Sheaf-Theoretic Stratification Learning from Geometric and Topological Perspectives

Adam Brown[1,2] · Bei Wang[3]

Abstract

We investigate a sheaf-theoretic interpretation of stratification learning from geometric and topological perspectives. 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 Bendich et al. (Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1355–1370, ACM, New York, 2012), and the cohomology stratification algorithm given in Nanda (Found. Comput. Math. **20**(2), 195–222, 2020). Additionally, we give examples of stratifications based on the geometric techniques of Breiding et al. (Rev. Mat. Complut. **31**(3), 545–593, 2018), illustrating how the sheaf-theoretic approach can be used to study stratifications from both topological and geometric perspectives. 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.

Keywords Topological data analysis · Sheaf · Stratification learning

Mathematics Subject Classification 55-08 · 51-08 · 32S60

Adam Brown
abrown@math.utah.edu; abrown@ist.ac.at

Bei Wang
beiwang@sci.utah.edu

1 Department of Mathematics, University of Utah, 155 South 1400 East, Salt Lake City, UT 84112, USA

2 Institute of Science and Technology Austria, Am Campus 1, Klosterneuburg 3400, Austria

3 School of Computing and SCI Institute, University of Utah, 72 South Central Campus Drive, Salt Lake City, UT 84112, USA

Springer

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*. Here, we use the word "stratification learning" loosely to mean an unsupervised, exploratory, clustering process that infers a decomposition of data into disjoint subsets that capture recognizable and meaningful structural information.

Previous work in mathematics has focused on the study of stratified spaces under smooth and continuous settings [14,30] 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 [16,19,29], and may not handle general scenarios with complex singularities. Recently, approaches from topological data analysis [3,5,28], which rely heavily on ingredients from computational [11] and intersection homology [2,4,12], 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 based on how the local homology of nearby sampled points map into one another [5]. Philosophically, our main goal is to find a stratification where any two points in the same strata (or cluster) can not be distinguished by homological methods, and any two points in different strata (different clusters) can be distinguished by homological methods. The majority of the paper will be spent developing a rigorous and computable interpretation of the purposely vague statement "distinguished by homological methods". Furthermore, we will see that our approach to computing the above stratification applies equally well to sheaves other than those based on local homology. As examples, we describe stratification learning with the combinatorially defined sheaf of maximal elements, and the geometrically motivated pre-sheaf of vanishing polynomials. This paper includes the full version of an extended abstract [7], and further extends these results by exploring alternatives to homological stratifications which lie in the sheaf-theoretic framework (Sects. 6 and 7). As our work is an interplay between sheaf theory and stratification, we briefly review various notions of stratification before describing our results.
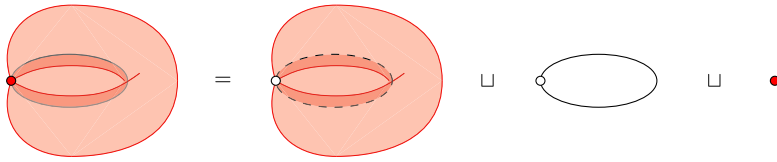
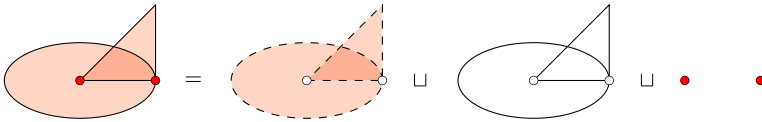**Fig. 1** Example of a topological stratification of a *pinched torus*

Fig. 2 Example of a homological stratification of a *sundial*

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 Fig. 1 for an example of a *pinched torus*, that is, a torus with points along a geodesic with fixed longitude identified, and a spanning disc glued along the equator.

Ideally, we would compute a topological stratification for a given space. However, if we are restricted to using only 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. A well-known example of a homological manifold that is not a topological manifold can be constructed from a homology 3-sphere with nontrivial fundamental group. The suspension of such a space is a homological manifold, but not a topological manifold, since the links of the suspension points have nontrivial fundamental groups [21, p. 326]. Specifically, the suspension of the Poincaré's homology sphere is a homological manifold but not a topological manifold. In this paper, we will avoid the difficult problem of computing topological stratifications, and instead aim to investigate stratifications which can be computed using homological methods. Therefore, we must first consider a definition of stratification that does not rely on topological conditions which are not distinguished by homology. We begin with an extremely loose definition of stratification (Definition 1.1) which only requires the properties necessary to discuss the constructibility of sheaves (defined in Sect. 2.2). We will then refine our definition of stratification by placing requirements on the constructibility of certain sheaves (Definition 1.4).

Definition 1.1 Given a topological space X, a *stratification* $\mathfrak{X}$ of X is a finite filtration of X by closed subsets X_i:

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

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 $\mathfrak{X}'$, or $\mathfrak{X}'$ *refines* $\mathfrak{X}$, if each stratum piece of $\mathfrak{X}'$ is contained in a stratum piece of $\mathfrak{X}$.

Figure 3 illustrates some examples of stratifications which are coarser than those in Figs. 1 and 2, as well as a different partition based on local homology transfer algorithm in [5] for the sundial (bottom).

Homological stratification There have been several approaches in topology literature to define 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 [13, Sect. 4]. There have been several approaches for building on the ideas of Goresky and MacPherson, with applications to computational geometry and topology in mind [4,24]. In this paper, we choose to adopt the perspective of homological stratifications found in [25], with a view toward sheaf theoretic generalizations and applications in topological data analysis.

Consider a filtration $\emptyset = X_{-1} \subset X_0 \subset \cdots \subset X_n = X$ of a topological space X. We can use relative homology to define a sheaf (the local homology sheaf) which associates to each open set U in X_i the relative homology groups $H_\bullet(X_i, X_i - U)$, for some i. Let $\mathcal{L}_i$ denote the local homology sheaf on the space X_i (see Sect. 5.1 for the definition of the local homology sheaf).

Definition 1.3 We say that a stratification is a *homological stratification* if $\mathcal{L}_n$ is locally constant (see Sect. 2.2 for the definition of locally constant) when restricted to $X_i - X_{i-1}$, for each i. A stratification is a *strong homological stratification* if for each i both $\mathcal{L}_i$ and $\mathcal{L}_n$ are locally constant when restricted to $X_i - X_{i-1}$. Finally, a stratification is a *very strong homological stratification* if for each i and every $k \geq i$, $\mathcal{L}_k$ is locally constant when restricted to $X_i - X_{i-1}$.

As mentioned in [25], it would be interesting to study the relationship between these definitions of homological stratification. We plan to pursue this in a future work. The cohomological stratification given in [24] can be considered as a cohomological analogue of the very strong homological stratification defined above. The utility of this definition is the extent to which it lends itself to the study of topological properties of individual strata. For example, it can be easily shown that the strata of such a stratification are R-(co)homology manifolds (R being the ring with which the local cohomology is computed). The trade off for using the very strong homological stratification is in the number of local (co)homology groups which need to be computed. This is by far the most computationally expensive aspect of the algorithm, and the very strong homological stratification requires one to compute new homology groups for each sheaf $\mathcal{L}_i$. By contrast, the homological stratification only requires the computation of local homology groups corresponding to the sheaf $\mathcal{L}_n$. In this paper, we choose to study the less rigid (and more computable) notion of homological stratification (see Sect. 4 for more details on computing homological stratifications).

Sheaf theoretic stratification The definition of homological stratification naturally lends itself to generalizations, which we now introduce (while delaying formal defini-
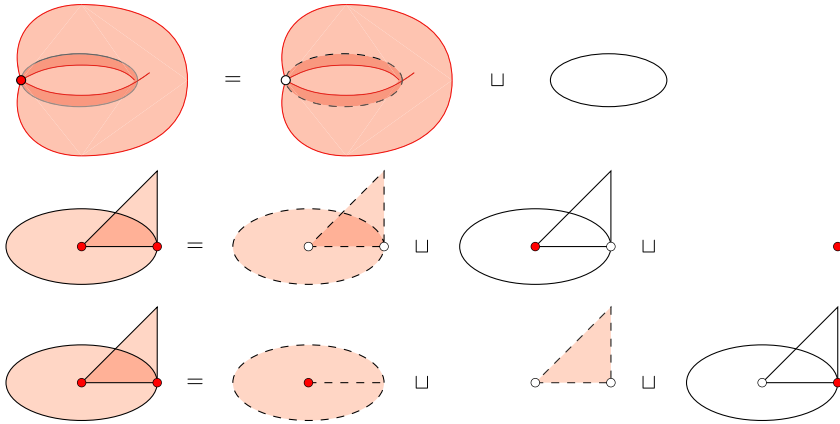
Fig. 3 Top: A topological stratification of a *pinched torus* which is coarser than the stratification shown in Fig. 1. Readers familiar with stratification theory will notice that while this is a topological stratification, it is not a locally conelike topological stratified space, also called a cs-space or cs-stratification [15, Def. 1.1]. Middle: An example of a topological stratification of a *sundial* which is coarser than the stratification shown in Fig. 2. One can check that the coarser stratification is no longer a homological stratification. Bottom: A partition of a *sundial* into stratum pieces, based on local homology transfer from [5], where we assume an arbitrarily dense sampling of points from the sundial. Notice that the resulting decomposition is not a stratification by our definition, since it is not induced by a filtration by closed subspaces

tion of constructible sheaves to Sect. 2.2). The primary observation which illuminates this generalization is that the fundamental mathematical structures (the association of a group or set to an open set and the restriction maps induced by inclusions of open sets) used to construct homological stratifications in [13,24,25] are exactly the formal properties which define a pre-sheaf. Therefore, we will explore analogous constructive stratifications where local homology is replaced with an arbitrary sheaf or pre-sheaf. The key notion used in the following definition is that of a *constructible* sheaf. Intuitively, constructibility means that the sheaf can be decomposed into sheaves which are locally constant on various pieces, or stratum, of the given topological space.

Definition 1.4 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 exist, 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. The extra structure needed to prove uniqueness is homogeneity of strata, and minimality of stratifications. By working with homogeneous stratifications, we are requiring strata to have a good notion of dimension. By defining a preorder on the set of homogeneous stratifications (Sect. 3), we are able to identify the output of the stratification algorithm described in Sect. 4 as the unique minimal homogeneous stratification with respect to the preorder.

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 fourfold:

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 (Sect. 3).
2. We give an algorithm for computing each of the above stratifications of a finite T_0-space based on a sheaf-theoretic language (Sect. 4).
3. In particular, when applying the local homology sheaf in our algorithm, we obtain a coarsest homological stratification (Sect. 5.2).
4. We give detailed examples of sheaf-theoretic stratifications based on combinatorial techniques (Sect. 6) and geometric techniques (Sect. 7).

We envision that our abstraction could give rise to a larger class of stratifications beyond homological stratification. For instance, we give examples of a "maximal element-stratification" when the sheaf is defined by considering maximal elements of an open set (Sect. 6) and a "vanishing polynomial" stratification when the (pre)-sheaf is defined with sets of vanishing polynomials (Sect. 7). Moreover, we see the geometric stratifications based on vanishing sets of polynomials as having natural applications to the mapper algorithm of [8,22,27] (see Sect. 7).

Comparison to prior work This paper can be viewed as a continuation of previous works which adapt the stratification and homology theory of Goresky and MacPherson to the realm of topological data analysis. In [25], Rourke and Sanderson give a proof of the topological invariance of intersection homology on PL homology stratifications, and give a recursive process for identifying a homological stratification (defined in [25, Sect. 5]). In [4], Bendich and Harer introduce a persistent version of intersection homology that can be applied to simplicial complexes. In [5], Bendich et al. 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 [24], Nanda uses the machinery of derived categories to study cohomological stratifications based on local cohomology.

Motivated by the results of [5,24], we aim to develop a computational approach to the stratifications studied in [25]. Our main results can be summarized as the generalization of homological stratifications of [25] to $\mathcal{F}$-stratifications, and a proof of existence and uniqueness of the minimal homogeneous $\mathcal{F}$-stratification of a finite simplicial complex. When $\mathcal{F}$ is the local homology sheaf, we recover the homological stratification described by [25]. While admitting a similar flavor as [24], our work differs from [24] in several important ways. The most obvious difference is our choice to work with homology and sheaves rather than cohomology and cosheaves. More importantly however, we reduce the number of local homology groups which need to be computed. We will investigate the differences between homological, strong homological, and very strong homological stratifications in a future work.

In [5], stratifications of point clouds are defined using persistent local homology and an equivalence based on local homology transfer. This method can easily be adapted to

the setting of simplicial complexes by defining two simplices τ and σ to be in the same stratum if there exists a sequence of face/coface relations $\tau \leq \gamma_1 \geq \cdots \leq \gamma_n \geq \sigma$, such that each face/coface relation induces an isomorphism of local homology groups. We choose not to take this approach here because the local homology transfer algorithm fails to produce reasonable stratifications for certain topological spaces. For example, the stratification of the *sundial* example (i.e., a stratified space with boundary) given by local homology transfer (see Fig. 3) is not technically a stratification by our definition, since the 1-dimensional stratum (which in this case is equal to X_1 in the induced filtration) is not closed. In comparison, the current algorithm correctly gives a coarsest homological stratification of this space. The algorithms described in this paper differ from local homology transfer in that we inductively define strata by requiring all restriction maps (transfer maps in a small neighborhood of a point/simplex) of a point/simplex in the said stratum to induce isomorphisms of local homology groups. In this sense, our algorithm assigns a point/simplex to the top stratum if the local homology is unchanged as we move small amounts in "any direction" of the point, while the homology transfer algorithm assigns two points to the same stratum if there exists at least one path connecting them which induces an isomorphism of local homology. Furthermore, the work in [5] uses persistent homology in an essential way so that it is amenable to point cloud data. The current work only brushes with the concept of persistence in Sect. 9. We plan to build on the results of [5,28], and extend the sheaf-theoretic stratification learning perspective described in this paper to the study of stratifications of point cloud data using persistent local homology.

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 given by the face/coface relation ($\tau \leq \sigma$ if simplex τ is a face of simplex σ). 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. The correspondence between finite T_0-spaces and simplicial complexes is detailed in [20]:

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 is 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 a topological space) 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.

As a concrete example, let K denote the finite T_0-space consisting of elements which are open simplices in a simplicial complex. In this setting, we can describe the T_0-space K using the more familiar language of simplicial complexes. This is also the setting that applies to most of the subsequent examples in this paper. For a simplex $\sigma \in K$, its minimal open neighborhood B_σ is its *star* consisting of all cofaces of σ, $\mathrm{St}\,\sigma = \{\tau \in K : \sigma \leq \tau\}$. Using the partial order induced by inclusions of minimal open neighborhoods, we set $\tau \leq \sigma$ if $B_\sigma \subseteq B_\tau$. By describing minimal open neighborhoods as open stars of simplices, we can state the partial order as $\tau \leq \sigma$ if $\mathrm{St}\,\sigma \subseteq \mathrm{St}\,\tau$. On the other hand, K is equipped with a partial order based on face relations, where $\tau \leq \sigma$ if simplex τ is a face of simplex σ. Therefore, the two partial orders (the face partial order and the open neighborhood inclusion partial order) coincide.

Given a finite T_0-space X with the above partial order, we say $x_0 \leq x_1 \leq \cdots \leq x_n$ (where $x_i \in X$) is a maximal chain in X if there is no totally ordered subset $Y \subset X$
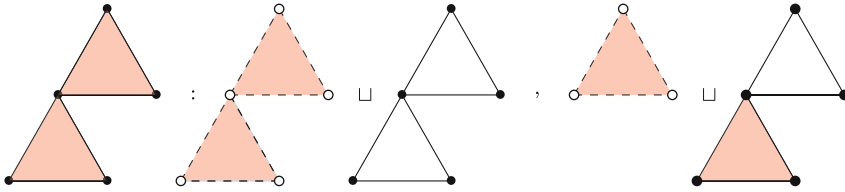
Fig. 4 An example of a homogeneous simplicial complex (left), a homogeneous stratification (middle), and a stratification which is not homogeneous (right)

consisting of elements $y_j \in Y$ such that $y_0 \leq \cdots \leq y_j \leq \cdots \leq y_k$ and $\bigcup_{i=0}^{n}\{x_i\} \subsetneq Y$. The cardinality of a chain $x_0 \leq x_1 \leq \cdots \leq x_n$ is $n + 1$. We say that a finite T_0-space has dimension m if the maximal cardinality of maximal chains is $m + 1$.

Definition 2.1 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 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 $m + 1$.

See Fig. 4 for an example. 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 data together 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. In this paper, we are primarily interested in sheaves on finite T_0-spaces, which are closely related to the cellular sheaves studied in [10,24,26].

Sheaves Suppose X is a topological space. Let **Top**(X) denote the category consisting of objects which are open sets in X with morphisms given by inclusion. Let $\mathcal{F}$ be a contravariant functor from **Top**(X) to $\mathcal{S}$, the category of sets. For open sets $U \subset V$ in X, we refer to the morphism $\mathcal{F}(U \subset V)\colon \mathcal{F}(V) \to \mathcal{F}(U)$ 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 *sheaf*[1] on X if $\mathcal{F}$ satisfies the following conditions 1.–3.; a *pre-sheaf* is a functor $\mathcal{E}$ (as above) which satisfies conditions 1.–2.:

---

[1] See [10, Chap. 2] and [18, Chap. 3] for various introductions to sheaf theory. In [10], Curry gives a very general definition of pre-sheaves which take values in a general category (possibly differing from the category of sets). Curry then discusses how certain properties of the chosen category (such as the existence of direct and inverse limits) imply properties for the corresponding pre-sheaves and sheaves. [18] gives a standard introduction to sheaves, constructible sheaves, and intersection homology.

1. $\mathcal{F}(U \subset U) = \mathrm{id}_U$;
2. if $U \subset V \subset W$, then $\mathcal{F}(U \subset W) = \mathcal{F}(U \subset V) \circ \mathcal{F}(V \subset W)$;
3. 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 pre-sheaf into a sheaf. In the setting of finite T_0-spaces, sheafification takes on a relatively simple form. Let $\mathcal{E}$ be a pre-sheaf on a finite T_0-space X. Then the sheafification of $\mathcal{E}$, denoted $\mathcal{E}^+$, is given by

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

For any pre-sheaf $\mathcal{E}$, 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 pre-sheaves will sheafify to the same sheaf if they agree on all minimal open neighborhoods. We will use this fact several times in Sect. 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 St $Y := \bigcup_{y \in Y} B_y$, 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 pre-sheaf $\mathcal{E}$ which maps an open set $U \subset X$ to $\mathcal{E}(U) := \mathcal{F}(\mathrm{St}\, f(U))$. We can avoid using direct limits in our definition of pull back because each point in a finite T_0-space has a minimal open neighborhood [10, Chap. 5]. The pull back of $\mathcal{F}$ along an inclusion map $\iota : U \hookrightarrow X$ 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 [18]. 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 2.2 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 (Definition 1.4, Proposition 3.1), the existence of coarsest $\mathcal{F}$-stratifications (Theorem 3.2), and the existence and uniqueness of minimal homogeneous $\mathcal{F}$-stratifications (Theorem 3.6). Of course, Theorem 3.2 immediately implies Proposition 3.1. We choose to include a separate statement of Proposition 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 Sect. 8 for technical details.

Proposition 3.1 *Let $\mathcal{F}$ be a sheaf on a finite T_0-space X. There exists an $\mathcal{F}$-stratification of X (see Definitions 1.4 and 2.2).*

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. Our process is a greedy algorithm as it seeks to make locally optimal choice at each stage. Rephrased using the language of simplicial complexes, the top-dimensional stratum will consist of all simplices such that the sheaf is constant when restricted to the open star of the simplex. We then remove the top-dimensional stratum from our space, and pull back the sheaf to the remaining points/simplices. Notice that the points/simplices which are maximal with respect to the natural partial order (described in Sect. 2.1) are guaranteed to be included in the top-dimensional stratum. We proceed inductively 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 Sect. 8.2 for the details of the above argument. □

Notice that if $\mathcal{F}$ is the trivial sheaf on X, then the coarsest $\mathcal{F}$-stratification of X will be the trivial stratification, consisting of a single stratum, even if the space X has mixed dimensionality. For example, the coarsest $\mathcal{F}$-stratification of the space X, as illustrated in Fig. 5, is the trivial stratification of X. To remedy this situation, we will introduce an extra constraint of homogeneity on a stratification, which will require each stratum piece to have "pure" dimension.
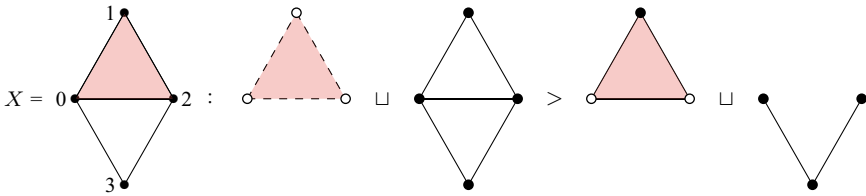
Fig. 5 An example of two inequivalent coarsest homogeneous $\mathcal{C}$-stratifications, where $\mathcal{C}$ is a constant sheaf on X. The stratification given on the right is the minimal homogeneous $\mathcal{C}$-stratification

Definition 3.3 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 for each i, the closure of the stratum S_i in X_i is homogeneous of dimension i (defined in Sect. 2.1).

Observe that if $\mathcal{F}$ is the constant sheaf on a simplicial complex X, as in Fig. 5, then the skeletal filtration of X (defined by setting X_i to be the set of i-simplices of the simplicial complex X) is automatically a homogeneous $\mathcal{F}$-stratification. However, the skeletal stratification is not guaranteed to be a coarsest $\mathcal{F}$-stratification.

We will introduce a lexicographical preorder on the set of homogeneous stratifications of a finite T_0-space X. While this preorder can be extended to the set of all stratifications, we choose to define the preorder on homogeneous stratifications in order to take advantage of the required regularity of the indexing of stratum. Specifically, the fact that the stratum $X_i - X_{i-1}$ is homogeneous of dimension i will guarantee compatible indexing of each homogeneous stratification of X. Let $\mathfrak{X}$ be a homogeneous stratification of X given by

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

We define a sequence $A^{\mathfrak{X}} := \{|X_n|, \ldots, |X_i|, \ldots, |X_0|\}$, where $|X_i|$ denotes the cardinality of the set X_i. Notice that the order of the sequence $A^{\mathfrak{X}}$ is the reverse of the natural filtration order corresponding to $\mathfrak{X}$. Given two homogeneous stratifications $\mathfrak{X}$ and $\mathfrak{X}'$ of X, let $m \in \{-1, \ldots, n\}$ be the smallest integer such that $|X'_j| - |X_j| = 0$ for all $j > m$. We say that $\mathfrak{X} < \mathfrak{X}'$ if $|X'_m| - |X_m|$ is positive; $\mathfrak{X} \leq \mathfrak{X}'$ if $m = -1$. In other words, we say that $\mathfrak{X} < \mathfrak{X}'$ if the first nonzero term in the sequence $A^{\mathfrak{X}'} - A^{\mathfrak{X}} = \{|X'_n| - |X_n|, \ldots, |X'_i| - |X_i|, \ldots, |X'_0| - |X_0|\}$ is positive. For two stratifications $\mathfrak{X}$ and $\mathfrak{X}'$ of the same space X, the first term, $|X'_n| - |X_n|$, of the sequence $A^{\mathfrak{X}'} - A^{\mathfrak{X}}$, is automatically zero, since $X_n = X'_n = X$.

Figure 5 illustrates two homogeneous stratifications of the simplicial complex X. The stratification on the left produces an integer sequence $A^{\mathfrak{X}'} = \{10, 9, 0\}$, while the stratification on the right produces the sequence $A^{\mathfrak{X}} = \{10, 5, 0\}$. The sequence of differences is then $A^{\mathfrak{X}'} - A^{\mathfrak{X}} = \{0, 4, 0\}$. The first nonzero term in the sequence is positive, so $\mathfrak{X} < \mathfrak{X}'$.

Proposition 3.4 *If $\mathfrak{X}$ and $\mathfrak{X}'$ are homogeneous stratifications of X such that $\mathfrak{X}$ is coarser than $\mathfrak{X}'$, then $\mathfrak{X} \leq \mathfrak{X}'$. Moreover, if $\mathfrak{X}$ is coarser than $\mathfrak{X}'$, and $A^{\mathfrak{X}} = A^{\mathfrak{X}'}$, then $\mathfrak{X} = \mathfrak{X}'$.*

Proof By Definition 1.2, each stratum piece of $\mathfrak{X}'$ is contained in a stratum piece of $\mathfrak{X}$. Assume that a connected component of $X_i' - X_{i-1}'$ is contained in a connected component of $X_j - X_{j-1}$ for some j. Since both stratifications are assumed to be homogeneous, we have that each connected component of $X_i' - X_{i-1}'$ is homogeneous of dimension i and each connected component of $X_j' - X_{j-1}'$ is homogeneous of dimension j. By the assumed containment of stratum pieces, we have that $i \leq j$.

Assume that $X_j = X_j'$ for all $j \geq i$. Then $X_j - X_{j-1} = X_j' - X_{j-1}'$ for all $j > i$. Since $X_j - X_{j-1} = X_j' - X_{j-1}'$ for all $j > i$, and $(X_j' - X_{j-1}') \cap (X_i' - X_{i-1}') = \emptyset$ for all $j > i$, we must have that $(X_j - X_{j-1}) \cap (X_i' - X_{i-1}') = \emptyset$ for all $j > i$. By the preceding argument, each connected component of $X_i' - X_{i-1}'$ is contained in $X_j - X_{j-1}$ for some $j \geq i$. Therefore, $X_i' - X_{i-1}' \subseteq X_i - X_{i-1}$. Combining the equality $X_i' = X_i$ with the containment $X_i' - X_{i-1}' \subseteq X_i - X_{i-1}$, we have that $X_{i-1} \subseteq X_{i-1}'$.

If $X_{i-1} \subsetneq X_{i-1}'$, then $|X_{i-1}'| - |X_{i-1}| > 0$, and $\mathfrak{X} < \mathfrak{X}'$. Otherwise, $X_{i-1} = X_{i-1}'$, and the proof proceeds inductively until either $X_k \subsetneq X_k'$ for some k, or $X_k = X_k'$ for all k. If $X_k \subsetneq X_k'$ then $\mathfrak{X} < \mathfrak{X}'$. If $X_k = X_k'$ for all k, then $\mathfrak{X} = \mathfrak{X}'$. □

We will use the preorder on the set of homogeneous $\mathcal{F}$-stratifications to identify a unique stratification which is minimal in the preorder.

Definition 3.5 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 $\mathfrak{X}'$. In other words, a homogeneous $\mathcal{F}$-stratification is *minimal* if it is minimal with respect to the lexicographic order on homogeneous $\mathcal{F}$-stratifications.

There are several examples which illustrate the necessity of introducing *minimal* homogeneous $\mathcal{F}$-stratifications, rather than studying only coarsest homogeneous $\mathcal{F}$-stratifications. Consider the simplicial complex K illustrated in Fig. 5. Let $\mathcal{C}$ denote the constant sheaf on the corresponding T_0-space X by assigning the one-dimensional vector space k to the star of each simplex σ, $\mathcal{C}(\text{St }\sigma) = k$. For the restriction maps, $\mathcal{C}(\text{St }\tau \subset \text{St }\sigma)$ is an isomorphism for each pair $\sigma < \tau$. Figure 5 shows that there are two coarsest homogeneous $\mathcal{C}$-stratifications of X. The stratification on the right side of Fig. 5 is the minimal homogeneous $\mathcal{C}$-stratification.

Theorem 3.6 *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 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 lexicographic 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 Sect. 8.3 for the remaining details. □

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 (Sect. 5) and the sheaf of maximal elements (Sect. 6).

Let X be a finite T_0-space, equipped with the natural partial order defined in Sect. 2.1. Instead of using the sheaf-theoretic language of Theorem 3.6, 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 $\mathcal{F}(B_y \subset B_x) \colon \mathcal{F}(B_x) \to \mathcal{F}(B_y)$ 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 2. in the definition of a sheaf (Sect. 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 partial ordering, computing δ can be interpreted as assigning a binary label to the edges of a Hasse diagram associated with the partial ordering (see Sect. 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

   2a. 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 $B_x \cap X_{d_i}$.
   2b. Set $d_{i+1} = \dim(X_{d_i} - S_{d_i})$.
   2c. Define $X_{d_{i+1}} = X_{d_i} - S_{d_i}$.
   2d. 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. It is worthwhile to note that in step 2a, we only want to consider adjacent pairs $w \leq y$ in the minimal open neighborhood of x (i.e., the open star of x) which have not been previously assigned to a stratum. In other words, instead of checking

$\delta(w \leq y)$ for every adjacent pair in the minimal open neighborhood of x, we only want to require $\delta(w \leq y) = 1$ for each adjacent pair which is contained in both the minimal open neighborhood of x and in the d_i-level of the filtration. Moreover, each element of X which is maximal with respect to the partial order will be automatically included in current strata defined by the algorithm. Hence the cardinality of the set of unassigned elements will decrease after each iteration of the algorithm.

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 2a. is:

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

The only modification in the algorithm for computing the unique minimal homogeneous $\mathcal{F}$-stratification is the additional requirement that $c(x, i) = 1$ for x to be included in the stratum X_{d_i}. This algorithm constructs a stratification which is minimal in the lexicographic order on homogeneous $\mathcal{F}$-stratifications because the algorithm iteratively maximizes the cardinality of each stratum, starting with the top-dimensional stratum. In Sect. 8.3, we give a detailed proof that since the cardinality of the stratum is maximized at each iteration of the algorithm, the resulting stratification must be minimal in the lexicographic order.

5 Stratification Learning with the Local Homology Sheaf

5.1 Local Homology Sheaf

We begin with an example that uses homology to cluster simplices of a T_0-space into strata. In order to capture local structure, we study a version of relative homology (i.e., the local homology) involving minimal open neighborhoods. In order to define relative (and local) homology, we will start with a chain complex associated to a finite T_0-space.

For a finite T_0-space X, consider the chain complex $C_\bullet(X)$, where $C_p(X)$ denotes the free R-module generated by $(p+1)$-chains in X, with chain maps $\partial_p \colon C_p(X) \to C_{p-1}(X)$ 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. Define the homology of X (with coefficients in the ring R) to be the homology groups of $C_\bullet(X)$, $H_i(X) = \ker \partial_i / \operatorname{im} \partial_{i+1}$. Since X is a finite T_0-space, each subset U of X is a finite T_0-space. We therefore define the homology of X relative to U, $H_\bullet(X, U)$, to be the homology groups of the quotient of chain complexes $C_\bullet(X)/C_\bullet(U)$.

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) [23, p. 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*.[2]

The following theorem, though straightforward, provides justification for applying the results of Sect. 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.–2. are satisfied in the definition of sheaf from Sect. 2. The inclusion of open sets $U \subset V$, and equivalently $X - V \subset X - U$, induce a morphism of graded R-modules,

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

We have the following commutative diagram of chain complexes

$$
\begin{array}{ccccccccc}
0 & \longrightarrow & C_\bullet(X - V) & \longrightarrow & C_\bullet(X) & \xrightarrow{p_1} & C_\bullet(X)/C_\bullet(X - V) & \longrightarrow & 0 \\
& & \downarrow & & \downarrow{\scriptstyle\mathrm{id}} & & \downarrow & & \\
0 & \longrightarrow & C_\bullet(X - U) & \longrightarrow & C_\bullet(X) & \xrightarrow{p_2} & C_\bullet(X)/C_\bullet(X - U) & \longrightarrow & 0
\end{array}
$$

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 that condition 3. 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 pre-sheaf in order to apply our algorithm, we will not include the details of this part of the proof. $\qquad \square$

---

[2] See [9] for an interesting approach to the computation of homology groups of finite T_0-spaces using spectral sequences.

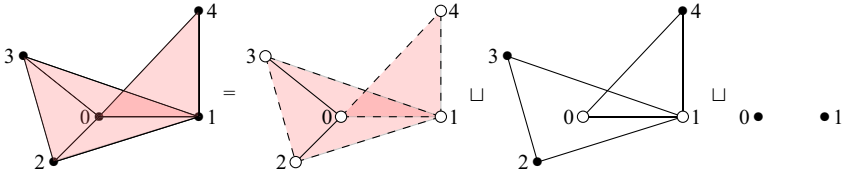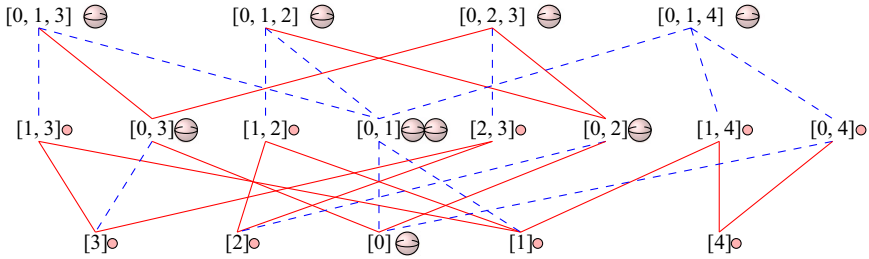**Fig. 6** A triangulated sundial and its stratification based on the local homology sheaf

Fig. 7 The Hasse diagram of the triangulated sundial. For any two adjacent simplices $\tau < \sigma$, an edge between τ and σ in the diagram is solid red if $\mathcal{L}(B_\tau) \to \mathcal{L}(B_\sigma)$ is an isomorphism; otherwise it is in dotted blue. On the right of each simplex τ is either a point, a sphere, or the wedge of two spheres, chosen so that $\mathcal{L}(B_\tau)$ is isomorphic to the reduced homology of the associated space

5.2 An Example Using the Local Homology Sheaf

If X is a T_0-space corresponding to a simplicial complex K, then the local homology groups in Sect. 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 Fig. 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 (Fig. 6). 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, St $\sigma = \{\tau \in K : \sigma \leq \tau\}$. The *closed star*, $\overline{\text{St}}\,\sigma$, is the smallest subcomplex that contains the star. The *link* consists of all simplices in the closed star that are disjoint from the star, Lk $\sigma = \{\tau \in \overline{\text{St}}\,\sigma : \tau \cap \text{St}\,\sigma = \emptyset\}$. 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 Fig. 7.

A sheaf on K can be considered as a labeling of each vertex in the Hasse diagram with a set and each edge with a morphism between the corresponding sets. Consider the local homology sheaf $\mathcal{L}$ on K which takes each open set[3] $U \subset K$ to $H_\bullet(K, K-U) \cong \tilde{H}_\bullet(K/(K-U)) \cong \tilde{H}_\bullet(\text{Cl}\,U/\text{Lk}\,U) \cong H_\bullet(\text{Cl}\,U, \text{Lk}\,U)$, where Lk $U := \text{Cl}\,U - U$. The local homology sheaf is naturally group-valued, but we make no use of the group structure here. So we will forget the group structure of local homology groups, and think of them purely as sets. The above isomorphisms follow from excision and the

---

[3] In the finite simplicial setting, U is the support of a union of open simplices in K.
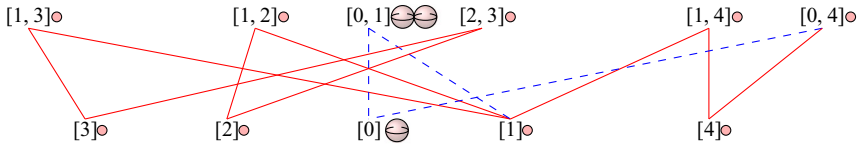
Fig. 8 The Hasse diagram after the top-dimensional stratum has been removed. We can consider this the beginning of the second iteration of the algorithm in Sect. 4

observation that $K-U$ (resp. Lk U) is a closed subcomplex of K (resp. Cl U), and therefore $(K, K-U)$ and $(Cl\, U, Lk\, U)$ form good pairs (see [17, p. 124]). Our algorithm described in Sect. 4 can then be interpreted as computing local homology sheaf associated with each vertex 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 $\mathcal{L}(\sigma \lessdot \tau)$ is an isomorphism for all pairs $\sigma < \tau$.

As illustrated in Fig. 7, first, we start with the 2-simplices. Automatically, we have that $\mathcal{L}$ is constant when restricted to any 2-simplex, and gives homology groups isomorphic to the reduced homology of a 2-sphere. For instance, the local homology group of the 2-simplex $\sigma = [0, 1, 3]$ is isomorphic to the reduced homology of a 2-sphere, $H_\bullet(\overline{St}\,\sigma, Lk\,\sigma) \cong \tilde{H}_\bullet(S^2)$.

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 Lk $B_{[1,3]} = [0] \cup [3] \cup [1] \cup [0, 3] \cup [0, 1]$, and $H_\bullet(Cl\, B_{[1,3]}, Lk\, B_{[1,3]})$ is isomorphic to the reduced homology of a single point space. Therefore the restriction map $\mathcal{L}(B_{[1,3]}) \rightarrow \mathcal{L}(B_{[0,1,3]})$ is not an isomorphism (illustrated as a dotted blue line in Fig. 7). 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 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 reduced 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 Fig. 7). This implies that $[0, 3] \in S_2 = X_2 - X_1$. Alternatively, we can consider the simplex $[0, 1]$ and see that $\mathcal{L}(B_{[0,1]}) \cong \tilde{H}_\bullet(\overline{St}\,[0, 1]/Lk\,[0, 1]) \cong \tilde{H}_\bullet(S^2 \vee S^2)$, the reduced homology of the wedge of two spheres. Therefore, for any 2-simplex τ, the restriction map $\mathcal{L}([0, 1] \lessdot \tau)$ cannot be an isomorphism. We conclude that $[0, 1]$ is not contained in the top-dimensional stratum. If we continue, we see that the top-dimensional stratum 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 Fig. 6.

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 Fig. 8). We get $S_1 = [0, 1] \cup [1, 3] \cup [1, 2] \cup [2, 3] \cup [0, 4] \cup [1, 4] \cup [2] \cup [3] \cup [4]$, which is visualized in Fig. 6. Finally, the stratum $S_0 = X_0$ consists of the vertices which have not been assigned to any strata. So $S_0 = [0] \cup [1]$.

We think it prudent to point out several observations related to the above example. First, we see that the local homology groups assigned to a simplex are trivial if and only if the simplex belongs to the boundary of our space. In this sense, local homology can detect which simplices are on the boundary of the space without relying on any particu-

lar geometric realization of the abstract simplicial complex (embedding of the abstract simplicial complex into Euclidean space). Secondly, we observe that (for this example) the coarsest $\mathcal{L}$-stratification we calculated is actually the unique minimal homogeneous $\mathcal{L}$-stratification. We will investigate this coincidence for $\mathcal{L}$-stratifications elsewhere, in an attempt to say if a coarsest $\mathcal{L}$-stratification is automatically homogeneous or minimal. For low-dimensional examples, we observe local homology based stratification we recover is actually a topological stratification. In general, local homology does not carry enough information to recover a stratification into manifold pieces, and examples exist in higher dimensions where $\mathcal{L}$-stratifications are not topological stratifications. One final remark related to the above example concerns the existence of restriction maps between isomorphic homology groups which are not isomorphisms. Suppose open sets $U \subset V$ in a finite T_0-space have the property that $\mathcal{L}(U)$ is isomorphic to $\mathcal{L}(V)$. A natural question to ask is if the restriction map $\mathcal{L}(U \subset V) \colon \mathcal{L}(V) \to \mathcal{L}(U)$ is necessarily an isomorphism. This happens to be true for the sundial example above, but it is not true in general. See Fig. 1 for an illustration of a *pinched torus*. The local homology of the pinched point and any point on the 1-dimensional strata are each isomorphic to the reduced homology of the wedge of two spheres. However, the restriction map from an open neighborhood of the pinched point to an open neighborhood of a 1-simplex in the one strata which is adjacent to the pinched point is not an isomorphism. The coarsest $\mathcal{L}$-stratification we obtain from our algorithm coincides with the stratification given in Fig. 1 (for a suitable triangulation of the pinched torus).

## 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 (defined below). Again, let $|K|$ be the polyhedron in Fig. 6 with labeled vertices.

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
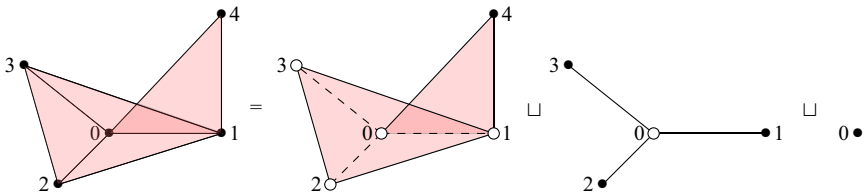
$$S_2 = X_2 - X_1 = [0, 1, 3] \cup [0, 1, 2] \cup [0, 2, 3] \cup [0, 1, 4] \cup [1, 3]$$
$$\cup \, [1, 2] \cup [2, 3] \cup [1, 4] \cup [0, 4].$$
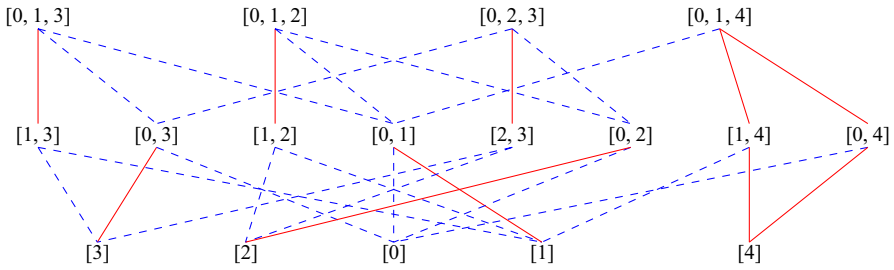
Fig. 9 The labeled Hasse diagram for the sundial with the sheaf of maximal elements. Red solid lines denote isomorphisms, while blue dotted lines denote non-isomorphisms
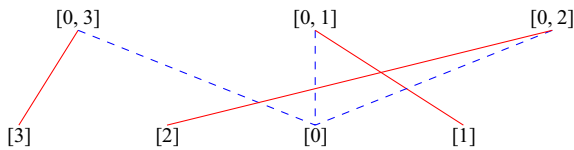
Fig. 10 The labeled Hasse diagram for the sundial with the sheaf of maximal elements after the top-dimensional stratum has been removed

Fig. 11 A triangulated sundial and its stratification based on the sheaf of maximal elements

For this example, the stratum S_2 is automatically homogeneous, hence the two algorithm variations in Sect. 4 (2a. and 2a′.) produce the same 2-stratum S_2, illustrated in Fig. 11. We construct the labeled Hasse diagram based on the sheaf of maximal elements, as shown in Fig. 9.

Next, we can calculate the strata S_1 by only considering restriction maps whose codomain is not contained in S_2 (see Fig. 10). 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 Fig. 11. Again, the stratum S_1 is automatically homogeneous, meaning that the algorithm variations 2a. and 2a′. produce the same out put.

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 stratum pieces where small neighborhoods of points in the same stratum piece intersect the same set of 2-simplices.

7 Stratification Learning Using Geometric Techniques

7.1 Pre-Sheaf of Vanishing Polynomials

In this section we will use *Learning Algebraic Varieties from Samples* [6] to stratify the nerve of an open cover of a point cloud data set. Suppose $X \subset \mathbb{R}^n$ is a finite set of points, and $\{U_i\}$ is a finite cover of X, such that $U_i \subset X$ for each i, and $X = \bigcup_i U_i$. We will proceed by outlining a geometric method for computing a stratification of the nerve of $\{U_i\}$, $\mathcal{N}$, viewed as a finite T_0-space.

We will begin by briefly reviewing the approach to learning algebraic varieties described in [6]. To each algebraic set $S \subset \mathbb{R}^n$, defined to be the set of solutions to a system of polynomial equations, we can associate an ideal of polynomial functions

$$I(S) := \{\text{polynomial function } p \text{ on } \mathbb{R}^n : p(x) = 0 \ \forall \, x \in S\}.$$

If Ω is a finite set of points sampled from S, then $I(\Omega) \supset I(S)$. The insight used in [6], is that certain finite-dimensional subspaces of polynomial functions will not be able to distinguish Ω from S. More precisely, we will start with a finite set of linearly independent polynomial functions $\mathcal{M}$, and consider the subspace $R_{\mathcal{M}}$ consisting of $\mathbb{R}$-linear combinations of elements in $\mathcal{M}$. To a given set $V \subset \mathbb{R}^n$, we will associate the subspace of polynomial functions in $R_{\mathcal{M}}$ which vanish on V:

$$I_{\mathcal{M}}(V) := \{p \in R_{\mathcal{M}} : p(x) = 0 \ \forall \, x \in V\}.$$

The goal is to carefully choose a finite set of polynomials $\mathcal{M}$ so that $I_{\mathcal{M}}(\Omega) = I_{\mathcal{M}}(S)$ and $I_{\mathcal{M}}(S)$ generates the ideal $I(S)$ in the ring of polynomial functions.

The pre-sheaf of vanishing polynomials $\mathcal{I}_{\mathcal{M}}$ is defined using the above association of a finite-dimensional vector space $I_{\mathcal{M}}(V)$ to various point sets $V \subset \mathbb{R}^n$. Given a collection of subsets $\{U_i\}_{i \in J}$, we define an abstract simplicial complex on the index set J by declaring a subset $\tau \subseteq J$ to be a simplex if and only if the corresponding intersection

$$V_\tau := \bigcap_{i \in \tau} U_i \subset X \subset \mathbb{R}^n$$

is non-empty. Finally, given an open set $W \subset \mathcal{N}$ (with respect to the Alexandroff topology defined in Sect. 2.1), define
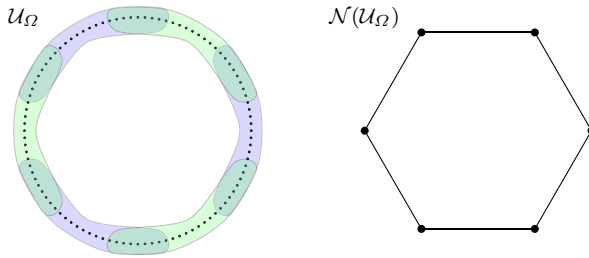
$$X_W = \bigcup_{\tau \in W} V_\tau.$$

Fig. 12 An open cover $\mathcal{U}_\Omega$ of Ω, with corresponding nerve $\mathcal{N}(\mathcal{U}_\Omega)$

The pre-sheaf of vanishing polynomials $\mathcal{I}_\mathcal{M}$ is defined by

$$\mathcal{I}_\mathcal{M}(W) := I_\mathcal{M}(X_W).$$

Remark We could sheafify this pre-sheaf, and get a sheaf of vanishing, locally polynomial functions. However, since the algorithm outlined in this paper only requires a pre-sheaf as an input, we will continue without taking into account the larger space of locally polynomial functions.

Proposition 7.1 *For every finite set of linearly independent polynomials $\mathcal{M}$, the contravariant functor $\mathcal{I}_\mathcal{M}$ from the category of open sets of $\mathcal{N}$ to the category of finite-dimensional vector spaces is a pre-sheaf.*

Proof Assume $W \subset Y \subset \mathcal{N}$ are two open sets. Then $X_W \subset X_Y$. If $p \in I_\mathcal{M}(X_Y)$, then by definition p vanishes on the set X_W. Therefore $I_\mathcal{M}(X_Y) \subset I_\mathcal{M}(X_W)$. The restriction map induced by $\mathcal{I}_\mathcal{M}$ is the inclusion

$$\mathcal{I}_\mathcal{M}(W \subset Y): I_\mathcal{M}(X_Y) \longhookrightarrow I_\mathcal{M}(X_W),$$
$$f \mapsto f.$$

To see that this map is well defined, we notice that if f vanishes on X_Y, and if $X_W \subset X_Y$, then f must vanish on X_W. It follows that $\mathcal{I}_\mathcal{M}(U \subset U) = \mathrm{id}_U$ and $\mathcal{I}_\mathcal{M}(U \subset W) \circ \mathcal{I}_\mathcal{M}(W \subset Y) = \mathcal{I}_\mathcal{M}(U \subset Y)$. $\qquad\square$

Remark Since the restriction maps $\mathcal{I}_\mathcal{M}(W \subset Y)$ are necessarily injective, we can conclude that if $\mathcal{I}_\mathcal{M}(W)$ is isomorphic to $\mathcal{I}_\mathcal{M}(Y)$, then the restriction map $\mathcal{I}_\mathcal{M}(W \subset Y)$ is an isomorphism. Therefore, computing the $\mathcal{I}_\mathcal{M}$-stratification reduces to computing the stalks $\mathcal{I}_\mathcal{M}(\mathrm{St}\,\tau)$ for each simplex τ (rather than computing the restriction maps).

7.2 Examples

Here we will provide examples of geometric stratifications of open covers of finite point sets in $\mathbb{R}^2$ and $\mathbb{R}^3$. These examples aim to illustrate some of the features, as well as subtleties, of the $\mathcal{I}_\mathcal{M}$-stratifications described above.

The circle S^1 Intuitively, we expect that the $\mathcal{I}_\mathcal{M}$-stratification will be trivial (i.e., will consist of a single stratum) when our underlying geometric space is sufficiently
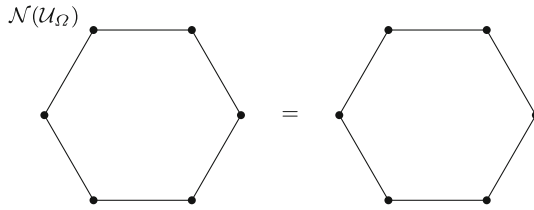
Fig. 13 The minimal homogeneous $\mathcal{I}_{\mathcal{M}}$-stratification of $\mathcal{N}(\mathcal{U}_{\Omega})$ is trivial

well behaved (an analytic manifold, for example). We will begin by checking this intuition for the $\mathcal{I}_{\mathcal{M}}$-stratification of a circle. Consider the finite set Ω consisting of 100 points on the unit circle S^1 in $\mathbb{R}^2$ spaced at regular intervals, with an open cover $\mathcal{U}_{\Omega} = \{U_1, \ldots, U_6\}$ consisting of six open sets (see Fig. 12).

Suppose $\mathcal{M} = \{1, x, y, x^2, y^2, xy\}$ and $U = \bigcap_{i \in I} U_i$ is a subset of Ω corresponding to a simplex of $\mathcal{N}(\mathcal{U}_{\Omega})$. The subspace of polynomials in $R_{\mathcal{M}}$ which vanish on U is equal to the kernel of the linear map

$$R_{\mathcal{M}} \to \mathbb{R}^{|U|},$$
$$p \mapsto (p(x_1, y_1), \ldots, p(x_k, y_k)),$$

where (x_i, y_i) are the elements of U with some fixed order. See [6, Sect. 5] for a discussion concerning how to optimize our choice of $\mathcal{M}$, possibly resulting in the above map being represented by a sparse matrix. For our example we will assume that the subspace of polynomials in $R_{\mathcal{M}}$ which vanish on U is the $\mathbb{R}$-*span* of $x^2 + y^2 - 1$:

$$I_{\mathcal{M}}(U) = \mathbb{R}\langle x^2 + y^2 - 1 \rangle = \{r(x^2 + y^2 - 1) : r \in \mathbb{R}\}.$$

Notice that this calculation follows for any open set $U \subset \mathcal{N}(\mathcal{U}_{\Omega})$. If V and U are two such sets with $V \subset U$, then

$$\mathcal{I}_{\mathcal{M}}(V \subset U) \colon I_{\mathcal{M}}(U) \to I_{\mathcal{M}}(V)$$

is an isomorphism. Therefore, our pre-sheaf $\mathcal{I}_{\mathcal{M}}$ is constant. This implies that the $\mathcal{I}_{\mathcal{M}}$-stratification of $\mathcal{N}(\mathcal{U}_{\Omega})$ is the trivial stratification (i.e., the stratification of $\mathcal{N}(\mathcal{U}_{\Omega})$ consisting of a single stratum). It should be noted that for degree reasons, the trivial stratification would be produced even if fewer points were sampled from the space. Specifically, as long as each cover element U_i contains at least four distinct points, then the resulting pre-sheaf $I_{\mathcal{M}}$ will be constant, and the $I_{\mathcal{M}}$-stratification will be trivial (Fig. 13).

A corner As a second example, we want to explore the extent to which the geometrically defined $\mathcal{I}_{\mathcal{M}}$-stratification is capable of detecting geometric singularities. Such a feature will provide an illustration of the key differences between $\mathcal{I}_{\mathcal{M}}$-stratifications and local homology stratifications. Let us consider the finite set $\Omega := \{(0.1n, 0) : n = 0, \ldots, 20\} \cup \{(0, 0.1n) : n = 0, \ldots, 20\} \subset \mathbb{R}^2$, with the open cover $\mathcal{U}_{\Omega}$ depicted in Fig. 14. Suppose (as in the previous example) that
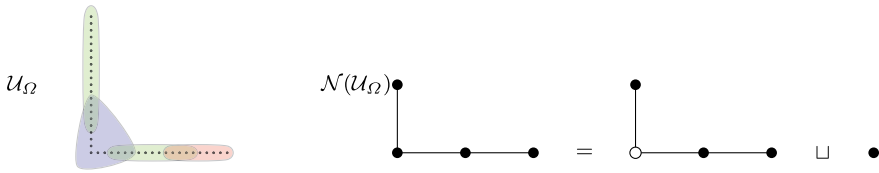
Fig. 14 A stratification of $\mathcal{N}(\mathcal{U}_\Omega)$ using the pre-sheaf of vanishing polynomials

$\mathcal{M} = \{1, x, y, x^2, y^2, xy\}$. If $W \in \mathcal{U}_\Omega$ consists of elements of the form $(x, 0)$, where $x \neq 0$, then $\mathcal{I}_\mathcal{M}(W) = \mathbb{R}\langle y \rangle \oplus \mathbb{R}\langle y^2 \rangle \oplus \mathbb{R}\langle xy \rangle$. Similarly, if $U \in \mathcal{U}_\Omega$ consists of elements of the form $(0, y)$, where $y \neq 0$, then $\mathcal{I}_\mathcal{M}(U) = \mathbb{R}\langle x \rangle \oplus \mathbb{R}\langle x^2 \rangle \oplus \mathbb{R}\langle xy \rangle$. If $V \subset \mathcal{U}_\Omega$ contains $(0, 0)$, then $\mathcal{I}_\mathcal{M}(V) = \mathbb{R}\langle xy \rangle$. The restriction map of vector spaces

$$\mathcal{I}(U \subset V): \mathbb{R}\langle xy \rangle \longrightarrow \mathbb{R}\langle x \rangle \oplus \mathbb{R}\langle x^2 \rangle \oplus \mathbb{R}\langle xy \rangle$$

is not an isomorphism. The resulting $\mathcal{I}_\mathcal{M}$-stratification of $\mathcal{N}(\mathcal{U}_\Omega)$ is illustrated in Fig. 14.

The curve $y^2 = x^3 + x^2$ Now we will give an example that aims to illustrate some of the more subtle properties of $\mathcal{I}_\mathcal{M}$-stratifications. Specifically, we will see a singularity that the local homology sheaf detects, but the $\mathcal{I}_\mathcal{M}$ pre-sheaf does not. This example will consist of a finite set of solutions of $y^2 = x^3 + x^2$. Suppose $\mathcal{M} = \{1, x, y, x^2, y^2, xy, x^3, x^2y, xy^2, y^3\}$. The set of all real solutions of $y^2 = x^3 + x^2$, denoted by X, is parameterized by the map

$$\phi: \mathbb{R} \to X \subset \mathbb{R}^2,$$
$$t \mapsto (t^2 - 1, t^3 - t).$$

Suppose $f \in \mathcal{I}_\mathcal{M}(U)$, for a given open set $U \subset X$ (in the subspace topology). The function

$$g(t) = f(\phi(t)) \in \mathbb{R}[t]$$

is a polynomial function from $\mathbb{R}$ to $\mathbb{R}$. Let $V = \phi^{-1}(U)$. Since f vanishes on U, g must vanish on V. Since V is an open subset of $\mathbb{R}$, we can conclude that g is the zero polynomial. Therefore f vanishes everywhere on X. Therefore, the $\mathcal{I}_\mathcal{M}$-stratification of X is the trivial stratification, even though X has a singular point at $(0, 0)$.

Informed by the calculation above, we turn our attention to the finite open cover of an ϵ-net of X (illustrated in Fig. 15). For each $U \in \mathcal{U}_\Omega$, the vector space of vanishing polynomials is

$$\mathcal{I}_\mathcal{M}(U) = \mathbb{R}\langle x^3 + x^2 - y^2 \rangle.$$

Moreover, the pre-sheaf $\mathcal{I}_\mathcal{M}$ is constant, and the resulting $\mathcal{I}_\mathcal{M}$-stratification is the trivial stratification, illustrated in Fig. 15. Notice that for this example (as well as the circle example) the pre-sheaf $\mathcal{I}_\mathcal{M}$ would be constant for most choices of cover $\mathcal{U}_\Omega$ of X. Specifically, if $\mathcal{U}_\Omega$ consists of two intersecting sets, then the corresponding nerve
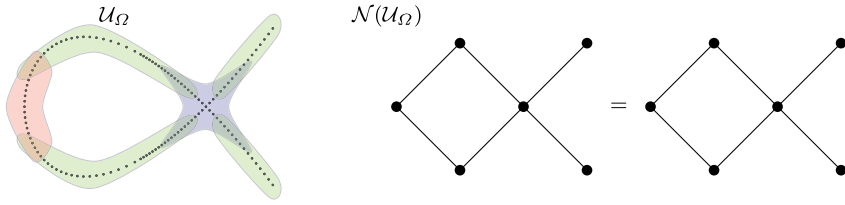
Fig. 15 The minimal homogeneous $\mathcal{I}_{\mathcal{M}}$-stratification of an open cover of a set of solutions to $y^2 = x^3 + x^2$. In this example, we obtain the trivial stratification
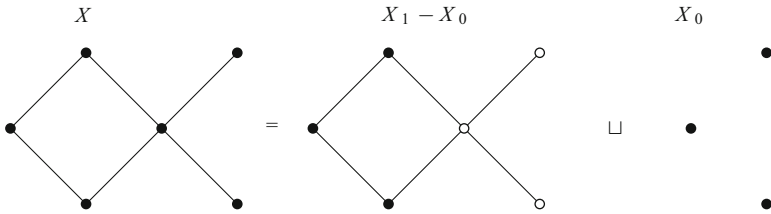
Fig. 16 The minimal homogeneous $\mathcal{L}$-stratification of a triangulation X of a bounded connected set of real solutions to $y^2 = x^3 + x^2$

$\mathcal{N}(\mathcal{U}_{\Omega})$ will consist of two vertices connected by an edge. The $\mathcal{I}_{\mathcal{M}}$-stratification of the 1-simplex $\mathcal{N}(\mathcal{U}_{\Omega})$ will be the trivial stratification.

To contrast with the $\mathcal{I}_{\mathcal{M}}$-stratification illustrated in Fig. 15, we include (Fig. 16) a local homology stratification of the simplicial complex $\mathcal{N}(\mathcal{U}_{\Omega})$. Unlike the polynomial based stratification, the local homology stratification distinguishes the singular point where the elliptic curve crosses over itself, as well as the boundary points of the simplicial complex.

Points sampled from the sundial In this example we study a higher-dimensional corner, as a comparison to the second example in this section. We will assume that we have a point sample Ω of the sundial (see Fig. 2), with an open cover $\mathcal{U}$ of the sundial whose nerve is the triangulation given in Fig. 17. For example, for each vertex τ in the prescribed triangulation, define a cover element U_{τ} to be the image of the open star of τ under the triangulation homeomorphism of Sect. 2. Assume that the point sample Ω and the set of polynomials $\mathcal{M}$ have the property that if U is an intersection of open sets in the open cover, then $\mathcal{I}(U) = \mathcal{I}_{\mathcal{M}}(U \cap \Omega)$ (here $\mathcal{I}(U)$ denotes the set of all real valued polynomials that vanish on U). We will additionally assume that the base of the sundial is contained in a 2-dimensional plane in $\mathbb{R}^3$, with the complement of the base contained in a subspace perpendicular to the 2-dimensional plane. The resulting $\mathcal{I}_{\mathcal{M}}$-stratification is given in Fig. 17.

Relation to the mapper construction As a final example, we will show how $\mathcal{I}_{\mathcal{M}}$-stratifications naturally apply to the mapper construction. The mapper algorithm, originally developed in [27], gives a topological description of the fibers of a continuous function. We will illustrate the fundamental concept of the algorithm through an example. Suppose Ω is an ϵ-net of points on the torus $\mathbb{T}$ in $\mathbb{R}^3$ (illustrated in Fig. 18). In other words, let Ω be any finite subset of $\mathbb{T}$ such that the Hausdorff distance between
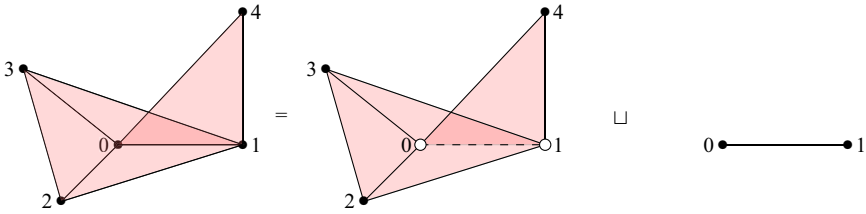
Fig. 17 A triangulated sundial and its minimal homogeneous $\mathcal{I}_{\mathcal{M}}$-stratification
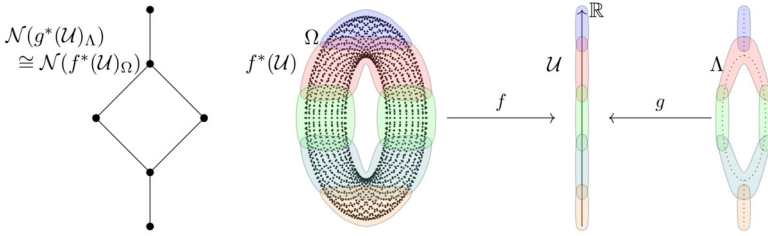

Fig. 18 The mapper construction applied to the height function on the torus as well as the height function on an ellipse with two branching lines

Ω and $\mathbb{T}$ is less than ϵ. Let

$$\Omega_\epsilon = \left\{ x \in \mathbb{R}^3 : \min_{\omega \in \Omega} \|x - \omega\| < \epsilon \right\}$$

be the ϵ-thickening of Ω, and suppose f is a continuous map from Ω_ϵ to $\mathbb{R}$. Let $\mathcal{U}$ be an open cover of $f(\Omega_\epsilon)$. The mapper construction of $\mathcal{U}$ and f, denoted $\mathcal{M}(\mathcal{U}, f) = \mathcal{N}(f^*(\mathcal{U}))$, is the nerve of the connected pull back $f^*(\mathcal{U})$ of the open cover $\mathcal{U}$, where

$$f^*(\mathcal{U}) = \{U \subset \Omega_\epsilon : U \text{ is a connected component of } f^{-1}(V) \text{ for some } V \in \mathcal{U}\}.$$

The $\mathcal{I}_{\mathcal{M}}$-stratification is naturally suited to work with the mapper construction studied in [8,22,27]. Let $f^*(\mathcal{U})_\Omega = \{U \cap \Omega : U \in f^*(\mathcal{U})\}$ be the open cover of the point set Ω (in the discrete topology) induced by $f^*(\mathcal{U})$. Additionally, choose $\mathcal{M}$ so that $R_{\mathcal{M}}$ is the set of all polynomials in three variables with degree less than or equal to 4,

$$R_{\mathcal{M}} = \{ax^{n_1} + by^{n_2} + cz^{n_3} : a, b, c \in \mathbb{R} \text{ and } 0 \leq n_1, n_2, n_3 \leq 4\}.$$

As an example, we will contrast the $\mathcal{I}_{\mathcal{M}}$-stratification obtained from Ω, f, and $\mathcal{U}$ with the $\mathcal{I}_{\mathcal{M}}$-stratification obtained from Λ, g, and $\mathcal{U}$, where Λ is an ϵ-net of points sampled from the ellipse with two branching lines embedded in $\mathbb{R}^3$ (illustrated in Fig. 18), and g is the height function on Λ_ϵ. To distinguish the pre-sheaf of vanishing polynomials defined using Ω and f from the pre-sheaf of vanishing polynomials defined using Λ and g, we will use the notation $\mathcal{I}_{\mathcal{M}}^\Omega$ and $\mathcal{I}_{\mathcal{M}}^\Lambda$, respectively. We can consider the $\mathcal{I}_{\mathcal{M}}^\Omega$-stratification of $\mathcal{N}(f^*(\mathcal{U})_\Omega)$, by taking each open set $V \in f^*(\mathcal{U})_\Omega$ to the vector space

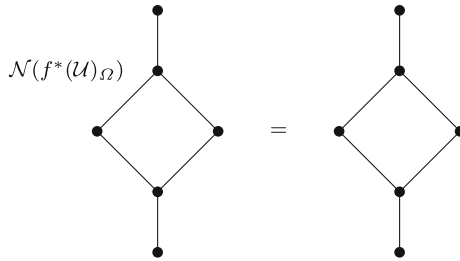**Fig. 19** The minimal homogeneous $\mathcal{I}_{\mathcal{M}}^{\Omega}$-stratification of $\mathcal{N}(f^*(\mathcal{U})_{\Omega})$
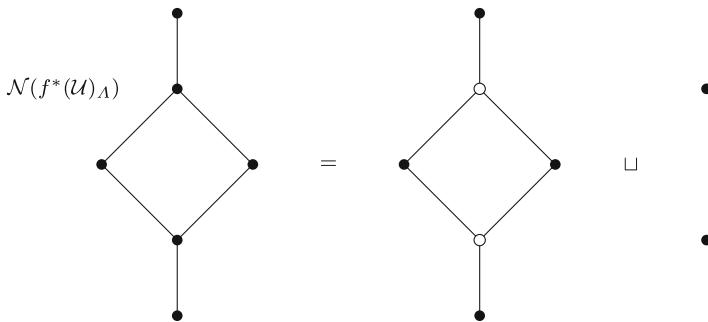

Fig. 20 The minimal homogeneous $\mathcal{I}_{\mathcal{M}}^{\Lambda}$-stratification of $\mathcal{N}(g^*(\mathcal{U})_{\Lambda})$

$I_{\mathcal{M}}(V)$ of polynomials in $\mathcal{M}$ that vanish on V. The $\mathcal{I}_{\mathcal{M}}^{\Omega}$-stratification of $\mathcal{N}(f^*(\mathcal{U})_{\Omega})$ will differ from the $\mathcal{I}_{\mathcal{M}}^{\Lambda}$-stratification of $\mathcal{N}(g^*(\mathcal{U})_{\Lambda})$, even though the underlying topological spaces are homeomorphic. Suppose R and r are the radii of the torus from which the points in Ω are sampled. For each open set $V \in f^*(\mathcal{U})_{\Omega}$,

$$\mathcal{I}_{\mathcal{M}}^{\Omega}(V) = \mathbb{R}\langle(x^2 + y^2 + z^2 + R^2 - r^2)^2 - 4R^2(x^2 + y^2)\rangle,$$

resulting in the trivial stratification (Fig. 19).

However, open sets $V \in g^*(\mathcal{U})_{\Lambda}$ containing only points on the branching lines will result in vector spaces generated (as vector spaces) by polynomials (in $R_{\mathcal{M}}$) of the form $x \cdot p(x, y, z)$. Alternatively, open sets $V \in g^*(\mathcal{U})_{\Lambda}$ which contain only points on the ellipse will result in the vector space $\mathbb{R}\langle cx^2 + dy^2 - 1\rangle$ for constants $c, d \in \mathbb{R}$. Therefore, the $\mathcal{I}_{\mathcal{M}}^{\Lambda}$-stratification (depicted in Fig. 20) is nontrivial.

8 Proofs of Our Main Results

We detail the proofs of our main theorems, that is, the existence of $\mathcal{F}$-stratifications (Proposition 3.1), the existence of coarsest $\mathcal{F}$-stratifications (Theorem 3.2), and the existence and uniqueness of minimal homogeneous $\mathcal{F}$-stratifications (Theorem 3.6).

8.1 Proof of Proposition 3.1

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 $X = \coprod_{x_i \in X} x_i$. In order to insure that the corresponding filtration is a filtration by closed subsets, we need to order our points so that x_i is minimal (with the poset ordering) in the complement of X_{i-1}. 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 $X = \coprod_{x_i \in X} x_i$. □

8.2 Proof of Theorem 3.2

This theorem can be proven immediately by noticing that there are only finitely many stratifications of X (X being a finite T_0-space with finitely 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 $X_{d_0} := X$. 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 complement 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 pre-sheaf $\mathcal{E}$ on B_x^i that 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 $\mathcal{E}(U \subset V) = \mathrm{id} \colon \mathcal{F}(B_x) \to \mathcal{F}(B_x)$ 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 pre-sheaf 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 in turn 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. $\qquad\square$

8.3 Proof of Theorem 3.6

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 $\text{Cl } B_x = \{y \in X_{d_0} : y \leq s \text{ for some } s \in B_x\}$ 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 $\text{Cl } x$ 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

$$H_{d_1} := \left\{x \in X_{d_1} : \text{Cl } B_x^{d_1} \text{ is homogeneous of dimension } d_1\right\} \quad \text{and}$$
$$C_{d_1} := \left\{x \in H_{d_1} : \mathcal{F}(B_w \subset B_y) \text{ is an iso. for all chains } x \leq y \leq w \text{ in } H_{d_1}\right\}.$$

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, $\text{Cl}(X_i - X_{i-1}) = \text{Cl } S_i$ 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 $\emptyset \subset X_0 \subset \cdots \subset X_{d_0} = X$. 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' corresponding 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 homogeneity of $X_i' - X_{i-1}'$). Suppose $a \in S_n$ and $b \in S_n'$ are 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 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 third part of the sheaf axiom (which is also true for pre-sheaves), 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 $\text{Cl } S_n''$ 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 use inductively the same argument as above to show that $X'_i = X_i$ for all i. Therefore the two stratifications are equal, which concludes the proof. □

9 Discussion

Many problems in computational geometry and topology are solved by finding suitable combinatorial models which reflect the geometric or topological properties of a particular space of interest. One way to study properties of a space such as the pinched torus in Fig. 1 is to begin by finding a triangulation of the space, which in some sense provides a combinatorial model which is amenable to computation. The corresponding triangulation can be thought of as a stratification of the space, by defining the d-dimensional stratum to be the collection of d-dimensional simplices. However, this stratification is too "fine" in a sense, as it breaks up the underlying space into too many pieces, resulting in each stratum piece retaining relatively little information about the underlying geometry of the total space. The results of this paper can be interpreted as a method for computing a coarsening of the stratification obtained from the triangulation of our underlying space, using homological techniques (or more generally, sheaf-theoretic techniques) to determine when two simplices should belong to the same coarser stratum.

There are two key features of the sheaf-theoretic stratification learning algorithm which should be highlighted. The first feature is that we avoid computations requiring the sheafification process. At first glance this may be surprising to those not familiar with cellular sheaves, since constructible sheaves cannot 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 pre-sheaf referenced in the definition of the pull back of a sheaf (Sect. 2.2). We can avoid this by noticing two facts. Suppose $\mathcal{E}$ is a pre-sheaf and $\mathcal{E}^+$ is the sheafification of $\mathcal{E}$. First, in the setting of finite T_0-spaces, we can deduce if $\mathcal{E}^+$ is constant by considering how it behaves on minimal open neighborhoods. Second, the behavior of $\mathcal{E}^+$ will agree with the behavior of the pre-sheaf $\mathcal{E}$ 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 pre-sheaf $\mathcal{E}$ applied to minimal open neighborhoods.

The second feature of our algorithm (which is made possible by the first one) 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 build an $\mathcal{F}$-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. The computation of local homology groups can therefore be distributed and computed independently. Additionally, once we have determined whether the local homology restriction maps are isomorphisms, we can quickly compute a coarsest $\mathcal{L}$-stratification, or a minimal homogeneous $\mathcal{L}$-stratification, without requiring any local homology groups to be recomputed. As we saw with the pre-sheaf of vanishing polynomials $\mathcal{I}_\mathcal{M}$, there are often cases of stratifications which rely only on the image of the sheaf applied to each minimal open set, making computations even more streamlined.

There are several interesting questions related to $\mathcal{F}$-stratifications that we will investigate in the future. We are interested in the stability of local homology based stratifications under refinements of triangulations of polyhedra. In this direction, it would be useful to view $\mathcal{L}$-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 asymptotics of persistent local homology based stratifications.

References

1. Alexandroff, P.: Diskrete Räume. Matematicheski Sbornik **2**(**44**)(3), 501–518 (1937)
2. Bendich, P.: Analyzing Stratified Spaces Using Persistent Versions of Intersection and Local Homology. PhD thesis, Duke University (2008)
3. Bendich, P., Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Morozov, D.: Inferring local homology from sampled stratified spaces. In: 48th Annual IEEE Symposium on Foundations of Computer Science, pp. 536–546. IEEE (2007)
4. Bendich, P., Harer, J.: Persistent intersection homology. Found. Comput. Math. **11**(3), 305–336 (2011)
5. Bendich, P., Wang, B., Mukherjee, S.: Local homology transfer and stratification learning. In: Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1355–1370. ACM, New York (2012)
6. Breiding, P., Kališnik, S., Sturmfels, B., Weinstein, M.: Learning algebraic varieties from samples. Rev. Mat. Complut. **31**(3), 545–593 (2018)
7. Brown, A., Wang, B.: Sheaf-theoretic stratification learning. In: 34th International Symposium on Computational Geometry. Leibniz Int. Proc. Inform., vol. 99, # 14. Leibniz-Zent. Inform., Wadern (2018)
8. Carrière, M., Oudot, S.: Structure and stability of the one-dimensional mapper. Found. Comput. Math. **18**(6), 1333–1396 (2018)
9. Cianci, N., Ottina, M.: A new spectral sequence for homology of posets. Topol. Appl. **217**, 1–19 (2017)
10. Curry, J.: Sheaves, Cosheaves and Applications. PhD thesis, University of Pennsylvania (2014)
11. Edelsbrunner, H., Harer, J.L.: Computational Topology: An Introduction. American Mathematical Society, Providence (2010)
12. Goresky, M., MacPherson, R.: Intersection homology theory. Topology **19**(2), 135–162 (1980)

13. Goresky, M., MacPherson, R.: Intersection homology II. Invent. Math. **72**(1), 77–129 (1983)
14. Goresky, M., MacPherson, R.: Stratified Morse Theory. Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 14. Springer, Berlin (1988)
15. Habegger, N., Saper, L.: Intersection cohomology of cs-spaces and Zeeman's filtration. Invent. Math. **105**, 247–272 (1991)
16. Haro, G., Randall, G., Sapiro, G.: Stratification learning: detecting mixed density and dimensionality in high dimensional point clouds. In: Advances in Neural Information Processing Systems (NIPS), vol. 19, pp. 553–560. MIT Press (2006)
17. Hatcher, A.: Algebraic Topology. Cambridge University Press, Cambridge (2002)
18. Kirwan, F., Woolf, J.: An Introduction to Intersection Homology Theory. Chapman & Hall/CRC, Boca Raton (2006)
19. Lerman, G., Zhang, T.: Robust recovery of multiple subspaces by geometric l_p minimization. Ann. Stat. **39**(5), 2686–2715 (2011)
20. McCord, M.C.: Singular homology groups and homotopy groups of finite topological spaces. Duke Math. J. **33**, 465–474 (1966)
21. Mio, W.: Homology manifolds. In: Surveys on Surgery Theory, vol. 1. Ann. of Math. Stud., vol. 145, pp. 323–343. Princeton Univ. Press, Princeton (2000)
22. Munch, E., Wang, B.: Convergence between categorical representations of Reeb space and mapper. In: 32nd International Symposium on Computational Geometry. Leibniz Int. Proc. Inform., vol. 51, # 53. Leibniz-Zent. Inform., Wadern (2016)
23. Munkres, J.R.: Elements of Algebraic Topology. Addison-Wesley, Menlo Park (1984)
24. Nanda, V.: Local cohomology and stratification. Found. Comput. Math. **20**(2), 195–222 (2020)
25. Rourke, C., Sanderson, B.: Homology stratifications and intersection homology. In: Proceedings of the Kirbyfest (Berkeley 1998). Geom. Topol. Monogr., vol. 2, pp. 455–472. Geom. Topol. Publ., Coventry (1999)
26. Shepard, A.D.: A Cellular Description of the Derived Category of a Stratified Space. PhD thesis, Brown University (1985)
27. Singh, G., Mémoli, F., Carlsson, G.: Topological methods for the analysis of high dimensional data sets and 3D object recognition. In: Eurographics Symposium on Point-Based Graphics, pp. 99–100. The Eurographics Association (2007)
28. Skraba, P., Wang, B.: Approximating local homology from samples. In: Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 174–192. ACM, New York (2014)
29. Vidal, R., Ma, Y., Sastry, S.: Generalized principal component analysis (GPCA). IEEE Trans. Pattern Anal. Mach. Intell. **27**(12), 1945–1959 (2005)
30. Weinberger, S.: The Topological Classification of Stratified Spaces. Chicago Lectures in Mathematics. University of Chicago Press, Chicago (1994)