \leq 0$  in  $\phi$  by  $P_i - \varepsilon_i \leq 0$ ) for  $0 \leq i \leq s+1$ .
-

---



---

```

5: for  $\mathcal{Q} \subset \mathcal{P}^*(\varepsilon)$ ,  $\text{card}(\mathcal{Q}) \leq k$  do
6:
    
$$Jac(\mathcal{Q}) \leftarrow \sum_{1 \leq i_1 < i_2 < \dots < i_{\text{card}(\mathcal{Q}')} \leq k} \det \left( \left( \frac{\partial Q}{\partial X_{i_j}} \right)_{Q \in \mathcal{Q}, 1 \leq i \leq k} \right)$$

7: end for
8: for  $\mathcal{Q} \subset \mathcal{P}^*(\varepsilon)$ ,  $\text{card}(\mathcal{Q}) = k + 1$  do
9:
    
$$\Sigma(\mathcal{Q}') \leftarrow \sum_{Q \in \mathcal{Q}} Q^2.$$

10: end for
11:
 $\mathcal{H} \leftarrow \{Jac(\mathcal{Q}) \mid \mathcal{Q} \subset \mathcal{P}^*(\varepsilon), \text{card}(\mathcal{Q}) \leq k\} \cup \{\Sigma(\mathcal{Q}) \mid \mathcal{Q} \subset \mathcal{P}^*(\varepsilon), \text{card}(\mathcal{Q}) = k + 1\}.$ 
12: Call Algorithm 14.1 (Block Elimination) from [9] with the block of variables
     $(X_1, \dots, X_k)$  and  $\mathcal{H}$  as input, and obtain  $\mathcal{G} = \text{BElim}_X(\mathcal{F})$  (following the
    same notation as in [9, Algorithm 14.1 (Block Elimination)]).
13: Call Algorithm 2 with  $\mathcal{G}$  as input and obtain an ordered list of Thom encodings
     $\mathcal{F} = ((f_0, \sigma_0), \dots, (f_N, \sigma_N))$ .
14: for  $0 \leq i \leq N$  do
15:     Call Algorithm 14.5 (Quantifier Elimination) [9] with input the formula
        
$$\tilde{\psi}(Y) := \forall Z((f_i(Z) = 0) \wedge (\text{sign}(\text{Der}(f_i))(Z) = \sigma_i)) \Rightarrow (Y \leq Z)$$

        to obtain an equivalent quantifier-free formula  $\tilde{\psi}_i(Y)$ .
16:     Call Algorithm 1 with  $\tilde{\psi}_i(Y)$  as input to obtain a closed formula  $\psi_i(Y)$ .
17:      $\phi_i \leftarrow \tilde{\phi} \wedge \psi_i(Y)$ .
18:      $\mathcal{Q}_i \leftarrow$  the set of polynomials appearing in  $\psi_i$ .
19: end for
20: Call Algorithm for simplicial replacement with input: the closed formulas
     $\phi_0, \dots, \phi_N$ ,  $R$  and  $\ell$ , and output the simplicial complexes  $K_i$ ,  $0 \leq i \leq N$ .
Complexity: The complexity of the algorithm is bounded by  $(sd)^{k^{O(\ell)}}$ , where
     $s = \text{card}(\mathcal{P})$ , and  $d = \max_{P \in \mathcal{P}} \deg(P)$ .

```

---

*Proof of correctness.* The correctness of the algorithm follows from Proposition 3.3, and the correctness of the following algorithms: Algorithm 14.1 (Block Elimination) in [9], Algorithm 2, Algorithm 14.5 (Quantifier Elimination) in [9], Algorithm 1, and the Algorithm for simplicial replacement [7, Theorem 1].  $\square$

*Complexity analysis.* The complexity bound follows from the complexity bounds of Algorithm 14.1 (Block Elimination) in [9], Algorithm 2, Algorithm 14.5 (Quantifier Elimination) in [9], Algorithm 1, and the Algorithm for simplicial replacement [7, Theorem 1].  $\square$

**4.5. Computing barcodes of semi-algebraic filtrations.** We can now describe our algorithm for computing the barcode of the filtration of a semi-algebraic set by the sub-level sets of a polynomial. First we need an algorithm for computing barcodes of finite filtrations of finite simplicial complexes.



**Algorithm 4** (Barcode of a finite filtration of finite simplicial complexes)**Input:**

1.  $\ell \in \mathbb{Z}_{\geq 0}$ .
2. A finite filtration  $\mathcal{F}$ ,  $K_0 \subset \cdots \subset K_N$  of finite simplicial complexes.

**Output:**

$$\mathcal{B}_p(\mathcal{F}), 0 \leq p \leq \ell.$$

**Procedure:**

- 1:  $K_{-1} \leftarrow \emptyset$ .
- 2:  $K_{N+1} \leftarrow K_N$ .
- 3: **for**  $-1 \leq i \leq j \leq N+1$  **do**
- 4:   Use Gaussian elimination to compute the persistent Betti numbers  $b_p^{i,j}(\mathcal{F})$ .
- 5: **end for**
- 6: **for**  $0 \leq p \leq \ell, 0 \leq i \leq j \leq N+1$  **do**
- 7:
- 8:   **if**  $j = N+1$  **then**
- 9:     
$$\mu_p^{i,j} \leftarrow b_p^{i,j}(\mathcal{F}) - b_p^{i-1,j}(\mathcal{F})$$
- 10:   **else**
- 11:     
$$\mu_p^{i,j} \leftarrow (b_p^{i,j-1}(\mathcal{F}) - b_p^{i,j}(\mathcal{F})) - (b_p^{i-1,j-1}(\mathcal{F}) - b_p^{i-1,j}(\mathcal{F}))$$
- 12:   **end if**
- 13:   (cf. Eqn. (3.5)).
- 14:   **end for**
- 15:   **for**  $0 \leq p \leq \ell$  **do**
- 16:     Output
- 17:     
$$\mathcal{B}_p(\mathcal{F}) = \{(i, j, \mu_p^{i,j}) \mid 0 \leq i \leq j \leq N, \mu_p^{i,j} > 0\} \cup$$
- 18:     
$$\{(i, \infty, \mu_p^{i,j}) \mid 0 \leq i \leq j = N+1, \mu_p^{i,j} > 0\}.$$
- 19:   **end for**

**Complexity:** The complexity of the algorithm is bounded polynomially in  $N$  times the number of simplices appearing in the complex  $K_N$ .

*Proof of correctness.* The correctness of the algorithm follows from Eqn. (3.5).  $\square$

*Complexity analysis.* The complexity of the algorithm follows from the complexity of Gaussian elimination.  $\square$

**Algorithm 5** (Computing persistent homology barcodes of semi-algebraic sets)**Input:**

- (A) A  $\mathcal{P}$ -closed formula  $\phi$ , with  $\mathcal{P}$  a finite subset of  $D[X_1, \dots, X_k]$ , such that  $\mathcal{R}(\phi, \mathbb{R}^k)$  is bounded.
- (B) A polynomial  $P \in D[X_1, \dots, X_k]$ .
- (C)  $\ell, 0 \leq \ell \leq k$ .

**Output:**

For each  $p, 0 \leq p \leq \ell$ ,  $\mathcal{B}_p(S, P)$ , where  $S = \mathcal{R}(\phi)$ .

**Procedure:**

- 1:  $\mathcal{P}' \leftarrow \mathcal{P} \cup \{\varepsilon(X_1^2 + \dots + X_k^2) - 1\}$ .
- 2:  $\phi' \leftarrow \phi \wedge \varepsilon^2(X_1^2 + \dots + X_k^2) - 1 \leq 0$ .
- 3:  $R \leftarrow R(\varepsilon)$ ,  $D \leftarrow D[\varepsilon]$ .
- 4: Call Algorithm 3 with input  $\ell, 1/\varepsilon, \mathcal{P}', \phi', P$ , to obtain a finite ordered set of Thom encodings  $(f_0, \sigma_0), \dots, (f_N, \sigma_N)$ , and a finite filtration  $\mathcal{F} = (K_0 \subset \dots \subset K_N)$ , where  $K_N$  is a finite simplicial complex.
- 5: Call Algorithm 4 with input  $\ell$  and the finite filtration  $\mathcal{F}$ , and output for each  $p, 0 \leq p \leq \ell$ ,  $\mathcal{B}_p(\mathcal{F})$ .
- 6: **for each**  $p, 0 \leq p \leq \ell$   
 Output

$$\mathcal{B}_p(S, P) = \bigcup_{(i,j,\mu) \in \mathcal{B}_p(\mathcal{F}), 0 \leq i \leq j \leq N} \{((f_i, \sigma_i), (f_j, \sigma_j), \mu)\} \cup \bigcup_{(i,\infty,\mu) \in \mathcal{B}_p(\mathcal{F})} \{((f_i, \sigma_i), \infty, \mu)\}.$$

**Complexity:** The complexity of the algorithm is bounded by  $(sd)^{k^{O(\ell)}}$ , where  $s = \text{card}(\mathcal{P})$ , and  $d = \max_{Q \in \mathcal{P} \cup \{P\}} \deg(Q)$ .

*Proof of correctness.* The correctness of the algorithm follows from the correctness of Algorithms 3 and 4.  $\square$

*Complexity analysis.* The complexity bound follows from the complexity bounds of Algorithms 3 and 4.  $\square$

*Proof of Theorem 1.* The theorem follows from the correctness and the complexity analysis of Algorithm 5.  $\square$

## 5. FUTURE WORK AND OPEN PROBLEMS

We conclude by stating some open problems and possible future directions of research in this area.

1. It would be very interesting (and challenging) to obtain an algorithm with singly exponential complexity that computes the entire barcode of a semi-algebraic filtration, and not restricted to dimension up to  $\ell$ . This would imply also an algorithm with singly exponential complexity for computing all the Betti numbers of a given semi-algebraic set, which is a challenging problem on its own [1].
2. Another open problem is to extend Algorithm 5 to the case of non-proper semi-algebraic maps using the proposed definition of barcodes for non-proper semi-algebraic maps (see Definition 2.8).
3. One very active topic in the area of persistent homology is the theory of multi-dimensional persistent homology [18]. In our setting this would imply studying the sub-level sets of two or more real polynomial functions simultaneously. While the so called persistence modules and associated barcodes can be defined analogously to the one-dimensional situation (see for example [18]), an analog of Proposition 3.1 is missing. It is thus an open problem to give an algorithm with

singly exponential complexity to compute the barcodes of “higher dimensional” semi-algebraic filtrations.

#### ACKNOWLEDGEMENTS

The authors are grateful to Ezra Miller for his comments on a previous version of this paper and for pointing out several related prior works.

#### REFERENCES

1. S. Basu, *Algorithms in real algebraic geometry: a survey*, Real algebraic geometry, Panor. Synthèses, vol. 51, Soc. Math. France, Paris, 2017, pp. 107–153. MR 3701212 [5](#), [31](#)
2. S. Basu, R. Pollack, and M.-F. Roy, *Computing roadmaps of semi-algebraic sets on a variety*, J. Amer. Math. Soc. **13** (2000), no. 1, 55–82. MR 1685780 (2000h:14048) [2](#)
3. ———, *Computing the Euler-Poincaré characteristics of sign conditions*, Comput. Complexity **14** (2005), no. 1, 53–71. MR 2134045 (2006a:14095) [2](#)
4. ———, *Computing the first Betti number of a semi-algebraic set*, Found. Comput. Math. **8** (2008), no. 1, 97–136. [2](#)
5. Saugata Basu, *On bounding the Betti numbers and computing the Euler characteristic of semi-algebraic sets*, Discrete Comput. Geom. **22** (1999), no. 1, 1–18. [2](#)
6. ———, *Computing the first few Betti numbers of semi-algebraic sets in single exponential time*, J. Symbolic Comput. **41** (2006), no. 10, 1125–1154. MR 2262087 (2007k:14120) [2](#), [3](#), [4](#)
7. Saugata Basu and Negin Karisani, *Efficient simplicial replacement of semi-algebraic sets*, arXiv e-prints, arXiv:2009.13365 (2020). [4](#), [27](#), [28](#), [29](#)
8. Saugata Basu and Sarah Percival, *Efficient computation of a semi-algebraic basis of the first homology group of a semi-algebraic set*, arXiv eprint, arxiv:2107.08947 (2021). [4](#)
9. Saugata Basu, Richard Pollack, and Marie-Françoise Roy, *Algorithms in real algebraic geometry*, second ed., Algorithms and Computation in Mathematics, vol. 10, Springer-Verlag, Berlin, 2006. MR 1998147 (2004g:14064) [4](#), [5](#), [14](#), [15](#), [17](#), [19](#), [24](#), [25](#), [26](#), [27](#), [29](#)
10. Saugata Basu and Nicolai Vorobjov, *On the number of homotopy types of fibres of a definable map*, J. Lond. Math. Soc. (2) **76** (2007), no. 3, 757–776. MR 2377123 [4](#), [14](#), [17](#), [18](#)
11. L. Blum, F. Cucker, M. Shub, and S. Smale, *Complexity and real computation*, Springer-Verlag, New York, 1998, With a foreword by Richard M. Karp. MR 1479636 (99a:68070) [5](#)
12. J. Bochnak, M. Coste, and M.-F. Roy, *Géométrie algébrique réelle (second edition in english: Real algebraic geometry)*, Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas ], vol. 12 (36), Springer-Verlag, Berlin, 1987 (1998). MR 949442 (90b:14030) [13](#)
13. Peter Bubenik and Jonathan A. Scott, *Categorification of persistent homology*, Discrete Comput. Geom. **51** (2014), no. 3, 600–627. MR 3201246 [3](#), [8](#)
14. Peter Bürgisser, Felipe Cucker, and Pierre Laires, *Computing the homology of basic semialgebraic sets in weak exponential time*, J. ACM **66** (2019), no. 1, Art. 5, 30, [Publication date initially given as 2018]. MR 3892564 [5](#)
15. Peter Bürgisser, Felipe Cucker, and Josué Tonelli-Cueto, *Computing the homology of semialgebraic sets. I: Lax formulas*, Found. Comput. Math. **20** (2020), no. 1, 71–118. MR 4056926 [5](#)
16. Peter Bürgisser, Felipe Cucker, and Josué Tonelli-Cueto, *Computing the homology of semi-algebraic sets. II: General formulas*, Found. Comput. Math. **21** (2021), no. 5, 1279–1316. [5](#)
17. J. Canny, *Computing road maps in general semi-algebraic sets*, The Computer Journal **36** (1993), 504–514. [2](#)
18. Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot, *The structure and stability of persistence modules*, SpringerBriefs in Mathematics, Springer, [Cham], 2016. MR 3524869 [8](#), [31](#)
19. M. Coste and M. Shiota, *Thom’s first isotopy lemma: a semialgebraic version, with uniform bound*, Real analytic and algebraic geometry (Trento, 1992), de Gruyter, Berlin, 1995, pp. 83–101. MR 1320312 (96i:14047) [18](#)

20. William Crawley-Boevey, *Decomposition of pointwise finite-dimensional persistence modules*, J. Algebra Appl. **14** (2015), no. 5, 1550066, 8 pp. MR 3323327 [8](#)
21. Hans Delfs and Manfred Knebusch, *On the homology of algebraic varieties over real closed fields*, J. Reine Angew. Math. **335** (1982), 122–163. MR 667464 [11](#)
22. Tamal Krishna Dey and Yusu Wang, *Computational topology for data analysis*, Cambridge University Press, 2022. [2](#), [11](#)
23. Herbert Edelsbrunner and John Harer, *Persistent homology—a survey*, Surveys on discrete and computational geometry, Contemp. Math., vol. 453, Amer. Math. Soc., Providence, RI, 2008, pp. 257–282. MR 2405684 (2009h:55003) [3](#), [6](#)
24. Herbert Edelsbrunner and John L. Harer, *Computational topology*, American Mathematical Society, Providence, RI, 2010, An introduction. MR 2572029 [8](#), [11](#), [14](#), [22](#)
25. Graham Ellis and Simon King, *Persistent homology of groups*, J. Group Theory **14** (2011), no. 4, 575–587. MR 2818950 [2](#)
26. A. Gabrielov and N. Vorobjov, *Approximation of definable sets by compact families, and upper bounds on homotopy and homology*, J. Lond. Math. Soc. (2) **80** (2009), no. 1, 35–54. MR 2520376 [2](#)
27. Robert Ghrist, *Barcodes: the persistent topology of data*, Bull. Amer. Math. Soc. (N.S.) **45** (2008), no. 1, 61–75. MR 2358377 [2](#), [3](#)
28. M. Goresky and R. MacPherson, *Stratified Morse theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 14, Springer-Verlag, Berlin, 1988. MR 932724 (90d:57039) [18](#)
29. D. Grigoriev and N. Vorobjov, *Counting connected components of a semi-algebraic set in subexponential time*, Comput. Complexity **2** (1992), no. 2, 133–186. [2](#)
30. R. Hardt, *Semi-algebraic local-triviality in semi-algebraic mappings*, Amer. J. Math. **102** (1980), no. 2, 291–302. MR 564475 (81d:32012) [4](#)
31. Masaki Kashiwara and Pierre Schapira, *Persistent homology and microlocal sheaf theory*, J. Appl. Comput. Topol. **2** (2018), no. 1-2, 83–113. MR 3873181 [4](#)
32. Yuri Manin and Matilde Marcolli, *Homotopy theoretic and categorical models of neural information networks*, 2020. [2](#)
33. Ezra Miller, *Data structures for real multiparameter persistence modules*, (2017). [4](#), [11](#)
34. ———, *Essential graded algebra over polynomial rings with real exponents*, arXiv:2008.03819 (2020). [4](#)
35. ———, *Homological algebra of modules over posets*, arXiv:2008.00063 (2020). [4](#)
36. J. Milnor, *On the Betti numbers of real varieties*, Proc. Amer. Math. Soc. **15** (1964), 275–280. MR 0161339 (28 #4547) [2](#)
37. I. G. Petrovskii and O. A. Oleĭnik, *On the topology of real algebraic surfaces*, Izvestiya Akad. Nauk SSSR. Ser. Mat. **13** (1949), 389–402. MR 0034600 (11,613h) [2](#)
38. R. Thom, *Sur l’homologie des variétés algébriques réelles*, Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse), Princeton Univ. Press, Princeton, N.J., 1965, pp. 255–265. MR 0200942 (34 #828) [2](#)
39. Shmuel Weinberger, *What is...persistent homology?*, Notices Amer. Math. Soc. **58** (2011), no. 1, 36–39. MR 2777589 [2](#), [6](#)

DEPARTMENT OF MATHEMATICS, PURDUE UNIVERSITY, WEST LAFAYETTE, IN 47906, U.S.A.  
 Email address: [sbasu@math.purdue.edu](mailto:sbasu@math.purdue.edu)

DEPARTMENT OF COMPUTER SCIENCE, PURDUE UNIVERSITY, WEST LAFAYETTE, IN 47906, U.S.A.  
 Email address: [nkarisan@cs.purdue.edu](mailto:nkarisan@cs.purdue.edu)