Sheaf-Theoretic Stratification Learning

Adam Brown*
University of Utah

Bei Wang†
University of Utah

arXiv:1712.07734v1 [cs.CG] 20 Dec 2017

Abstract

In this paper, we investigate a sheaf-theoretic interpretation of stratification learning. Motivated by the work of Alexandroff [1] and McCord [14], we aim to redirect efforts in the computational topology of triangulated compact polyhedra to the much more computable realm of sheaves on partially ordered sets. Our main result is the construction of stratification learning algorithms framed in terms of a sheaf on a partially ordered set with the Alexandroff topology. We prove that the resulting decomposition is the *unique minimal* stratification for which the strata are homogeneous and the given sheaf is constructible. In particular, when we choose to work with the local homology sheaf, our algorithm gives an alternative to the local homology transfer algorithm given in [5], and the cohomology stratification algorithm given in [16]. We envision that our sheaf-theoretic algorithm could give rise to a larger class of stratification beyond homology-based stratification. This approach also points toward future applications of sheaf theory in the study of topological data analysis by illustrating the utility of the language of sheaf theory in generalizing existing algorithms.

*E-mail: abrown@math.utah.edu.
†E-mail: beiwang@sci.utah.edu.

1 Introduction

Our work is motivated by the following question: Given potentially high-dimensional point cloud samples, can we infer the structures of the underlying data? In the classic setting of *manifold learning*, we often assume the support for the data is from a low-dimensional space with manifold structure. However, in practice, a significant amount of interesting data contains mixed dimensionality and singularities. To deal with this more general scenario, we assume the data are sampled from a mixture of possibly intersecting manifolds; the objective is to recover the different pieces, often treated as clusters, of the data associated with different manifolds of varying dimensions. Such an objective gives rise to a problem of particular interest in the field of *stratification learning*.

Previous work in mathematics has focused on the study of stratified spaces under smooth and continuous settings [10, 20] without computational considerations of noisy and discrete datasets. Statistical approaches that rely on inferences of mixture models and local dimension estimation require strict geometric assumptions such as linearity [11, 13, 19], and may not handle general scenarios with complex singularities. Recently, approaches from topological data analysis [3, 5, 18], which rely heavily on ingredients from computational [7] and intersection homology [8, 2, 4], are gaining momentum in stratification learning.

Topological approaches transform the smooth and continuous setting favored by topologists to the noisy and discrete setting familiar to computational topologists in practice. In particular, the local structure of a point cloud (sampled from a stratified space) can be described by a multi-scale notion of local homology [3]; and the point cloud data could be clustered by strata based on how the local homology of nearby sampled points map into one another [5]. Recently, Nanda [16] employs the notion of local cohomology in the language of cellular cosheaves to recover a canonical stratification of a given regular CW complex.

As our work is an interplay between sheaf theory and stratification, we briefly review various notions of stratification before describing our main results.

1.1 Stratifications

Given a topological space X, a *topological stratification* of X is a finite filtration, that is, an increasing sequence of closed subspaces $\emptyset = X_{-1} \subset X_0 \subset \cdots \subset X_d = X$, such that for each i, $X_i - X_{i-1}$ is a (possibly empty) open i-dimensional topological manifold. See Figure 1 for an illustration in the case of a *pinched torus* example, that is, a pinched torus with a spanning disc stretched across the hole.

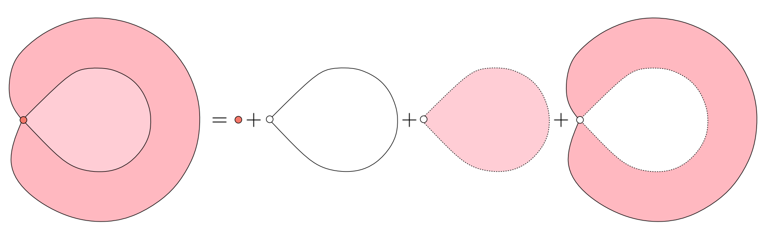

Figure 1: A stratification of a *pinched torus* example.

Ideally, we would like to compute a topological stratification for a given space. However, if we are restricted to only using homological methods, this is a dubious task. Topological invariants like homology are too rough to detect when a space such as $X_i - X_{i-1}$ is an open i-manifold. Therefore, we must adapt our definition of stratification to a more general setting. We begin with

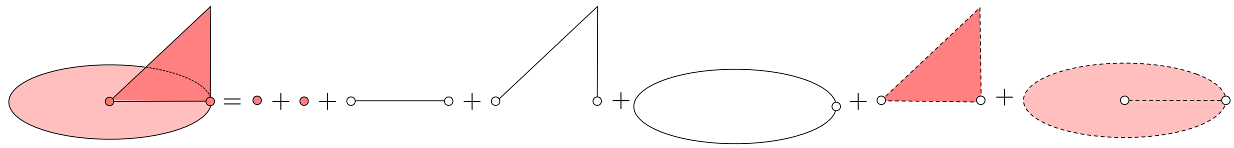

Figure 2: A stratification of a *sundial* example.

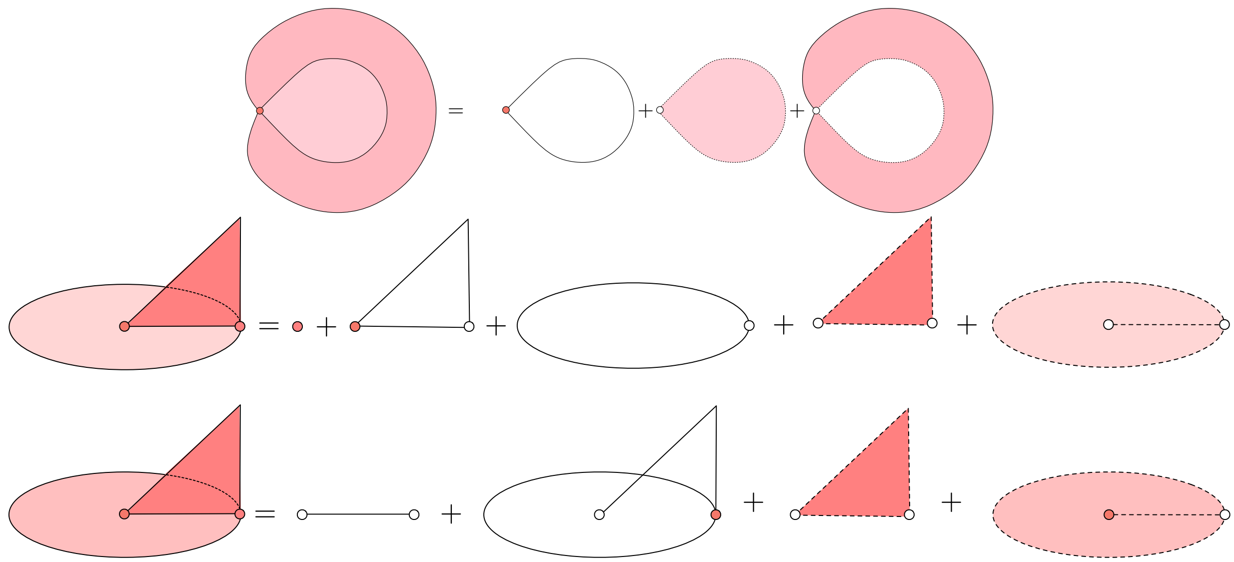

Figure 3: Top: A stratification of a *pinched torus* based on cohomology stratification in [16] in comparison to Figure 1. Middle: A stratification of a *sundial* based on [16] in comparison to Figure 2. Bottom: A different stratification of a *sundial* based on local homology transfer in [5].

an extremely loose definition of stratification which only requires the properties necessary to discuss constructibility of sheaves.

Definition 1.1. Given a topological space X, a *stratification* $\mathfrak{X}$ of X is a finite filtration

$$\emptyset = X_{-1} \subset X_0 \subset \cdots \subset X_d = X,$$

such that for each i, $X_i - X_{i-1}$ is a locally closed subspace of X. We say a subset $U \subset X$ is *locally closed* if it is the intersection of an open and a closed set in X. We refer to the space $X_i - X_{i-1}$ as *stratum*, denoted by S_i, and a connected component of S_i as a *stratum piece*.

Suppose we have two stratifications of the topological space X, denoted $\mathfrak{X}$ and $\mathfrak{X}'$. We say that $\mathfrak{X}$ is equivalent to $\mathfrak{X}'$ if each stratum piece of $\mathfrak{X}$ is equal to a stratum piece of $\mathfrak{X}'$.

Definition 1.2. Given two inequivalent stratifications of X, $\mathfrak{X}$ and $\mathfrak{X}'$, we say $\mathfrak{X}$ is *coarser* than $\mathfrak{X}'$ if each stratum piece of $\mathfrak{X}'$ is contained in a stratum piece of $\mathfrak{X}$.

Figure 3 illustrates examples of stratifications based on cohomology stratification framework [16] for the pinched torus (top) and sundial (middle); as well as a different stratification based on local homology transfer algorithm in [5] for the sundial (bottom).

Homological stratification. There have been several approaches in topology literature to defining homological stratifications. While proving the topological invariance of intersection homology, Goresky and MacPherson defined a type of homological stratification which they call a $\bar{p}$-stratification [9]. Their approach has been cemented as a powerful tool for studying topological pseudomanifolds. However, many topological spaces derived from data science are not pseudomanifolds. For example, a pseudomanifold must satisfy the assumption that the union of top dimensional strata are dense in the space. While this is not troubling for examples coming from complex algebraic varieties, it does rule out examples such as a 2-dimensional plane with a line intersecting the plane at a point.

There have been several approaches for adapting the ideas of Goresky and MacPherson to the setting of topological data analysis. One could change the definition of intersection homology so that it can be used to study a broader class of spaces, as in [4]. Alternatively, one could broaden the definition of homological stratifications. This is the approach we will take. It is a natural choice to define a homological stratification to be a filtration of a space such that the local homology groups are isomorphic for each pair of points in a stratum. This approach could be described as a version (albeit an oversimplified version) of the cohomological stratification given in [16]. The utility of this approach is the extent to which it lends itself to the study of topological properties of individual strata. For example, it is can be easily shown that the strata of such a stratification are R-(co)homology manifolds (R being any ring).

However, we choose to adopt a more refined notion of homological stratification, following Rourke and Sanderson [17]. We say that a stratification is a *homological stratification* if the local homology sheaf is locally constant when restricted to each stratum. This definition utilizes not only information about local homology groups, but information about the restriction maps between local homology groups as well. Therefore, in many situations, we will recover a finer stratification than would be obtained in [16].

$\mathcal{F}$-stratification. The definition of homological stratification naturally lends itself to generalizations, which we now introduce (while delaying formal definition of constructible sheafs to Section 2.2).

Definition 1.3. Suppose $\mathcal{F}$ is a sheaf on a topological space X. An *$\mathcal{F}$-stratification* ("sheaf-stratification") of X is a stratification such that $\mathcal{F}$ is constructible with respect to $X = \amalg S_i$. A *coarsest $\mathcal{F}$-stratification* is an $\mathcal{F}$-stratification such that $\mathcal{F}$ is not constructible with respect to any coarser stratification.

For general topological spaces, a coarsest $\mathcal{F}$-stratification may not exists, and may not be unique if it does exist. The main focus of this paper will be proving existence and uniqueness results for certain coarsest $\mathcal{F}$-stratifications.

1.2 Our Contribution

In this paper, we study stratification learning using the tool of constructible sheaves. As a sheaf is designed to systematically track locally defined data attached to the open sets of a topological space, it seems to be a natural tool in the study of stratification based on local structure of the data. Our contributions are three-fold:

1. We prove the existence of coarsest $\mathcal{F}$-stratifications and the existence and uniqueness of the minimal homogeneous $\mathcal{F}$-stratification for finite T_0-spaces (Section 3).
2. We give an algorithm for computing each of the above stratification of a finite T_0-space based on a sheaf-theoretic language (Section 4).

3. In particular, when applying the local homology sheaf in our algorithm, we obtain a coarsest homological stratification (Section 5.2).

4. In addition, we envision that our abstraction could give rise to a larger class of stratification beyond homological stratification. For instance, we give an example of a "maximal element-stratification" when the sheaf is defined by considering maximal elements of an open set (see Section 6).

Comparison to prior work. This paper can be viewed as a continuation of previous works that aim to adapt the stratification and homology theory of Goresky and MacPherson to the realm of topological data analysis. In [17], Rourke and Sanderson give a proof of the topological invariance of intersection homology on PL homology stratifications, and give an recursive process for identifying a homological stratification (defined in Section 5 of [17]). In [4], Bendich and Harer introduce a persistent version of intersection homology that can be applied to simplicial complexes. In [5], Bendich, Wang, and Mukherjee provide computational approach that yields a stratification of point clouds by computing transfer maps between local homology groups of an open covering of the point cloud. In [16], Nanda uses the machinery of derived categories to study cohomological stratifications based on local cohomology.

Motivated by the discrepancies in the stratifications given by [16] and [5] (see Figure 3), we return to the approach of [17]. Our main results can be summarized as the generalization of homological stratifications of [17] to $\mathcal{F}$-stratifications, and a proof of existence and uniqueness of the coarsest $\mathcal{F}$-stratification of finite T_0-spaces. When $\mathcal{F}$ is the local homology sheaf, we recover the homological stratification described by [17]. While the results described in this paper give the same stratification as [5] for many examples (such as the pinched torus), the current algorithm gives reasonable stratifications for a larger class of topological spaces. For example, the stratification of the *sundial* example (i.e. a stratified space with boundary) given by [5] (see Figure 3) is not technically a stratification by our definition, since the 2-dimensional strata is not locally closed. In comparison, the current algorithm correctly gives the unique coarsest stratification of this space. While admitting a similar flavor as [16], our work differs from [16] in several important ways. Primarily, we give a different (finer) stratification for many examples (see Figures 1, 2, and 3). In addition, our algorithm only requires local homology to be calculated once. By comparison, the algorithm described in [16] requires local homology to be recalculated each time a stratum is defined. Therefore our algorithm is much less computationally expensive, while retaining the distributivity of the algorithm in [16].

2 Preliminaries

2.1 Compact Polyhedra, Finite T_0-spaces and Posets

Our broader aim is to compute a clustering of a finite set of points sampled from a compact polyhedron, based on the coarsest $\mathcal{F}$-stratification of a finite T_0-space built from the point set. In this paper, we avoid discussion of sampling theory, and assume the finite point set forms the vertex set of a triangulated compact polyhedron. The finite T_0-space is the set of simplices of the triangulation, with the corresponding partial order. To describe this correspondence in more detail, we first consider the connection between compact polyhedra and finite simplicial complexes. We then consider the correspondence between simplicial complexes and T_0-topological spaces.

Compact polyhedra and triangulations. A *compact polyhedron* is a topological space which is homeomorphic to a finite simplicial complex. A *triangulation* of a compact polyhedron is a finite simplicial complex K and a homeomorphism from K to the polyhedron.

T_0-spaces. A T_0-*space* is a topological space such that for each pair of distinct points, there exists an open set containing one but not the other. Its correspondence with simplicial complex is detailed in [14]:

1. For each finite T_0-space X there exists a (finite) simplicial complex K and a weak homotopy equivalence $f : |K| \to X$.

2. For each finite simplicial complex K there exists a finite T_0-space X and a weak homotopy equivalence $f : |K| \to X$.

Here, weak homotopy equivalence is a continuous map which induces isomorphisms on all homotopy groups.

T_0-spaces have a natural partial order. In this paper, we study certain topological properties of a compact polyhedron by considering its corresponding finite T_0-space. The last ingredient, developed in [1], is a natural partial order defined on a given finite T_0-space. We can define this partial ordering on a finite T_0-space X by considering minimal open neighborhoods of each point (i.e. element) $x \in X$. Let X be a finite T_0-space. Each point $x \in X$ has a minimal open neighborhood, denoted B_x, which is equal to the intersection of all open sets containing x.

$$B_x = \bigcap_{U \in \mathcal{N}_x} U$$

where $\mathcal{N}_x$ denotes the set of open sets containing x. Since X if a finite space, there are only finitely many open sets. In particular, $\mathcal{N}_x$ is a finite set. So B_x is defined to be the intersection of finitely many open sets, which implies that B_x is an open neighborhood of x. Moreover, any other open neighborhood V of x must contain B_x as a subset. We can define the partial ordering on X by setting $x \leq y$ if $B_y \subseteq B_x$.

Conversely, we can endow any poset X with the Alexandroff topology as follows. For each element $\tau \in X$, we define a minimal open neighborhood containing τ by $B_\tau := \{\gamma \in X : \gamma \geq \tau\}$. The collection of minimal open neighborhoods for each $\tau \in X$ forms a basis for a topology on X. We call this topology the Alexandroff topology. Moreover, a finite T_0-space X is naturally equal (as topological spaces) to X viewed as a poset with the Alexandroff topology. Therefore, we see that each partially ordered set is naturally a T_0-space, and each finite T_0-space is naturally a partially ordered set. The purpose for reviewing this correspondence here is to give the abstractly defined finite T_0-spaces a concrete and familiar realization.

Given a finite T_0-space X with the above partial order, we say $x_0 \leq x_1 \leq \cdots \leq x_n$ is a maximal chain in X if there is no set $\{y_i\} \subset X$ such that $y_0 \leq y_1 \leq \cdots \leq y_k$ and $\{x_i\} \subsetneq \{y_i\}$. The cardinality of a chain is the cardinality of the corresponding subset of X. We say that a finite T_0-space has dimension m if the maximal cardinality of maximal chains is $m + 1$. An m-dimensional simplicial complex is called *homogeneous* if each simplex of dimension less than m is a face of a simplex of dimension m. Motivated by this definition and the correspondence between simplicial complexes and T_0-spaces, we say an m-dimensional finite T_0-space is *homogeneous* if each maximal chain has cardinality $m + 1$.

The correspondences allow us to study certain topological properties of compact polyhedra by using the combinatorial theory of partially ordered sets. In particular, instead of using the more complicated theory of sheaves on the geometric realization $|K|$ of a simplicial complex K, we will continue by studying sheaves on the corresponding finite T_0-space, denoted by X.

2.2 Constructible Sheaves

Intuitively, a sheaf assigns some piece of data to each open set in a topological space X, in a way that allows us to glue together data to recover some information about the larger space. This process can be described as the mathematics behind understanding global structure by studying local properties of a space.

Sheaves. Let $\mathcal{C}$ be an abelian category and X be a topological space. Let $\mathcal{F}$ be a contravariant functor from $\mathbf{Top}(X)$ (the category of open sets of X) to $\mathcal{C}$. For open sets $U \subset V$ in X, we refer to the morphism $\mathcal{F}(U \subset V) := \mathcal{F}(\iota : U \to V) : \mathcal{F}(V) \to \mathcal{F}(U)$ induced by $\mathcal{F}$ and the inclusion $U \subset V$, as a *restriction* map from V to U. We say that $\mathcal{F}$ is a $\mathcal{C}$-valued *sheaf* on X (see [12] Chapter 3) if $\mathcal{F}$ satisfies following conditions 1-4; a *presheaf* is a functor $\mathcal{E}$ (as above) which satisfies conditions 1-3:

1. $\mathcal{F}(\emptyset) = 0$;
2. $\mathcal{F}(U \subset U) = \mathrm{id}_U$.
3. If $U \subset V \subset W$, then $\mathcal{F}(U \subset W) = \mathcal{F}(U \subset V) \circ \mathcal{F}(V \subset W)$.
4. If $\{V_i\}$ is an open cover of U, and $s_i \in \mathcal{F}(V_i)$ has the property that $\forall i, j$, $\mathcal{F}((V_i \cap V_j) \subset V_i)(s_i) = \mathcal{F}((V_j \cap V_i) \subset V_j)(s_j)$, then there exists a unique $s \in \mathcal{F}(U)$ such that $\forall i$, $\mathcal{F}(V_i \subset U)(s) = s_i$.

There is a useful process known as *sheafification*, which allows us to transform any presheaf into a sheaf. In the setting of finite T_0-spaces, sheafification takes on a relatively simple form. Let $\mathcal{E}$ be a presheaf on a finite T_0-space X. Then the sheafification of $\mathcal{E}$, denoted $\mathcal{E}^+$, is given by

$$\mathcal{E}^+(U) = \left\{ f : U \to \coprod_{x \in U} \mathcal{E}(B_x) \middle| f(x) \in \mathcal{E}(B_x) \text{ and } f(y) = \mathcal{F}(B_y \subset B_x)(f(x)) \text{ for all } y \geq x \right\}$$

For any presheaf $\mathcal{E}$, it can be seen that $\mathcal{E}^+$ is necessarily a sheaf. We only need to know the values $\mathcal{E}(B_x)$ for minimal open neighborhoods B_x, and the corresponding restriction maps between minimal open neighborhoods $\mathcal{E}(B_x \subset B_y)$, in order to define the sheafification of $\mathcal{E}$. The result is that two presheaves will sheafify to the same sheaf if they agree on all minimal open neighborhoods. We will use this fact several times in Section 3. Unless otherwise specified, for the remaining of this paper, we use X to denote a T_0-space.

Pull back of a sheaf. For notational convenience, define for each subset $Y \subset X$ the *star* of Y by $\mathrm{st}(Y) := \cup_{y \in Y} B_y$, where B_y is the minimal open neighborhood of $y \in X$. We can think of the star of Y as the smallest open set containing Y.

Let X and Y be two finite T_0-spaces. The following property can be thought of as a way to transfer a sheaf on Y to a sheaf on X through a continuous map $f : X \to Y$. Let $\mathcal{F}$ be a sheaf on Y. Then the *pull back* of $\mathcal{F}$, denoted $f^{-1}\mathcal{F}$, is defined to be the sheafification of the presheaf $\mathcal{E}(U) := \mathcal{F}(\mathrm{st}(f(U)))$, for each $U \subset X$.

When we have an inclusion map $\iota : U \hookrightarrow X$, the pull back $\iota^{-1}\mathcal{F}$ is called the *restriction* of $\mathcal{F}$ to U, and is denoted $\mathcal{F}|_U$.

Constant and locally constant sheaves. Now we can define classes of well-behaved sheaves, constant and locally constant ones, which we can think of intuitively as analogues of constant functions based on definitions common to algebraic geometry and topology [12]. A sheaf $\mathcal{F}$ is a *constant sheaf* if $\mathcal{F}$ is isomorphic to the pull back of a sheaf $\mathcal{G}$ on a single point space $\{x\}$, along the projection map $p : X \to x$. A sheaf $\mathcal{F}$ is *locally constant* if for all $x \in X$, there is a neighborhood U of x such that $\mathcal{F}|_U$ (the restriction of $\mathcal{F}$ to U), is a constant sheaf.

Definition 2.1. A sheaf $\mathcal{F}$ on a finite T_0-space X is *constructible* with respect to the decomposition $X = \coprod S_i$ of X into finitely many disjoint locally closed subsets, if $\mathcal{F}|_{S_i}$ is locally constant for each i.

3 Main Results

In this section we state three of our main theorems, namely, the existence of $\mathcal{F}$-stratifications (Theorem 3.1), the existence of coarsest $\mathcal{F}$-stratifications (Theorem 3.2), and the existence and uniqueness of minimal homogeneous $\mathcal{F}$-stratifications (Theorem 3.3). Of course, Theorem 3.2 immediately implies Theorem 3.1. We choose to include a separate statement of Theorem 3.1 however, as we wish to illustrate the existence of $\mathcal{F}$-stratifications which are not necessarily the coarsest. We include proof sketches here and refer to Section 7 for technical details.

Theorem 3.1. *Let $\mathcal{F}$ be a sheaf on a finite T_0-space X. There exists an $\mathcal{F}$-stratification of X (see Definition 1.3 and Definition 2.1).*

Proof Sketch. $\mathcal{F}$ is constructible with respect to the decomposition $X = \coprod_{x\in X} x$. □

Theorem 3.2. *Let $\mathcal{F}$ be a sheaf on a finite T_0-space X. There exists a coarsest $\mathcal{F}$-stratification of X.*

Proof Sketch. We can prove Theorem 3.2 easily as follows. There are only finitely many stratifications of our space X, which implies that there must be an $\mathcal{F}$-stratification with a minimal number of strata pieces. Such a stratification must be a coarsest stratification, since any coarser stratification would have fewer strata pieces.

However, the above proof is rather unenlightening if we are interested in computing the coarsest $\mathcal{F}$-stratification. Therefore we include a constructive proof of the existence of a coarsest $\mathcal{F}$-stratification which we sketch here. We can proceed iteratively, by defining the top-dimensional stratum to be the collection of points (i.e. elements) so that the sheaf is constant when restricted to the minimal open neighborhoods of the said points. We then remove the top-dimensional stratum from our space, and pull back the sheaf to the remaining points. We proceed until all the points in our space have been assigned to a stratum. We can see that this is a coarsest $\mathcal{F}$-stratification by arguing that this algorithm, in some sense, maximizes the size of each stratum piece, and thus any coarser $\mathcal{F}$-stratification is actually equivalent to the one constructed above. We refer the reader to Section 7.2 for the details of the above argument. □

To uniquely identify a stratification by its properties, we will need to introduce a notion of a minimal homogeneous $\mathcal{F}$-stratification.

Definition 3.1. Suppose $\mathcal{F}$ is a sheaf on a finite T_0-space X. A *homogeneous $\mathcal{F}$-stratification* is an $\mathcal{F}$-stratification such that each stratum S_i is homogeneous of dimension i (defined in Section 2.1).

We will introduce a lexicographical preorder on the set of homogeneous $\mathcal{F}$-stratifications of a finite T_0-space X. Let $\mathfrak{X}$ be a homogeneous $\mathcal{F}$-stratification of X given by

$$\emptyset = X_{-1} \subset X_0 \subset X_1 \subset \cdots \subset X_n = X$$

We define a sequence $A^{\mathfrak{X}} := \{|X_n|, \cdots, |X_i|, \cdots, |X_0|\}$. Given two stratifications $\mathfrak{X}$ and $\mathfrak{X}'$, we say that $\mathfrak{X}' > \mathfrak{X}$ if the first non-zero term of the sequence $A^{\mathfrak{X}'} - A^{\mathfrak{X}} = \{|X_i'| - |X_i|\}$ is positive; $\mathfrak{X} = \mathfrak{X}'$ if there are no non-zero terms. Notice that if $\mathfrak{X}$ and $\mathfrak{X}'$ are homogeneous stratifications such that $\mathfrak{X}$ is coarser than $\mathfrak{X}'$, then we necessarily have that $\mathfrak{X} \leq \mathfrak{X}'$. We say that a stratification $\mathfrak{X}$ is a *minimal homogeneous $\mathcal{F}$-stratification* if $\mathfrak{X} \leq \mathfrak{X}'$ for every other homogeneous $\mathcal{F}$-stratification $\mathfrak{X}'$.

Definition 3.2. A homogeneous $\mathcal{F}$-stratification is *minimal* if it is minimal with respect to the lexocographic order on homogeneous $\mathcal{F}$-stratifications.

Theorem 3.3. *Let K be a finite simplicial complex, and X be a finite T_0-space consisting of the simplices of K endowed with the Alexandroff topology. Let $\mathcal{F}$ be a sheaf on X. There exists a unique minimal homogeneous $\mathcal{F}$-stratification of X.*

Proof Sketch. The idea for this proof is very similar to that of the Theorem 3.2. We construct a stratification in a very similar way, with the only difference being that we must be careful to only construct homogeneous strata. The argument for the uniqueness of the resulting stratification uses the observation that this iterative process maximizes the size of the current stratum (starting with the top-dimensional stratum) before moving on to define lower-dimensional strata. Thus the resulting stratification is minimal in the lexocographic order. The top-dimensional stratum of any other minimal homogeneous $\mathcal{F}$-stratification then must equal the top stratum constructed above, since these must both include the set of top-dimensional simplices, and have maximal size. An inductive argument then shows the stratifications are equivalent. Again, we refer readers to Section 7.3 for the remaining details. $\square$

4 A Sheaf-Theoretic Stratification Learning Algorithm

We outline an explicit algorithm for computing the coarsest $\mathcal{F}$-stratification of a space X given a particular sheaf $\mathcal{F}$. We give two examples of stratification learning using the local homology sheaf (Section 5) and the sheaf of maximal elements (Section 6).

Let X be a finite T_0-space, equipped with a partial ordering. Instead of using the sheaf-theoretic language of Theorem 3.3, we frame the computation in terms of X and an "indicator function" δ. For every $x, y \in X$ with a relation $x \leq y$, δ assigns a binary value to the relation. That is, $\delta(x \leq y) = 1$ if the restriction map $\mathcal{F}(B_y \subset B_x) : \mathcal{F}(B_x) \to \mathcal{F}(B_y)$ is an isomorphism, and $\delta(x \leq y) = 0$ otherwise. We say a pair $w \leq y$ is *adjacent* if $w \leq z \leq y$ implies $z = w$ or $z = y$ (in other words, there are no elements in between w and y). Due to condition 3 in the definition of a sheaf (Section 2.2), δ is fully determined by the values $\delta(w \leq y)$ assigned to each adjacent pair (w, y). If $a_1 \leq a_2 \leq \cdots \leq a_k$ is a chain of adjacent elements (a_i is adjacent to a_{i+1} for each i), we have that $\delta(a_1 \leq a_k) = \delta(a_1 \leq a_2) \cdot \delta(a_2 \leq a_3) \cdots \delta(a_{k-1} \leq a_k)$. As X is equipped with a finite partially ordering, computing δ can be interpreted as assigning a binary label to the edges of a Hasse diagram associated with the partial ordering (see Section 5 for an example).

For simplicity, we assume that δ is pre-computed, with a complexity of $O(m)$ where m denotes the number of adjacent relations in X. When X corresponds to a simplicial complex K, m is the number of nonzero terms in the boundary matrices of K. δ can, of course, be processed on-the-fly, which may lead to more efficient algorithm. In addition, determining the value of δ is a local computation for each $x \in X$, therefore it is easily parallelizable.

Computing a coarsest $\mathcal{F}$-stratification. If we are only concerned with calculating a coarsest $\mathcal{F}$-stratification as described in Theorem 3.2, we may use the algorithm below.

1. Set $i = 0$, $d_0 = \dim X$, $X_{d_0} = X$, and initialize $S_j = \emptyset$, for all $0 \leq j \leq d_0$.
2. While $d_i \geq 0$, do
   (a) For each $x \in X_{d_i}$, set $S_{d_i} = S_{d_i} \cup x$ if $\delta(w \leq y) = 1$, $\forall$ adjacent pairs $w \leq y$ in $B_x \cap X_{d_i}$
   (b) Set $d_{i+1} = \dim(X_{d_i} - S_{d_i})$

(c) Define $X_{d_{i+1}} = X_{d_i} - S_{d_i}$

(d) Set $i = i + 1$

3. Return S

Here, i is the step counter; d_i is the dimension of the current strata of interest; the set S_{d_i} is the stratum of dimension d_i. d_i decreases from $\dim(X)$ to 0. To include an element x to the current stratum S_{d_i}, we need to check δ for adjacent relations among all x's cofaces.

Computing the unique minimal homogeneous $\mathcal{F}$-stratification. If we would like to obtain the unique minimal homogeneous $\mathcal{F}$-stratification, then we need to modify step 2a. Let $c(x,i) = 1$ if all maximal chains in X_{d_i} containing x have cardinality d_i, and $c(x,i) = 0$ otherwise. Then the modified version of 2.a. is:

2.a. For each $x \in X_{d_i}$, set $S_{d_i} = S_{d_i} \cup x$ if $\delta(w \leq y) = 1, \forall$ adj. pairs $w \leq y$ in $B_x \cap X_{d_i}$ and $c(x,i) = 1$

5 Stratification Learning with Local Homology Sheaf

5.1 Local Homology Sheaf

For a finite T_0-space X, consider the chain complex $C_\bullet(X)$, where $C_p(X)$ consists of free R-modules generated by $(p+1)$-chains in X, with chain maps $\partial_p : C_p(X) \to C_{p-1}(X)$ given by

$$\partial_p(a_0 \leq \cdots \leq a_p) = \sum (-1)^i (a_0 \leq \cdots \leq \hat{a}_i \leq \cdots \leq a_p)$$

where $\hat{a}_i$ means that the element a_i is to be removed from the chain.

We would like to remark on the decision to refer to this sheaf as the local homology sheaf. If X is a more general topological space (CW space, simplicial complex, manifold, etc), then the *local homology* of X at $x \in X$ is defined to be the direct limit of relative homology $H_\bullet(X, X-x) := \varinjlim H_\bullet(X, X-U)$ (where the direct limit is taken over all open neighborhoods U of x with the inclusion partial order) [15] (page 196). In our setting, the local homology of X (a finite T_0-space) at a point $x \in X$ is given by $H_\bullet(X, X - B_x)$. Here we avoid using notions of direct limit by working with topological spaces that have minimal open neighborhoods. This motivates our decision to refer to the sheaf defined by relative homology $H_\bullet(X, X - U)$ for each open set U (see Theorem 5.1), as the *local homology sheaf* [1].

The following theorem, though straightforward, provides justification for applying the results of Section 4 to local homology computations.

Theorem 5.1. *The functor $\mathcal{L}$ from the category of open sets of a finite T_0-space to the category of graded R-modules, defined by*

$$\mathcal{L}(U) := H_\bullet(X, X - U)$$

where R is the ring of coefficients of the relative homology, is a sheaf on X.

Proof. We first show that conditions 1-3 are satisfied in the definition of sheaf from Section 2. The inclusion of open sets $U \subset V$, and equivalently $X - V \subset X - U$, induce a morphism of graded R-modules,

$$\mathcal{L}(U \subset V) : H_\bullet(X, X - V) \to H_\bullet(X, X - U).$$

We have the following commutative diagram of chain complexes

[1] See [6] for an interesting approach to the computation of homology groups of finite T_0-spaces.

$$\begin{tikzcd}
0 \arrow[r] & C_\bullet(X-V) \arrow[r] \arrow[d] & C_\bullet(X) \arrow[r, "p_1"] \arrow[d, "\mathrm{id}"] & C_\bullet(X)/C_\bullet(X-V) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & C_\bullet(X-U) \arrow[r] & C_\bullet(X) \arrow[r, "p_2"] & C_\bullet(X)/C_\bullet(X-U) \arrow[r] & 0
\end{tikzcd}$$

where the map $C_\bullet(X)/C_\bullet(X-V) \to C_\bullet(X)/C_\bullet(X-U)$ is defined by $p_2 \circ p_1^{-1}$, and is well-defined since $X - V \subset X - U$. For a triple $U \subset V \subset W$, we have the restriction maps

$$H_\bullet(X, X-W) \to H_\bullet(X, X-V) \to H_\bullet(X, X-U)$$

whose composition is equal to $H_\bullet(X, X-W) \to H_\bullet(X, X-U)$. This can be seen by applying our construction of the restriction map above to three short exact sequences of chain complexes. In order to prove condition *iv* in the definition of a sheaf is satisfied, we could apply Mayer-Vietoris sequences for relative homology groups. But considering that we only need to think of $\mathcal{L}$ as a presheaf in order to apply our algorithm, we will not include the details of this part of the proof. $\square$

5.2 An Example of Stratification Learning Using Local Homology Sheaf

If X is a T_0-space corresponding to a simplicial complex K, then the local homology groups in Section 5.1 are isomorphic to the simplicial homology groups of K. We now give a detailed example of stratification learning using local homology sheaf for the sundial example from Figure 2. We will abuse notation slightly, and use K to denote the finite T_0-space consisting of elements which are open simplices corresponding to the triangulated sundial (Figure 4). We choose this notation so that we can describe our T_0-space using the more familiar language of simplicial complexes. For a simplex $\sigma \in K$, its minimal open neighborhood B_σ is its *star* consisting of all cofaces of σ, $\mathrm{St}(\sigma) = \{\tau \in K \mid \sigma \leq \tau\}$. The *closed star*, $\overline{\mathrm{St}}(\sigma)$, is the smallest subcomplex that contains the star. The *link* consists of all simplices in the closed star that are disjoint from the star, $\mathrm{Lk}(\sigma) = \{\tau \in \overline{\mathrm{St}}(\sigma) \mid \tau \cap \mathrm{St}(\sigma) = \emptyset\}$. K is equipped with a partial order based on face relations, where $x < y$ if x is a proper face of y. This partial order gives rise to a Hasse diagram illustrated in Figure 5.

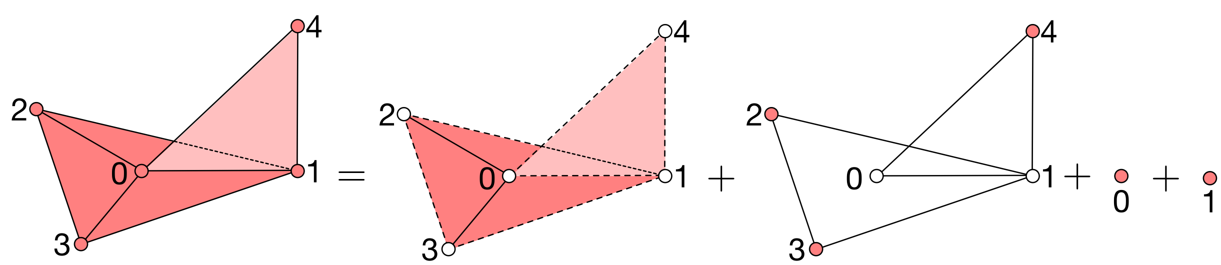

Figure 4: A triangulated sundial and its stratification based on the local homology sheaf.

A sheaf on K can be considered as a labeling of each vertex in the Hasse diagram with an object in an abelian category $\mathcal{C}$ and each edge with a morphism between the corresponding objects. Consider the local homology sheaf $\mathcal{L}$ on K which takes each open set $U \subset K$ [2] to $H_\bullet(K, K-U) = H_\bullet(\mathrm{Cl}(U), \mathrm{Lk}(U))$, where $\mathrm{Lk}(U) := \mathrm{Cl}(U) - U$. Our algorithm described in Section 4 can then be interpreted as computing local homology sheaf associated with each vertex

[2] In the finite simplicial setting, U is the support of a union of open simplices in K.

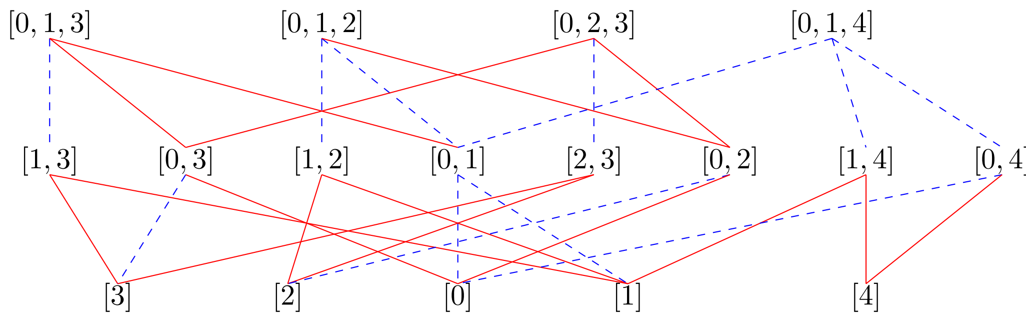

Figure 5: The Hasse diagram of the triangulated sundial. For any two simplex $\tau^{(p-1)} < \sigma^{(p)}$, an edge between τ and σ in the diagram is in solid red, if $\mathcal{L}(B_\tau) \to \mathcal{L}(B_\sigma)$ is an isomorphism; otherwise it is in dotted blue.

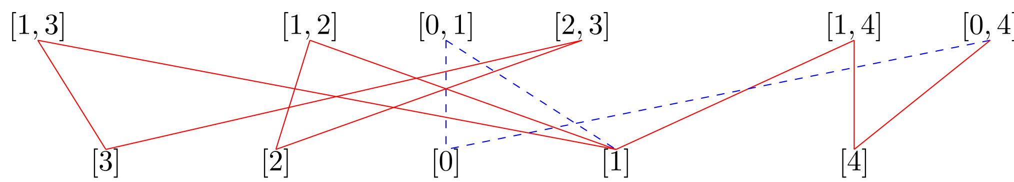

Figure 6: The Hasse diagram after the top dimensional stratum has been removed. We can consider this the beginning of the second iteration of the algorithm in Section 4.

in the Hasse diagram, and determining whether each edge in the diagram is an isomorphism. Our algorithm works by considering an element σ in the Hasse diagram to be in the top-dimensional strata if all of the edges above σ are isomorphisms, that is, if $\mathcal{L}(\tau < \sigma)$ is an isomorphism $\forall \sigma > \tau$.

First, we start with 2-simplexes. Automatically, we have that $\mathcal{L}$ is constant when restricted to any 2-simplex, and gives homology groups isomorphic to the homology of a 2-sphere. For instance, the local homology groups of a 2-simplex $\sigma = [0, 1, 3]$ is isomorphic to $H_\bullet(\mathrm{Cl}(\sigma), \mathrm{Lk}(\sigma)) \cong H_\bullet(S^2)$.

Second, we consider the restriction of $\mathcal{L}$ to the minimal open neighborhood of a 1-simplex. For instance, consider the 1-simplex $[1, 3]$; $B_{[1,3]} = [1, 3] \cup [0, 1, 3]$. It can be seen that $\mathrm{Lk}(B)_{[1,3]} = [0] \cup [3] \cup [1] \cup [0, 3] \cup [0, 1]$, and $H_\bullet(\mathrm{Cl}(B_{[1,3]}), \mathrm{Lk}(B_{[1,3]}))$ is isomorphic to the homology of a single point space. Therefore the restriction map $\mathcal{L}(B_{[1,3]}) \to \mathcal{L}(B_{[0,1,3]})$ is not an isomorphism (illustrated as a dotted blue line in Figure 5). On the other hand, let us consider the 1-simplex $[0, 3]$, where $B_{[0,3]} = [0, 3] \cup [0, 1, 3] \cup [0, 2, 3]$. We have that $\mathrm{Lk}(B_{[0,3]}) = [0] \cup [1] \cup [2] \cup [3] \cup [0, 1] \cup [0, 2] \cup [1, 3] \cup [2, 3] \cup [1, 3]$. Therefore $\mathcal{L}(B_{[0,3]})$ is isomorphic to the homology of a 2-sphere. Moreover, both of the restriction maps corresponding to $B_{[0,1,3]} \subset B_{[0,3]}$ and $B_{[0,2,3]} \subset B_{[0,3]}$ are isomorphisms (illustrated as solid red lines in Figure 5). This implies that $[0, 3] \subset S_2 = X_2 - X_1$. If we continue, we see that the top dimensional strata is given by $S_2 = [0, 1, 3] \cup [0, 1, 2] \cup [0, 2, 3] \cup [0, 1, 4] \cup [0, 2] \cup [0, 3]$, see Figure 4.

Next, we can calculate the stratum $S_1 = X_1 - X_0$ by only considering restriction maps whose codomain is not contained in S_2 (see Figure 6). We get $S_1 = [0, 1] \cup [1, 3] \cup [1, 2] \cup [2, 3] \cup [0, 4] \cup [1, 4] \cup [2] \cup [3] \cup [4]$. We can visualize S_1 in Figure 4. This stratification is finer than the one following Nanda's framework [16], as $[0]$ is contained in the 1-stratum in Figure 3 (middle), but is not contained in S_1 presently.

Finally, the strata $S_0 = X_0$ consists of the vertices which have not been assigned to a strata yet. So $S_0 = [0] \cup [1]$. Observe that (for this example) the coarsest $\mathcal{L}$-stratification we calculated is actually the unique minimal homogeneous $\mathcal{L}$-stratification. We would like to investigate this coincidence for $\mathcal{L}$-stratifications elsewhere.

6 Stratification Learning with Sheaf of Maximal Elements

We will now consider a stratification of the triangulated sundial given by the sheaf of maximal elements. Again, let $|K|$ be the polyhedron in Figure 4 with labeled vertices. We construct the labeled Hasse diagram based on the sheaf of maximal elements, as shown in Figure 7.

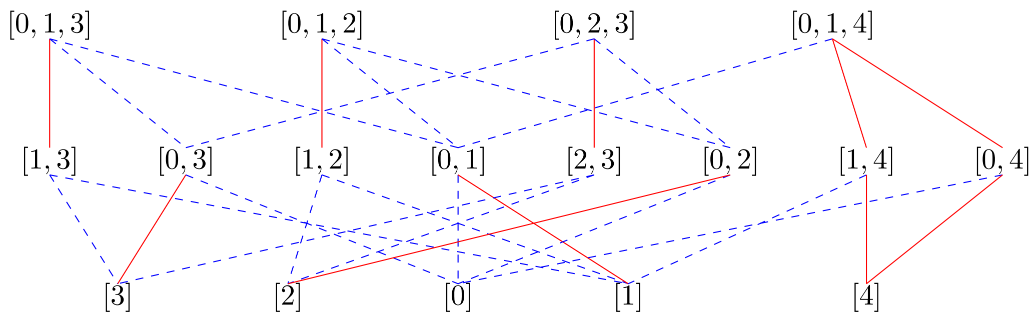

Figure 7: The labeled Hasse diagram for the sundial with the sheaf of maximal elements. Red solid lines denote isomorphisms; while blue dotted lines denote non-isomorphisms.

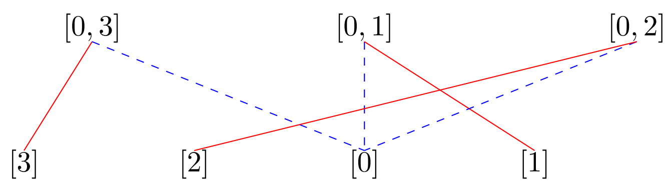

Figure 8: The labeled Hasse diagram for the sundial with the sheaf of maximal elements after the top dimensional stratum has been removed.

Sheaf of maximal elements. Consider the sheaf $\mathcal{F}$ on the space K which takes each open set $U \subset K$ to the free $\mathbb{Z}$-module generated by maximal elements of U. For $V \subset U$, $\mathcal{F}(U) \to \mathcal{F}(V)$ maps an element $u \in U$ to $u \in V$ if $u \in V$ and 0 otherwise.

Stratification learning using sheaf of maximal elements. Now for the triangulated sundial, $\mathcal{F}$ is automatically constant when restricted to any of the 2-simplices.

Let us consider the restriction of $\mathcal{F}$ to the minimal open set containing $[1,3]$, that is, $B_{[1,3]} = [1,3] \cup [0,1,3]$. The restriction map $\mathcal{F}(B_{[1,3]}) \to \mathcal{F}(B_{[0,1,3]})$ sends the only maximal element in $B_{[1,3]}$ to the only maximal element in $B_{[0,1,3]}$, and therefore is an isomorphism from $\mathbb{Z}$ to $\mathbb{Z}$.

Let us consider a more subtle example. The minimal open set containing $[0,3]$ is given by $B_{[0,3]} = [0,3] \cup [0,1,3] \cup [0,2,3]$. We see that there are two distinct maximal elements, $[0,1,3]$ and $[0,2,3]$. Therefore $\mathcal{F}(B_{[0,3]}) = \mathbb{Z}^2$. This means that neither of the restriction maps can be isomorphisms, since $\mathcal{F}(B_{[0,1,3]})$ and $\mathcal{F}(B_{[0,2,3]})$ are each isomorphic to $\mathbb{Z}$.

If we continue, we see that the top stratum is given by

$$S_2 = X_2 - X_1 = [0,1,3] \cup [0,1,2] \cup [0,2,3] \cup [0,1,4] \cup [1,3] \cup [1,2] \cup [2,3] \cup [1,4] \cup [0,4].$$

$S_2 \subset |K|$ is illustrated in Figure 9.

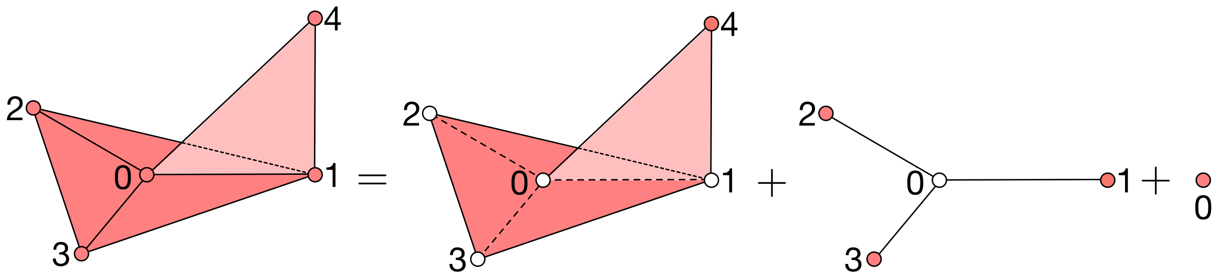

Figure 9: A triangulated sundial and its stratification based on the sheaf of maximal elements.

Next, we can calculate the strata S_1 by only considering restriction maps whose codomain is not contained in S_2 (see Figure 8). We get

$$S_1 = X_1 - X_0 = [0, 1] \cup [0, 2] \cup [0, 3] \cup [1] \cup [2] \cup [3]$$

We can consider the Hasse diagram and corresponding visualization of S_1, as illustrated in Figure 9.

Finally, the strata X_0 in the coarsest $\mathcal{F}$-stratification consists of the points which have not been assigned to a strata yet. So

$$S_0 = X_0 = [0].$$

Intuitively, we are using this relatively simple sheaf to cluster the space $|K|$ into p-skeletons ($p = 0, 1, 2$), with a subtle difference being that if a lower dimensional simplex appears as a face of a unique p-simplex, it will be included in the p-skeleton.

7 Proofs of Our Main Results

We detail the proofs of our main theorems, that is, the existence of $\mathcal{F}$-stratifications (Theorem 3.1), the existence of coarsest $\mathcal{F}$-stratifications (Theorem 3.2), and the existence and uniqueness of minimal homogeneous $\mathcal{F}$-stratifications (Theorem 3.3).

7.1 Proof of Theorem 3.1

Proof. We can take the finest filtration of X, so that each $X_i - X_{i-1}$ consists of a single point (i.e. element) $S_i = X_i - X_{i-1} = x_i \in X$. Then $X = \coprod_{x_i \in X} x_i$, and each point $x_i \in X$ is locally closed since it is the intersection of an open and a closed set. Let B_x denote the minimal open neighborhood of x, then $x = B_x \cap (\cup_{y>x} B_y)^c$, for all $y > x$. Since X is equipped with a finite partially ordered set, this gives a decomposition $X = \coprod_{x_i \in X} x_i$ of X into finitely many locally closed subsets. Now we wish to show that $\mathcal{F}|_{x_i}$ is locally constant for each $x_i \in X$. This is trivial, and in fact $\mathcal{F}|_{x_i}$ is a constant sheaf, since it is a sheaf defined on a topological space consisting of a single point. Therefore $\mathcal{F}$ is constructible with respect to the single point decomposition $X = \coprod_{x_i \in X} x_i$. □

7.2 Proof of Theorem 3.2

Proof. This theorem can be proved immediately by noticing that there are only finitely many stratifications of X (X being a finite T_0-space with finite many points). Since the set of $\mathcal{F}$-stratifications is nonempty, there must be an $\mathcal{F}$-stratification with a minimal number of strata pieces, and such a stratification must be a coarsest $\mathcal{F}$-stratification. However, for the purposes of developing an algorithm, we will prove this constructively by defining each X_i in a coarsest $\mathcal{F}$-stratification. Let d_0 be the dimension of X and define $X_{d_0} := X$. Define

$$S_{d_0} := \{x \in X_{d_0} : \mathcal{F}(B_w \subset B_y) \text{ is an isomorphism for all chains } x \leq y \leq w\}$$

Set d_1 to be the dimension of $X_{d_0} - S_{d_0}$. Then define X_{d_1} to be the compliment of S_{d_0} in X_{d_0}:

$$X_{d_1} := X_{d_0} - S_{d_0}$$

Now each $d_0 + 1$ chain in X_{d_0} terminates with an element x of S_{d_0} because $\mathcal{F}|_{B_x}$ is automatically constant when x is the terminal element of a maximal chain. The dimension of X_{d_1} is strictly less than d_0, since each $d_0 + 1$ chain in X ends with an element of S_{d_0}, and thus is not a chain in $X - S_{d_0}$. Define $X_i := X$ for each i such that $d_1 < i < d_0$. Now X_{d_1} is itself a finite T_0-space. Let $B_x^{d_1}$ denote the minimal open neighborhood of x in X_{d_1}. Then we can use the same condition as above to define S_{d_1}:

$$S_{d_1} := \{x \in X_{d_1} : \mathcal{F}(B_w \subset B_y) \text{ is an iso. for all chains } x \leq y \leq w \text{ in } X_{d_1}\}$$

Again notice that S_{d_1} is not empty since terminal elements of maximal chains are guaranteed to be elements of S_{d_1}. Continue to define d_i to be the dimension of $X_{d_{i-1}} - S_{d_{i-1}}$ and $X_{d_i} := X_{d_{i-1}} - S_{d_{i-1}}$ inductively until $d_i = 0$. To fill out the missed indices, define S_j to be empty if $d_i < j < d_{i-1}$ and $X_j := X_{d_i}$ if $d_i \leq j < d_{i-1}$.

Notice that each S_i is an open subset of X_i. Therefore X_{i-1} is closed in X_i, and S_i is an open set in X_i (and therefore locally closed in X). So we have constructed a stratification of X.

Now we will focus on showing that $\mathcal{F}|_{S_i}$ is locally constant. If S_i is non-empty, then $S_i = S_{d_k}$ for some k. If we want to show that $\mathcal{F}|_{S_i}$ is locally constant, we need to check that $(\mathcal{F}|_{S_i})|_{B_x^i}$ is locally constant for each $x \in S_i$ (where $B_x^i = B_x \cap X_i$). Consider the presheaf $\mathcal{E}$ on B_x^i, which maps each open set $U \subset B_x^i$ to $\mathcal{F}(B_x)$, and each morphism $U \subset V$ to the identity morphism. So we have $\mathcal{E}(U) = \mathcal{F}(B_x)$ for all $U \subset B_x^i$, and $\mathcal{E}(U \subset V) = \mathrm{id} : \mathcal{F}(B_x) \to \mathcal{F}(B_x)$ for all $U \subset V \subset B_x^i$. Notice that the sheafification of $\mathcal{E}$ is by definition a constant sheaf. Let $\mathcal{E}'$ be the presheaf on B_x^i defined by $\mathcal{E}'(U) = \mathcal{F}(\mathrm{St}(U))$ and $\mathcal{E}'(U \subset V) = \mathcal{F}(\mathrm{St}(U) \subset \mathrm{St}(V))$. Notice that the sheafification of $\mathcal{E}'$ is by definition $(\mathcal{F}|_{S_i})|_{B_x^i}$. We want to show that the sheafification of $\mathcal{E}$ is isomorphic to the sheafification of $\mathcal{E}'$. Recall that it is enough to show that $\mathcal{E}$ and $\mathcal{E}'$ agree on minimal open sets B_y^i, and give the same restriction maps between minimal open sets. We have the equalities (as morphisms) $\mathcal{E}'(B_y^i \subset B_w^i) = \mathcal{F}(B_y \subset B_w) = \mathcal{F}(B_x \subset B_x) = \mathcal{E}(B_y^i \subset B_w^i)$, which we obtain by applying our definition of $\mathcal{E}'$, the assumption (made in our definition of S_i) that $\mathcal{F}(B_y \subset B_w)$ is an isomorphism for all $x \leq y \leq w \in X_i$, and the definition of $\mathcal{E}$. These equalities further imply that $\mathcal{E}'(B_y^i) = \mathcal{E}(B_y^i)$. So we have shown that the sheafification of $\mathcal{E}$ is isomorphic to the sheafification of $\mathcal{E}'$, which is a constant sheaf. Therefore $(\mathcal{F}|_{S_i})|_{B_x}$ is constant, which implies that $\mathcal{F}|_{S_i}$ is locally constant, which implies that $\mathcal{F}$ is constructible with respect to the decomposition $X = \coprod S_i$. So we have constructed an $\mathcal{F}$-stratification.

Now suppose that there exists a coarser $\mathcal{F}$-stratification

$$\emptyset \subset X'_0 \subset \cdots \subset X'_n = X$$

We will continue by using the notation S_i° (respectively $S_j'^\circ$) to denote a connected component of S_i (respectively S'_j). Suppose $S_i^\circ \subsetneq S_j'^\circ$. Let $x \in S_j'^\circ - S_i^\circ$. Since $\mathcal{F}$ is locally constant when restricted to $S_j'^\circ$, we have that $\mathcal{F}$ is constant when restricted to $B_x \cap S_j'^\circ$. Notice that $B_x \cap S_i^\circ \subset B_x \cap S_j'^\circ$. Therefore $\mathcal{F}$ is constant when restricted to $B_x \cap S_i^\circ$. Since S_i° is an open subset of X_i, we have that

$B_x \cap X_i = B_x \cap S_i^\circ$. So we can finally conclude that $\mathcal{F}$ is constant when restricted to $B_x \cap X_i$. However, by the definition of S_i above, we see that x must be an element of S_i. Therefore $S_i^\circ \subset S_j'^\circ \subset S_i$, which implies that $S_i^\circ = S_j'^\circ$. Therefore each stratum piece of the stratification $\emptyset \subset X_0 \subset \cdots \subset X_n = X$ is equal to a stratum piece of the stratification $\emptyset \subset X_0' \subset \cdots \subset X_n' = X$. So we can conclude that these two stratifications are equivalent, which concludes the proof. □

7.3 Proof of Theorem 3.3

Proof. We will prove this constructively by defining each X_i in a minimal homogeneous $\mathcal{F}$-stratification, and then showing that any minimal homogeneous $\mathcal{F}$-stratification is necessarily equal to the stratification constructed below. In many ways, this proof is similar to the proof of Theorem 3.2. Let d_0 be the dimension of K and $X_{d_0} = X$. Define

$$H_{d_0} := \{x \in X_{d_0} : \mathrm{Cl}(B_x) \text{ is homogeneous of dimension } d_0\}$$

(H for homogeneous) and

$$C_{d_0} := \{x \in H_{d_0} : \mathcal{F}(B_w \subset B_y) \text{ is an iso. for all } x \leq y \leq w\}$$

(C for constant) where $\mathrm{Cl}(B_x) = \{y \in X_{d_0} : y \leq s \text{ for some } s \in B_x\}$ is the closure of B_x. Then define $S_{d_0} = H_{d_0} \cap C_{d_0}$. Set d_1 to be the dimension of $X_{d_0} - S_{d_0}$. Then define X_{d_1} to be $X_{d_0} - S_{d_0}$. Now each $d_0 + 1$ chain in X_{d_0} terminates with an element x of S_{d_0} because $\mathrm{Cl}(x)$ is homogeneous of dimension d_0 by our assumption that X_{d_0} consists of simplices of a simplicial complex. We have that d_1 is strictly less than d_0, since each $d_0 + 1$ chain in X_{d_0} ends with an element of S_{d_0}, and thus is not a chain in X_{d_1}. Define $X_i := X_{d_1}$ for each i such that $d_1 < i < d_0$. Now X_{d_1} is itself a finite T_0-space. Let $B_x^{d_1}$ denote the minimal open neighborhood of x in X_{d_1}. Then we can use the same condition as above to define

$$H_{d_1} := \{x \in X_{d_1} : \mathrm{Cl}(B_x^{d_1}) \text{ is homogeneous of dimension } d_1\}$$

and

$$C_{d_1} := \{x \in H_{d_1} : \mathcal{F}(B_w \subset B_y) \text{ is an iso. for all chains } x \leq y \leq w \text{ in } H_{d_1}\}$$

As before, let $S_{d_1} = H_{d_1} \cap C_{d_1}$, and notice that S_{d_1} is not empty since X_{d_1} corresponds to a sub-simplicial complex of K. Continue to define H_{d_k}, C_{d_k}, S_{d_k}, and $X_{d_{k+1}}$ inductively until $d_k = 0$. To fill out the missed indices, define S_i to be empty if $d_j < i < d_{j-1}$ and $X_i := X_{d_j}$ if $d_j < i < d_{j-1}$.

Notice that each S_i is an open subset of X_i. Therefore X_{i-1} is closed in X_i, and $X_i - X_{i-1}$ is an open set in X_i (and therefore locally closed in X). Additionally, $\mathrm{Cl}(X_i - X_{i-1}) = \mathrm{Cl}(S_i)$ is either empty or is homogeneous of dimension i. So we have constructed a homogeneous stratification of X. Now we wish to show that this is a homogeneous $\mathcal{F}$-stratification. It remains to show that $\mathcal{F}|_{S_i}$ is locally constant. This follows the same argument as in the proof of Theorem 3.2. So $\mathcal{F}$ is constructible with respect to the stratification given by the filtration $0 \subset X_0 \subset \cdots \subset X_{d_0} = X$. We will denote this stratification by $\mathfrak{X}$.

Suppose that there exists a minimal homogeneous $\mathcal{F}$-stratification

$$\emptyset \subset X_0' \subset \cdots \subset X_{d_0}' = X$$

denoted by $\mathfrak{X}'$. Now S_i' must contain all of the elements of X_i' which correspond to i-simplices in K. Moreover, for each element $x \in S_i'$, there exists $y \in X_i'$ corresponding to an i-simplex in K, such that $x \leq y$ (due to the homgeneity of $X_i' - X_{i-1}'$). Suppose $a \in S_n$ and $b \in S_n'$ such that $a \leq b$ (an analogous argument follows for $b \leq a$). Since b is necessarily a face of an n-simplex

$\tau \in X'_n = X_n = X$, we have $a \leq b \leq \tau$. Since τ is a an n-simplex, we have that $\tau \in S_n$. Since $\mathcal{F}$ is assumed to be locally constant when restricted to S_i and S'_i, we have that ρ_{B_a,B_τ} and ρ_{B_b,B_τ} are isomorphisms. By the sheaf axioms, we have that $\rho_{B_b,B_\tau} \circ \rho_{B_a,B_b} = \rho_{B_a,B_\tau}$. Therefore, $\mathcal{F}(B_b \subset B_a)$ is an isomorphism. So if we set $S''_n := S_n \cup S'_n$, then $\mathrm{Cl}(S''_n)$ is homogeneous of dimension n and $\mathcal{F}|_{S''_n}$ is locally constant. However, by our construction of S_n, we can see that S_n is the maximal set with these properties. So $S_n \subset S''_n$ implies that $S_n = S''_n$. This implies that $S'_n \subset S_n = X - X_{n-1}$. If $S'_n \subsetneq S_n$, then we would have that $\mathfrak{X} < \mathfrak{X}'$, which would contradict the minimality of $\mathfrak{X}'$. So we must have that $S_n = S'_n$, which implies that $X'_{n-1} = X_{n-1}$. This allows us to inductive use the same argument above to show that $X'_i = X_i$ for all i. Therefore the two stratifications are equal, which concludes the proof. $\square$

8 Discussion

We would like to highlight two key features of our sheaf-theoretic stratification learning algorithm. The first feature is that we avoid computations which require the sheafification process. At first glance this is surprising, since constructible sheaves can not be defined without referencing sheafification, and our algorithm builds a stratification for which a given sheaf is constructible. In other words, each time we want to determine the restriction of a sheaf to a subspace, we need to compute the sheafification of the presheaf referenced in the definition of the pull back of a sheaf (Section 2.2). We can avoid this by noticing two facts. First, in the setting of finite T_0-spaces, we can deduce if a given sheaf $\mathcal{E}^+$ is constant by considering how it behaves on minimal open neighborhoods. Second, the behavior of $\mathcal{E}^+$ will agree with the behavior of the presheaf $\mathcal{E}$ on minimal open neighborhoods. Symbolically, this is represented by the equalities $\mathcal{E}^+(B_x) = \mathcal{E}(B_x)$ and $\mathcal{E}^+(B_w \subset B_x) = \mathcal{E}(B_w \subset B_x)$ for all pairs of minimal open neighborhoods $B_w \subset B_x$ (where B_x is a minimal open neighborhood of x, and B_w is a minimal open neighborhood of an element $w \in B_x$). Therefore, we can determine if $\mathcal{E}^+$ is constant, locally constant, or constructible, while only using computations involving the presheaf $\mathcal{E}$ applied to minimal open neighborhoods.

The second feature of our algorithm (which is made possible by the first) is that the only sheaf-theoretic computation required is checking if $\mathcal{F}(B_w \subset B_x)$ is an isomorphism for each pair $B_w \subset B_x$ in our space. This is extremely relevant for implementations of the algorithm, as it minimizes the number of expensive computations required to find a coarsest $\mathcal{F}$-stratification. For example, if our sheaf is the local homology sheaf, we will only need to compute the restriction maps between local homology groups of minimal open neighborhoods. This allows us to distribute the computation in determining local homology. Additionally, once we have determined whether the local homology restriction maps are isomorphisms, we can quickly compute a coarsest $\mathcal{F}$-stratification, or a minimal homogeneous $\mathcal{F}$-stratification, without requiring any local homology groups to be recomputed.

There are several interesting questions related to $\mathcal{F}$-stratifications that we would like to investigate in the future. We hope to study $\mathcal{F}$-stratifications for natural sheaves (other than the local homology sheaf) on finite T_0-spaces. The primary objective would be to find easily computable sheaves that yield intuitive stratifications of interest to data analysts. We would also like to study the stability of $\mathcal{F}$-stratifications under refinements of triangulations of polyhedra. In this direction, it would be interesting to view $\mathcal{F}$-stratifications from the perspective of persistent homology. If we are given a point cloud sampled from a compact polyhedron, it would be natural to ask about the convergence of $\mathcal{F}$-stratifications and the properties of the strata under a filtration of the simplicial complex. Finally, we are also intrigued by the results of [6], and possible implementations of our algorithm using spectral sequences.

Acknowledgements

This work was partially supported by NSF IIS-1513616.

References

[1] P. S. Alexandroff. Diskrete Räume. *Mathematiceskii Sbornik*, 2:501–518, 1937.

[2] Paul Bendich. *Analyzing Stratified Spaces Using Persistent Versions of Intersection and Local Homology*. PhD thesis, Duke University, 2008.

[3] Paul Bendich, David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Dmitriy Morozov. Inferring local homology from sampled stratified spaces. *Proceedings 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 536–546, 2007.

[4] Paul Bendich and John Harer. Persistent intersection homology. *Foundations of Computational Mathematics*, 11:305–336, 2011.

[5] Paul Bendich, Bei Wang, and Sayan Mukherjee. Local homology transfer and stratification learning. *Proceedings 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), arXiv:1008.3572*, pages 1355–1370, 2012.

[6] Nicolás Cianci and Miguel Ottina. A new spectral sequence for homology of posets. *Topology and its Applications*, 217(Supplement C):1 – 19, 2017.

[7] Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2010.

[8] M Goresky and R MacPherson. Intersection homology I. *Topology*, 19:135–162, 1982.

[9] M Goresky and R MacPherson. Intersection homology II. *Inventiones Mathematicae*, 71:77–129, 1983.

[10] Mark Goresky and Robert MacPherson. *Stratified Morse Theory*. Springer-Verlag, 1988.

[11] Gloria Haro, Gregory Randall, and Guillermo Sapiro. Stratification learning: Detecting mixed density and dimensionality in high dimensional point clouds. *Advances in NIPS*, 17, 2005.

[12] Frances Clare Kirwan. *An introduction to intersection homology theory*. Chapman & Hall/CRC, 2006.

[13] Gilad Lerman and Teng Zhang. Probabilistic recovery of multiple subspaces in point clouds by geometric lp minimization. *Annals of Statistics*, 39(5):2686–2715, 2010.

[14] M. McCord. Singular homology groups and homotopy groups of finite topological spaces. *Duke Mathematical Journal*, 33:465–474, 1978.

[15] James R. Munkres. *Elements of algebraic topology*. Addison-Wesley, Redwood City, CA, USA, 1984.

[16] Vidit Nanda. Local cohomology and stratification. ArXiv: 1707.00354, 2017.

[17] Colin Rourke and Brian Sanderson. Homology stratifications and intersection homology. *Geometry and Topology Monographs*, 2:455–472, 1999.

[18] Primoz Skraba and Bei Wang. Approximating local homology from samples. *Proceedings 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), arXiv:1206.0834*, pages 174–192, 2014.

[19] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (GPCA). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1945 – 1959, 2005.

[20] Shmuel Weinberger. *The topological classification of stratified spaces.* University of Chicago Press, Chicago, IL, 1994.