Abstract

Motivated by the measurement of local homology and of functions on noisy domains, we extend the notion of persistent homology to sequences of kernels, images, and cokernels of maps induced by inclusions in a filtration of pairs of spaces. Specifically, we note that persistence in this context is well defined, we prove that the persistence diagrams are stable, and we explain how to compute them.

Keywords. Persistent homology, kernels, images, cokernels, persistence diagrams, vineyards, algorithms, stratified spaces.

1 Introduction

Natural phenomena are often modeled in terms of spaces and functions on these spaces. We argue that it is almost always more appropriate to use a multi-scale hierarchy instead of a single space. One reason is the prevalent multi-scale organization we find in nature, another is that the data we gather about nature is necessarily incomplete and requires interpolation. The multi-scale aspect helps bridging the gap between the data about natural phenomena and the idealized mathematical concepts we use for exploration. In particular, we consider homology groups, which are algebraic structures that define and count holes in a topological space [11]. Their multi-scale extensions are persistent homology groups introduced in [10, 13]. Similar to homology which not only counts but also defines, persistent homology not only measures but also creates the hierarchy.

In all previous settings, the hierarchy is defined by a nested sequence of spaces, \( \mathcal{X}_0 \subseteq \mathcal{X}_1 \subseteq \ldots \subseteq \mathcal{X}_m \), and persistent homology arises from considering the corresponding sequence of homology groups, \( H(\mathcal{X}_0) \to H(\mathcal{X}_1) \to \ldots \to H(\mathcal{X}_m) \), connected from left to right by homomorphic maps induced by inclusion. Persistence tracks when a homology class is born and when it dies. This can also be done for an arbitrary sequence of vector spaces connected by homomorphic maps. Motivation for studying such more general sequences is derived from recent investigations in computational topology. First, Bendich et al. describe a multi-scale assessment of local homology for the purpose of reconstructing a stratified space from a point sample [2]. We will see how sequences of kernels can be used to refine their construction. Second, we use sequences of images to introduce a notion of persistence that filters out noise induced by imprecise specifications of domains. This contrasts standard persistence which can handle imprecise function values but not imprecise domains. As an application, we will approximate the persistence diagram of a function knowing only its values at a finite set of points. The main contributions of this paper are two-fold:

- an algorithm that computes the persistence diagrams of sequences of kernels, images, and cokernels in time at most cubic in the size of the simplicial complexes representing the data;
- applications of the algebraic and algorithmic results to measuring local homology and to approximating persistence diagrams of noisy functions on noisy domains.

Outline. The remainder of this paper is structured as follows. Section 2 introduces the algebra of persistent homology including its extension to sequences of kernels, images and cokernels. Section 3 explains the algorithms for computing the corresponding persistence diagrams for a nested sequence of pairs of spaces and Section 4 proves their correctness. Section 5 presents the two applications of our algebraic and algorithmic results. Section 6 concludes the paper.
2 Algebra

Beginning with a review of persistent homology, this section extends this concept to sequences of kernels, images, and cokernels. It also proves that the persistence diagrams of these extensions are stable.

Persistent homology. This concept is a recent addition to classical homology theory and was originally introduced for ordered simplicial complexes [10]. We follow the exposition in [6] in which we have a topological space $X$ and a continuous function $f : X \to \mathbb{R}$. The sublevel set defined by $a \in \mathbb{R}$ consists of all points with function value at most the threshold, $X_a = f^{-1}(-\infty, a]$. We use the algebraic language of homology theory to characterize how $X_a$ is connected, see e.g. [11]. Adding chains with modulo-2 arithmetic, we write $H_p(X_a)$ for the dimension $p$ homology group over $\mathbb{Z}/2\mathbb{Z}$ of $X_a$ and $H(X_a) = (\ldots, H_p(X_a), H_{p+1}(X_a), \ldots)$ for the infinite series obtained by collecting the groups for all dimensions. Of course, only the groups for $p$ between 0 and the dimension of $X_a$ are possibly non-trivial. To simplify language, we will often ignore the difference between a single homology group and the entire series. Given $a \leq b$, the inclusion between the sublevel sets, $X_a \subseteq X_b$, induces a homomorphism, $f_{a,b} : H(X_a) \to H(X_b)$. For $a = b$ this is an isomorphism and for $a < b$ it may or may not be an isomorphism. A value $a \in \mathbb{R}$ is a homological critical value of $f$ if there is no sufficiently small $\varepsilon > 0$ for which $f_{a-\varepsilon,a+\varepsilon}$ is an isomorphism. We assume that $f$ is tame, that is, it has only finitely many critical values and every sublevel set has only finite rank homology groups.

Let $a_1 < a_2 < \ldots < a_m$ be the critical values of $f$ and consider an interleaved sequence $s_{i-1} < a_i < s_i$ for all $i$. This gives a sequence of spaces, $X_0 \subseteq X_1 \subseteq \ldots \subseteq X_m = X$, where we simplify notation by writing $X_i = X_{s_i}$, and a corresponding sequence of homology groups connected by homomorphisms,

$$H(X_0) \to H(X_1) \to \ldots \to H(X_m).$$

Persistence concerns itself with the history of individual homology classes within this sequence. Specifically, a class $\gamma$ in $H(X_i)$ is born at $a_i$ if it is not in the image of $f_{s_{i-1},s_i} = f_{a_i}$. More precisely, an entire coset is born at $a_i$. Furthermore, if $\gamma$ is born at $a_i$ we say it dies entering $a_j$ if $f_{a_j-1}(\gamma)$ is not contained in the image of $f_{a_j-1,a_j-1}$ but $f_{a_j}(\gamma)$ is contained in the image of $f_{a_j-1,a_j}$. The images of the maps $f_{a_j}$ are referred to as persistent homology groups since they consist of all homology groups born at or before $a_j$ that live beyond $a_j$. Nothing about the definition of birth and death is specific to homology groups. In other words, persistence makes perfect sense for any sequence of vector spaces connected by homomorphisms.

It is convenient to represent the fact that $\gamma$ is born at $a_i$ and dies entering $a_j$ by drawing the point $(a_i, a_j)$ in the two-dimensional plane. By collecting the points for all $p$-dimensional classes we get the dimension $p$ persistence diagram which we denote as $Dgm_p(f)$. Since birth necessarily happens before death all points lie above the diagonal. It is also possible that a class $\gamma$ is born at $a_i$ but does not die since it represents a class of $X_m = X$. In this case, we draw $\gamma$ as the point $(a_i, \infty)$ in the diagram. For technical reasons that will become clear later, we consider all points on the diagonal to be part of the persistence diagram. Similar to homology groups we get a diagram for each dimension and we write $Dgm(f)$ for the infinite series of diagrams. Same as for homology groups and for the maps between them we will simplify language by ignoring the difference between a single diagram and an entire series.

Kernels, images, and cokernels. For the extension of persistence to kernels, images, and cokernels we consider two functions, $f : X \to \mathbb{R}$ and a majorizing function $g : Y \to \mathbb{R}$ defined on a subspace $Y \subseteq X$, that is, $f(y) \leq g(y)$ for all $y \in Y \subseteq X$. Assuming both functions are tame, we order the collection of critical values of $f$ and $g$ and interleave them with a sequence of real values $s_i$. The corresponding sequences of sublevel sets give rise to two parallel sequences of homology groups,

$$H(X_0) \to H(X_1) \to \ldots \to H(X_m),$$

$$H(Y_0) \to H(Y_1) \to \ldots \to H(Y_m),$$

where $X_i = f^{-1}(\infty, s_i]$ and $Y_i = g^{-1}(\infty, s_i]$. The two sequences are connected by homomorphisms $j_i : H(Y_i) \to H(X_i)$ induced by the inclusions $Y_i \subseteq X_i$. We call this the two function setting, in contrast to the more special one function setting in which $g$ is the restriction of $f$ to $Y$. More about the relationship between the two settings later. We are interested in the kernels, images, and cokernels of the connecting homomorphisms,

$$\begin{align*}
\ker j_i &= \{ \gamma \in H(Y_i) \mid j_i(\gamma) = 0 \in H(X_i) \}; \\
\im j_i &= \{ j_i(\gamma) \in H(X_i) \mid \gamma \in H(Y_i) \}; \\
\cok j_i &= H(X_i)/\im j_i.
\end{align*}$$

We note that the “coimage” of the map $j_i$, in symbols $H(Y_i)/\ker j_i$, is isomorphic to the image of this map, and therefore does not deserve any special attention. Figure 1 illustrates this construction for two contiguous spaces in both sequences. The square defined by the four homology groups commutes. It follows that the inclusion $Y_i \subseteq Y_{i+1}$ induces a homomorphism $\ker j_i \to \ker j_{i+1}$. Similarly, the inclusion $X_i \subseteq X_{i+1}$ induces a homomorphism $\im j_i \to \im j_{i+1}$ and another homomorphism $\cok j_i \to \cok j_{i+1}$. We thus get...
sequences of kernels, images, and cokernels,

\[
\begin{align*}
\text{Ker}(g \to f) & : \ker j_0 \to \ker j_1 \to \ldots \to \ker j_m; \\
\text{Im}(g \to f) & : \im j_0 \to \im j_1 \to \ldots \to \im j_m; \\
\text{Cok}(g \to f) & : \cok j_0 \to \cok j_1 \to \ldots \to \cok j_m,
\end{align*}
\]

all connected from left to right by homomorphisms. Homology classes are born and die in these sequences same as in the sequences of homology groups. We can therefore define persistent kernels, persistent images, and persistent cokernels as well as construct the corresponding persistence diagrams, which we denote as \(\text{Dgm}({\ker g \to f})\), \(\text{Dgm}({\im g \to f})\), and \(\text{Dgm}({\cok g \to f})\).

**Birth-death combinations.** We consider the generic case in which changes happen one at a time. An event thus corresponds to a birth, a death, on no change in the kernel, in the image, and in the cokernel, giving rise to 27 different combinations. But the ranks of these groups are not independent, that is,

\[
\begin{align*}
\text{rank } \ker j_i + \text{rank } \im j_i &= \text{rank } H(Y_i); \\
\text{rank } \im j_i + \text{rank } \cok j_i &= \text{rank } H(X_i),
\end{align*}
\]

for all \(i\). We can therefore relate the births and deaths in the three sequences using the births and deaths in the sequences of homology groups of the \(Y_i\) and of the \(X_i\). The first equation eliminates two of the nine combinations for kernels and images. Another combination is eliminated by \(\ker j_i\) being a subgroup of \(H(Y_i)\), hence a death in the kernel implies a death in the homology group. Table 1 lists the remaining six cases. Cases A to F have been described above and the example for Case P also works for Case Q. Case R occurs when \(X_{i-1} = Y_i = X_i\) is a circle and \(Y_{i-1}\) is a point on that circle. Cases Q and R do not happen in the one function setting in which every change has a non-zero effect on the rank of the homology group of \(X_i\).

**Mapping cylinder.** We reduce the two function setting to the one function setting using a construction that will be exploited by the algorithm described in Section 3. Specifically, the mapping cylinder of the pair \(Y \subseteq X\) is the space \(X' = X \cup (Y \times [0,1])\) obtained by gluing \(Y \subseteq X\) to \(X \times \{0\} \subseteq Y \times \{0,1\}\). This is illustrated in Figure 2. The function \(f' : X' \to \mathbb{R}\) agrees with \(f\) on \(X\) and with \(g\) on \(Y' = Y \times \{1\}\),

\[
\begin{array}{|c|c|c|c|}
\hline
\text{Case} & \ker j_i & \im j_i & H(Y_i) \\
\hline
A & \text{birth} & \text{death} & \text{—} \\
B & \text{—} & \text{birth} & \text{birth} \\
C, D & \text{—} & \text{—} & \text{—} \\
E & \text{—} & \text{death} & \text{death} \\
F & \text{death} & \text{—} & \text{death} \\
P & \text{birth} & \text{—} & \text{birth} \\
\hline
\end{array}
\]

Table 1: The five cases in the two function setting relating kernels and images. Except for Case P they also occur in the one function setting.

Case C occurs when \(Y_{i-1} = X_{i-1} = Y_i\) is a point and \(X_i\) is again obtained by adding an arc that forms a circle. We also retain ranks in Case D which occurs when \(X_{i-1}\) is a circle, \(Y_{i-1} = Y_i\) is a point on this circle, and \(X_i\) is obtained by adding a spanning disk. Case E occurs when \(Y_{i-1} = X_{i-1}\) is a circle and \(Y_i = X_i\) is obtained by adding a spanning disk. Case F occurs when \(X_{i-1}\) is a disk, \(Y_{i-1}\) is its boundary circle, and we get \(Y_i\) and \(X_i\) by adding another spanning disk to both spaces. Finally, Case P occurs when \(X_{i-1} = X_i\) is a disk, \(Y_{i-1}\) is a point of its boundary circle, and \(Y_i\) is obtained by adding the rest of the circle. This last case happens in the two function setting but not in the one function setting because it requires points that are added to the sublevel set of \(g\) strictly after they are added to the sublevel set of \(f\).

Similarly, the second equation eliminates two of the nine combinations for images and cokernels. Another combination is eliminated by \(\im j_i\) being a subgroup of \(H(X_i)\), hence a death in the image implies a death in the homology group. Table 2 lists the remaining six cases. Cases A to F have been described above and the example for Case P also works for Case Q. Case R occurs when \(X_{i-1} = Y_i = X_i\) is a circle and \(Y_{i-1}\) is a point on that circle. Cases Q and R do not happen in the one function setting in which every change has a non-zero effect on the rank of the homology group of \(X_i\).

\[
\begin{array}{|c|c|c|c|}
\hline
\text{Case} & \im j_i & \cok j_i & H(X_i) \\
\hline
A, E & \text{death} & \text{—} & \text{death} \\
B & \text{—} & \text{birth} & \text{birth} \\
C, F & \text{—} & \text{birth} & \text{birth} \\
D & \text{—} & \text{death} & \text{death} \\
Q & \text{—} & \text{—} & \text{—} \\
R & \text{birth} & \text{—} & \text{—} \\
\hline
\end{array}
\]

Table 2: The five cases in the two function setting relating cokernels and images. Except for Cases Q and R they also occur in the one function setting.
linearly interpolating in between, that is, 
\[ f'(x) = f(x) \] for every \( x \in X \) and \( f'(y, t) = (1 - t)f(y) + tg(y) \) for every \( y \in Y \) and every \( t \in [0, 1] \). The pair of functions \( f' \) and \( g' = f'|_Y \) induces homomorphisms \( j'_i : H(Y'_i) \to H(X'_i) \); see Figure 3. The corresponding sequences of kernels, images, and cokernels are

\[
\begin{align*}
\text{Ker}(g' - f') \quad & : \quad \ker j'_0 \to \ker j'_1 \to \ldots \to \ker j'_m; \\
\text{Im}(g' - f') \quad & : \quad \text{im} j'_0 \to \text{im} j'_1 \to \ldots \to \text{im} j'_m; \\
\text{Cok}(g' - f') \quad & : \quad \text{cok} j'_0 \to \text{cok} j'_1 \to \ldots \to \text{cok} j'_m.
\end{align*}
\]

We claim that they contain the same information as the sequences \( \text{Ker}(g - f) \), \( \text{Im}(g - f) \), and \( \text{Cok}(g - f) \).

**Stability.** An important property of the persistence diagrams is their stability originally proved in [6]. More precisely, the bottleneck distance between the diagrams of two functions \( f \) and \( f'' \) is bounded from above by the difference between the functions, \( d_B(Dgm(f), Dgm(f'')) \leq ||f - f''||_\infty \). Here \( d_B \) is the maximum of \( ||u - \gamma(u)||_\infty \), where \( u \) is a point in the diagram of \( f \) and \( \gamma \) is a dimension-preserving bijection between the diagrams of \( f \) and \( f'' \). Recall that the points on the diagonal belong to the diagrams and can therefore be used in the effort to find a matching \( \gamma \) that minimizes the length of the longest edge. The proof of stability given in [6] can be adapted to the setting in this paper. Specifically, we consider the maps \( j_a : H(Y_a) \to H(X'_a) \) and \( j''_a : H(Y''_a) \to H(X''_{a+\varepsilon}) \), where \( \varepsilon \) is the larger of the two differences between functions, \( ||f - f''||_\infty \) and \( ||g - g''||_\infty \), and \( Y_a, X_a, Y''_a, X''_{a+\varepsilon} \) are the interval sets of \( g, f, g'', f'' \) for thresholds \( a \) and \( a + \varepsilon \). To adapt the proof we need that the maps induced by the inclusions \( Y_a \subseteq Y''_a \) and \( X_a \subseteq X''_{a+\varepsilon} \) send the kernel of \( j''_a \) into the kernel of \( j'_{a+\varepsilon} \). But this follows from the commutativity of the diagram

\[
\begin{array}{c}
H(X'_a) \to H(X'_{a+\varepsilon}) \\
\uparrow j_a \\
H(Y'_a) \to H(Y''_{a+\varepsilon})
\end{array}
\]

Similarly, we need that the inclusions \( Y''_a \subseteq Y_{a+\varepsilon} \) and \( X''_{a+\varepsilon} \subseteq X_{a+\varepsilon} \) send ker \( j''_a \) into ker \( j_{a+\varepsilon} \) which follows by symmetry. With this property the original proof of stability goes through and we refer to [6] for details. The arguments for the images and the cokernels are the same and we state the results.

**Stability Theorem.** Let \( f, f'' : X \to \mathbb{R} \) and \( g, g'' : Y \to \mathbb{R} \) with \( f(y) \leq g(y) \) and \( f''(y) \leq g''(y) \) for every \( y \in Y \subseteq X \) and \( \varepsilon = \max\{||f - f''||_\infty, ||g - g''||_\infty\} \). Then the bottleneck distance between the persistence diagrams is bounded from above by the difference between the functions:

\[
d_B(Dgm(grp g - f), Dgm(grp g'' - f'')) \leq \varepsilon,
\]

for \( \text{grp} \in \{\text{ker}, \text{im}, \text{cok}\} \), provided \( f, g, f'', \) and \( g'' \) are continuous and tame and there is a triangulation of \( X \) in which \( Y \) arises as a subcomplex.
3 Algorithms

In this section, we describe the algorithms for computing the persistence diagrams of the sequences of kernels, images, and cokernels. At their core is the reduction of a matrix as introduced in [8] which we describe first.

Matrix reduction. To get a finite representation of the data we now substitute a function on a simplicial complex for the continuous function on a topological space. In particular, we assume a (finite) simplicial complex $K$ and an injective function $f : K \to \mathbb{R}$ that maps each simplex to a real number. The only additional requirement is $f(\sigma) < f(\tau)$ whenever $\sigma$ is a face of $\tau$. Equivalently, $K_a = f^{-1}(-\infty, a]$ is a subcomplex of $K$ for every $a \in \mathbb{R}$. We index the simplices of $K$ such that $f(\sigma_1) < f(\sigma_2) < \ldots < f(\sigma_m)$ and let $K_i$ be the complex consisting of the first $i$ simplices in the sequence. To compute persistence for the sequence of complexes $K_i$ we let $D$ be the $m$-by-$m$ incidence matrix defined by $D[i, j] = 1$ if $\sigma_i$ is a co-dimension one face of $\sigma_j$ and $D[i, j] = 0$ otherwise. We reduce $D$ using left-to-right modulo-2 column additions until the lowest one of every non-zero column is in a unique row. Initializing $R$ and $V$ to the incidence and the identity matrices and letting $\text{low}(i)$ be the row index of the lowest one in column $i$ of $R$, or 0 if the entire column is zero, we can formalize the algorithm as follows.

$$R = D; V = I;$$
for $i = 1$ to $m$
do
  while $\exists k < i$ with $\text{low}(k) = \text{low}(i) \neq 0$
do
    add column $k$ to column $i$ in $R$ as well as in $V$
  endwhile
endfor.

Equivalently, the reduced matrix is obtained by multiplying the incidence matrix from the right with an upper-triangular matrix, $R = DV$, such that the map from the non-zero columns of $R$ to the row indices of their lowest ones is injective. As proved in [8], $R$ is not unique but the map is. By construction, the rows of $R$ and $V$ correspond to individual simplices, same as the rows of $D$, but the columns of $R$ and $V$ correspond to chains. Specifically, column $i$ of $R$ stores the boundary of the chain stored in column $i$ of $V$. We call $\sigma_i$ positive if its addition to $K_{i-1}$ gives birth to a homology class. Equivalently, column $i$ of $V$ stores a cycle and column $i$ of $R$ is zero. Symmetrically, we call $\sigma_i$ negative if its addition to $K_{i-1}$ gives death to a homology class. Equivalently, column $i$ of $V$ stores a chain that is not a cycle and column $i$ of $R$ is non-zero. The significance of the lowest one in this column of $R$ is that the negative $\sigma_i$ is paired with the positive $\sigma_\ell$, with $\ell = \text{low}(i)$, which gives birth to the class that $\sigma_\ell$ kills.

The persistence diagrams of $f$ can be obtained from the reduced matrix. Specifically, each lowest one, $\ell = \text{low}(i)$, corresponds to a pair of simplices, $\sigma_\ell, \sigma_i$, and we draw the point $(f(\sigma_\ell), f(\sigma_i))$ in the diagram whose dimension is the same as that of $\sigma_\ell$.

Partial and reordered matrices. We prepare the computation of the persistence diagrams by reducing five matrices. By the Mapping Cylinder Lemma, we can restrict ourselves to the one function setting. We therefore assume two simplicial complexes, $L \subseteq K$, let $f : K \to \mathbb{R}$ be an injective function whose sublevel sets are subcomplexes of $K$, and let $g$ be the restriction of $f$ to $L$. We write $D_f$ for the incidence matrix of $K$ whose rows and columns are ordered by $f$. Similarly, we write $D_g$ for the incidence matrix of $L$ whose rows and columns are ordered by $g$.

Step 1 Reduce the two incidence matrices to get $R_f = D_fV_f$ and $R_g = D_gV_g$.

Step 2 Reorder the rows of $D_f$ leaving the columns untouched to get a new matrix $D_{im}$. Specifically, its rows correspond to the simplices in $L$, ordered by $g$, followed by the simplices in $K - L$, ordered by $f$. Reduce the new matrix to get $R_{im} = D_{im}V_{im}$; see Figure 5.

$$\begin{align*}
R & = D; V = I; \\
& \text{for } i = 1 \text{ to } m \text{ do} \\
& \quad \text{while } \exists k < i \text{ with } \text{low}(k) = \text{low}(i) \neq 0 \text{ do} \\
& \quad \quad \text{add column } k \text{ to column } i \text{ in } R \text{ as well as in } V \\
& \quad \text{endwhile} \\
& \text{endfor.}
\end{align*}$$

Figure 4: The reduced matrix equals the incidence matrix times the chain matrix. All three are upper-triangular.

$$\begin{align*}
R & = D; V = I; \\
& \text{for } i = 1 \text{ to } m \text{ do} \\
& \quad \text{while } \exists k < i \text{ with } \text{low}(k) = \text{low}(i) \neq 0 \text{ do} \\
& \quad \quad \text{add column } k \text{ to column } i \text{ in } R \text{ as well as in } V \\
& \quad \text{endwhile} \\
& \text{endfor.}
\end{align*}$$

Figure 5: Matrices computed in the reduction of the incidence matrix of $K$ with reordered rows. The matrix $V_{im}$ is upper-triangular with all ones in the diagonal.

Step 3 Delete some of the columns from $V_{im}$ and reorder the rows to get a new matrix $D_{ker}$. Specifically, keep the columns that represent cycles and remove all others. Furthermore, reorder the rows so they correspond to the simplices in $L$, ordered by $g$, followed by the simplices in $K - L$, ordered by $f$. Finally, reduce the new matrix to get $R_{ker} = D_{ker}V_{ker}$.

Step 4 Starting again with $D_f$, replace some of the columns to get a new matrix $D_{cok}$. Specifically, substitute

$$\begin{align*}
R & = D; V = I; \\
& \text{for } i = 1 \text{ to } m \text{ do} \\
& \quad \text{while } \exists k < i \text{ with } \text{low}(k) = \text{low}(i) \neq 0 \text{ do} \\
& \quad \quad \text{add column } k \text{ to column } i \text{ in } R \text{ as well as in } V \\
& \quad \text{endwhile} \\
& \text{endfor.}
\end{align*}$$
columns in $V_g$ that represent cycles for the corresponding columns in $D_f$, adding zeros to compensate for the simplices in $K - L$, which are missing in $V_g$. Reduce the new matrix to get $R_{cok} = D_{cok} V_{cok}$.

We note that reducing $D_f$ is redundant because the type information it furnishes is also available from $R_{im}$. We still use $R_f$ because this clarifies which information is used where.

**Births, deaths, and pairs.** We use the reduced matrices to compute the persistence diagrams of the sequences of kernels, images, and cokernels. Specifically, $R_f$ and $R_g$ (and in one case $R_{im}$) decide which simplices give birth and which give death and $R_{ker}$, $R_{im}$, $R_{cok}$ determine how the births match up with the deaths. We begin with the sequence of kernels and recall the relevant Cases A and F in Table 1.

**Algorithm for kernels:**

**Birth.** A simplex $\sigma$ gives birth in $\text{Ker}(g \rightarrow f)$ iff $\sigma \in K - L$, $\sigma$ is negative in $R_f$, and the lowest one in its column in $R_{im}$ corresponds to a simplex in $L$.

**Death.** A simplex $\tau$ gives death in $\text{Ker}(g \rightarrow f)$ iff $\tau \in L$, $\tau$ is negative in $R_g$, and $\tau$ is positive in $R_f$. In this case, the lowest one in the column of $\tau$ in $R_{ker}$ corresponds to a simplex $\sigma \in K - L$ that gives birth in $\text{Ker}(g \rightarrow f)$. Then $(\sigma, \tau)$ is a pair.

A dimension $p$ homology class is given birth to in the kernel by a $(p+1)$-simplex and it dies at the hand of another $(p+1)$-simplex. The dimension $p$ persistence diagram thus consists of all points $(f(\sigma), f(\tau))$ encoding pairs of $(p+1)$-simplices identified in the Death case as well as all points $(f(\sigma), \infty)$ encoding unpaired $(p+1)$-simplices identified in the Birth case. We continue with the sequence of images and recall the relevant Cases A, B, E in Tables 1 and 2.

**Algorithm for images:**

**Birth.** A simplex $\sigma$ gives birth in $\text{Im}(g \rightarrow f)$ iff $\sigma \in L$ and $\sigma$ is positive in $R_g$.

**Death.** A simplex $\tau$ gives death in $\text{Im}(g \rightarrow f)$ iff $\tau$ is negative in $R_f$ and the lowest one in its column in $R_{im}$ corresponds to a simplex $\sigma \in L$. Then $(\sigma, \tau)$ is a pair.

Note that the Death case splits into Case A with $\tau \in K - L$ and Case E with $\tau \in L$. The dimension $p$ persistence diagram consists of all points $(f(\sigma), f(\tau))$ encoding pairs of $p$- and $(p+1)$-simplices identified in the Death case as well as points $(f(\sigma), \infty)$ encoding unpaired $p$-simplices identified in the Birth case. We continue with the sequence of cokernels and recall the relevant Cases C, F and D in Table 2.

**Algorithm for cokernels:**

**Birth.** A simplex $\sigma$ gives birth in $\text{Cok}(g \rightarrow f)$ iff $\sigma$ is positive in $R_f$ and it is either in $K - L$ or negative in $R_g$.

**Death.** A simplex $\tau$ gives death in $\text{Cok}(g \rightarrow f)$ iff $\tau$ is negative in $R_f$ and the lowest one in its column in $R_{im}$ corresponds to a simplex in $K - L$. In this case, the lowest one in the column of $\tau$ in $R_{cok}$ corresponds to a simplex $\sigma$ that gives birth in $\text{Cok}(g \rightarrow f)$. Then $(\sigma, \tau)$ is a pair.

The dimension $p$ persistence diagram consists of all points $(f(\sigma), f(\tau))$ encoding pairs of $p$- and $(p+1)$-simplices identified in the Death case as well as points $(f(\sigma), \infty)$ encoding unpaired $p$-simplices identified in the Birth case. The running time of the three algorithms is $O(m^3)$, same as the original persistence algorithm given in [10]. Furthermore, it is possible to extend the algorithm in [8] so that it maintains the reduced matrices in time $O(m)$ per transposition of contiguous simplices in the ordered sequences; details can be found in Appendix A. This is the method of choice for computing the vineyard of a pair of 1-parameter families of functions $f$ and $g$, as they arise in applications considered in Section 5.

## 4 Correctness

We prove the correctness of the algorithms inductively, by considering one simplex at a time. For each index $i$, we consider the actual births, deaths, and pairs that occur in the sequences up to $j_i$, and the computed births, deaths, and pairs reported by the algorithm working on the simplices up to $\sigma_i$. To prove that the corresponding sets are the same at the end, for $i = m$, we show they are the same throughout, for all $i$. We do this in two steps, first proving that the algorithms are necessary and second that they are sufficient. In other words, we first prove that the computed information is correct and second that it is complete.

**Preparation.** We begin with a few preliminary observations. Recall that Tables 1 and 2 list the possible combinations of births and deaths under the simplifying assumption that each group has at most one change happening at any one time. This is indeed the situation if we add individual simplices to a growing complex; see Appendix B. We can therefore use the two tables in the correctness proof, but since we only consider the one function setting, we can further simplify and combine them into Table 3. We get $K_i$ by

<table>
<thead>
<tr>
<th>Case</th>
<th>$\ker(j_i)$</th>
<th>$\im(j_i)$</th>
<th>$\cok(j_i)$</th>
<th>$\text{H}(L_i)$</th>
<th>$\text{H}(K_i)$</th>
<th>$\sigma_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>birth</td>
<td>death</td>
<td>cok</td>
<td>$K - L$</td>
<td>L</td>
<td>$K - L$</td>
</tr>
<tr>
<td>B</td>
<td>$\ldots$</td>
<td>birth</td>
<td>birth</td>
<td>$K - L$</td>
<td>L</td>
<td>$K - L$</td>
</tr>
<tr>
<td>C</td>
<td>$\ldots$</td>
<td>birth</td>
<td>$\ldots$</td>
<td>$K - L$</td>
<td>L</td>
<td>$K - L$</td>
</tr>
<tr>
<td>D</td>
<td>$\ldots$</td>
<td>death</td>
<td>$\ldots$</td>
<td>$K - L$</td>
<td>L</td>
<td>$K - L$</td>
</tr>
<tr>
<td>E</td>
<td>death</td>
<td>$\ldots$</td>
<td>death</td>
<td>$L$</td>
<td>$\ldots$</td>
<td>L</td>
</tr>
<tr>
<td>F</td>
<td>death</td>
<td>birth</td>
<td>death</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$L$</td>
</tr>
</tbody>
</table>

Table 3: In the one function setting there are six cases in which the addition of $\sigma_i$ changes the kernel, the image, or the cokernel.
adding $\sigma_i$ to $K_{i-1}$. If $\sigma_i \in L$ then $L_i = L_{i-1} \cup \{\sigma_i\}$ else $L_i = L_{i-1}$. In Cases B, E, F, the addition of $\sigma_i$ changes the homology of $L_{i-1}$, which can only happen if $\sigma_i \in L$. In the remaining three cases, the addition of $\sigma_i$ changes the homology of $K_{i-1}$ but not that of $L_{i-1}$, which can only happen if $\sigma_i \in K - L$. Note also that the change in the homology of $K_{i-1}$ is unambiguous in all cases, that is, $\sigma_i$ is positive in Cases B, C, F, and negative in Cases A, D, E. We note that each death is paired with a unique birth but some births remain unpaired until the very end. It is convenient to rephrase the pairing condition in a form that is most directly useful in the argument below. We say a cycle appears in $\text{grp} j_i$ if the class it represents is born at that group, where $\text{grp} \in \{\text{ker}, \text{im}, \text{cok}\}$ as usual. We note that the cycle might exist in the complex before it appears in the group.

**Death Lemma.** Let $l < i$ be indices and $z$ a cycle that appears first in $\text{grp} j_i$ and is zero in $\text{grp} j_l$. If there is no index $i' < i$ for which there is a cycle that first appears in $\text{grp} j_l$ and is zero in $\text{grp} j_{i'}$, then the class represented by $z$ is born at $\text{grp} j_l$ and dies entering $\text{grp} j_i$.

Next we consider the incidence matrices used to compute the persistence diagrams of the sequences of kernels, images, and cokernels. To simplify language, we let $M[i]$ be the column of matrix $M$ that corresponds to $\sigma_i$ or, alternatively, the set of simplices whose corresponding rows have a one in this column. We will refer to it as column $i$ of $M$ and note that in some cases it is not the $i$-th column from the left, for example when $M = R_{\text{ker}}$.

**Observation.** Recall that $R_f, R_g, R_{\text{ker}}, R_{\text{im}}, R_{\text{cok}}$ are the reduced incidence matrices computed by the algorithms in Section 3.

(i) If $R_g[i] = 0$ then $R_f[i] = 0$.
(ii) $R_f[i] = 0$ iff $R_{\text{im}}[i] = 0$.
(iii) If $\sigma_i \in L$ and $R_f[i] \neq 0$ then the lowest one in $R_{\text{im}}[i]$ corresponds to a simplex in $L$.
(iv) The columns of $R_{\text{ker}}$ are all non-zero.
(v) If $\sigma_i \in K - L$ then the lowest one in $R_{\text{ker}}[i]$ corresponds to $\sigma_i$.
(vi) If $\sigma_i \in L$, $R_g[i] \neq 0$, and $R_f[i] = 0$ then the lowest one in $R_{\text{ker}}[i]$ corresponds to a simplex in $K - L$.

**Proof.** Except for added zeros the columns of a simplex in $L$ are the same in $D_g$ and in $D_f$. This implies (i). The reordering of rows does not change the rank of the matrix. This implies (ii). By Observation (i), $R_g[i] \neq 0$ follows from $\sigma_i \in L$ and $R_f[i] \neq 0$. Since the columns of $\sigma_i$ in $D_g$ and $D_{\text{im}}$ are the same, except for added zeros as before, the lowest one in $R_{\text{im}}[i]$ cannot be lower than that in $R_g[i]$. This implies (iii). The matrix $V_{\text{im}}$ is upper triangular with a diagonal of ones. It thus has full rank and so does $D_{\text{ker}}$ which consists of a subset of the columns in $V_{\text{im}}$. This implies (iv). We get $D_{\text{ker}}$ from this subset of columns by reordering the rows, moving simplices in $L$ up and simplices in $K - L$ down. The reordering maintains the relative order of the simplices in $K - L$. This implies (v).

Finally, we prove (vi) by contradiction, assuming the lowest one in $R_{\text{ker}}[i]$ corresponds to a simplex $\sigma_i \in L$. Since $R_f[i] = 0$ we have $R_{\text{im}}[i] = 0$. It follows that column $i$ of $V_{\text{im}}$, part of $D_{\text{ker}}$, after reordering the rows. Because of the upper triangular structure of $V_{\text{im}}$, the diagonal ones may be moved by the reordering but they are not cancelled in the reduction. By assumption, the lowest one in $R_{\text{ker}}[i]$ corresponds to a simplex in $L$ which can therefore only be $\sigma_i$, that is, $l = i$. But then $R_{\text{ker}}[i]$ stores a cycle in $L$, a contradiction to $R_g[i] \neq 0$.

**Inductive step.** We assume inductively that the actual and the computed sets of births, deaths, and pairs are the same up to index $i - 1$. This is clearly true for $i - 1 = 0$, when all sets are empty.

**Necessity.** Using this as the induction hypothesis, we first show that the computed sets of births, deaths, and pairs are subsets of the corresponding actual sets up to index $i$.

**Images.** The algorithm for the images reports a birth for the new simplex iff $\sigma_i \in L$ and $R_g[i] = 0$. In this case, $V_{\text{im}}[i]$ stores a cycle that represents a new class in the homology group of $L_i$ as well as in $\text{im} j_i$. Hence, $\sigma_i$ gives rise to an actual birth in the sequence of images.

The algorithm reports a death iff $R_f[i] \neq 0$ and the lowest one in $R_{\text{im}}[i]$ corresponds to a simplex $\sigma_i \in L$. The reduced column is a sum of boundaries in $K_i$,

$$R_{\text{im}}[i] = \sum_{\sigma_i \in V_{\text{im}}[i]} D_{\text{im}}[k].$$

This cycle first appears in $\text{im} j_i$ and it is zero in $K_i$ and therefore also in $\text{im} j_i$. Furthermore, there is no $i' < i$ for which there is a chain that first appears in $\text{im} j_i$ and is zero in $\text{im} j_{i'}$. Otherwise, $(\sigma_{i'}, \sigma_{i})$ would be a pair and the lowest one in $R_{\text{im}}[i']$ would correspond to $\sigma_i$ by inductive hypothesis. But then $R_{\text{im}}$ could be reduced further, a contradiction. By the Death Lemma, $\sigma_i$ gives rise to an actual death and $(\sigma_i, \sigma_i)$ is an actual pair in the sequence of images.

**Cokernels.** The algorithm for the cokernels reports a birth iff $R_f[i] = 0$ and either $\sigma_i \in K - L$ or else $\sigma_i \in L$ and $R_g[i] \neq 0$. In this case, we have indeed a new class in the cokernel, namely the one represented by $V_{\text{im}}[i]$.

The algorithm reports a death iff $R_f[i] \neq 0$ and the lowest one in $R_{\text{im}}[i]$ corresponds to a simplex in $K - L$. By Observation (iii), this implies $\sigma_i \in K - L$. Letting $\sigma_{i'}$ correspond to the lowest one in $R_{\text{cok}}[i]$, the algorithm reports $(\sigma_i, \sigma_{i'})$ as a pair. In this case, the reduced column is a sum of boundaries in $K_i$ and cycles in $L_i$,

$$R_{\text{cok}}[i] = \sum_{\sigma_{i'} \in V_{\text{cok}}[i]} D_{\text{cok}}[k].$$
This cycle appears first in $\text{cok } j_l$ and it is zero in $K_l$ and therefore also in $\text{cok } j_i$. Furthermore, there is no index $i' < i$ for which there is a chain that first appears in $\text{cok } j_l$ and is zero in $\text{cok } j_{i'}$. As before, we use induction and the fact that $R_{\text{cok}}$ is reduced to prove this claim. By the Death Lemma, $\sigma_i$ gives rise to an actual death and $(\sigma_i, \sigma_i)$ is an actual pair in the sequence of cokernels.

**Kernels.** The algorithm for the kernels reports a birth iff $\sigma_i \in K - L$, $R_f[i] \neq 0$, and the lowest one in $R_{\text{im }}[i]$ corresponds to a simplex in $L$. In this case, $R_{\text{im }}[i]$ is a cycle in $L$ and $V_{\text{im }}[i]$ is a chain whose boundary is this cycle. Furthermore, $i$ is the smallest index for which such a chain exists, else we could use induction to show that $R_{\text{im }}$ can be further reduced. Since $\sigma_i$ belongs to $V_{\text{im }}[i]$, this chain does not belong to $L$. Hence, $R_{\text{im }}[i]$ represents a class in the kernel and $\sigma_i$ gives rise to an actual birth.

The algorithm reports a death iff $\sigma_i \in L$, $R_g[i] \neq 0$, and $R_f[i] = 0$. By Observation (vi), the lowest one in $R_{\text{ker }}[i]$ corresponds to a simplex $\sigma_i \in K - L$. The algorithm reports $(\sigma_i, \sigma_i)$ as a pair. To prove that there is an actual death, we recall that the columns in $D_{\text{ker }}[k]$ are cycles in $K$. We write each cycle as a sum of two chains, one in $L$ and the other in its complement, $D_{\text{ker }}[k] = \sum \lambda_t + \sum \kappa_t$, where the $\lambda_t$ belong to $L$ and the $\kappa_t$ belong to $K - L$. The two chains share their boundary, which we denote as $z_k = \partial \sum \lambda_t = \partial \sum \kappa_t$. Clearly, $z_k$ is a cycle in $L$. Because it bounds the sum of the $\kappa_t$, the cycle belongs to the kernel, and because it bounds the sum of the $\lambda_t$, the cycle is zero in $L$ and therefore also zero in the kernel. Consider the cycle

$$z = \sum_{\sigma_i \in V_{\text{ker }}[i]} z_k.$$  

We claim that the class it represents in the kernel is born at $\ker j_l$. Indeed, if it were born earlier there would be $l' < l$ and a chain $c \in K_{l'}$ whose boundary is $z$. But then $\sigma_i$ would be positive and by Observation (v) it would be the lowest one of its own column and not that of column $i$. Now, $V_{\text{ker }}[i]$ provides the sum we need to finish the proof using the Death Lemma. As before, we use the induction hypothesis and the fact that $R_{\text{ker }}$ is reduced to conclude that there is no index $i' < i$ for which a cycle appears in $\ker j_l$ and is zero in $\ker j_{l'}$. Therefore, $\sigma_i$ gives rise to an actual death and $(\sigma_i, \sigma_i)$ is an actual pair in the sequence of kernels.

**Sufficiency.** We second show that the algorithms are complete, that is, the actual births, deaths, and pairs are subsets of the corresponding computed sets. Since there is a bijection between the deaths and the pairs, it suffices to prove the containment for the births and the deaths. We use Table 3 to do this by exhaustive case analysis.

**Case 1** \(\sigma_i \in K - L\).

**Case 1.1** \(R_f[i] = 0\). This is Case C in Table 3. The only change is a birth in the cokernel and this is correctly reported by the algorithms.

**Case 1.2** \(R_f[i] \neq 0\). Let $\sigma_i$ correspond to the lowest one in $R_{\text{im }}[i]$.

**Case 1.2.1** $\sigma_i \in K - L$. From the above analysis we know that this corresponds to a death in the cokernel. This is Case D in Table 3. There are no other changes and this is correctly reported by the algorithms.

**Case 1.2.2** $\sigma_i \in L$. From the above analysis we know that this corresponds to a birth in the kernel and a death in the image. This is Case A in Table 3. There are no other changes and this is correctly reported by the algorithms.

**Case 2** $\sigma_i \in L$.

**Case 2.1** $R_f[i] = 0$.

**Case 2.1.1** $R_g[i] \neq 0$. This is Case F in Table 3. There is a death in the kernel, a birth in the cokernel, and no change in the image, and this is correctly reported by the algorithms.

**Case 2.1.2** $R_g[i] = 0$. This is Case B in Table 3. The only change is a birth in the image and this is correctly reported by the algorithms.

**Case 2.2** $R_f[i] \neq 0$. By Observation (i), this implies $R_g[i] \neq 0$. This is Case E in Table 3. The only change is a death in the image which is correctly reported by the algorithms.

We conclude that the actual births and deaths are subsets of the computed births and deaths, and similarly that the actual pairs are a subset of the computed pairs.

We now have the containment relations in both directions which implies that corresponding sets of computed and actual births, deaths, and pairs are in fact the same. This concludes the proof that the algorithms in Section 3 correctly compute the persistence diagrams of the sequences of kernels, images, and cokernels.

**5 Applications**

In this section, we use kernels to measure local homology and images to approximate persistence diagrams of noisy functions specified on noisy domains. There are additional applications that are sufficiently straightforward that we can leave the details to the interested reader. One is the denoising of alpha-beta witness complexes as introduced in [1]; see also [9]. By considering maps from one complex to another, more tolerantly constructed complex we can preserve persistent features without accidentally introducing new ones. By varying the scale parameter, $\alpha$, we thus get a persistence diagram that is less noisy than the diagram of the sequence of complexes for fixed tolerance parameter $\beta$. Another application is the computation of rank invariants for a doubly-filtered space $\mathcal{X}_{i,k}$ as considered in [3]. Here we get a simple algorithm by encoding the ranks of the images of
The absolute homology of \( S_r \) expresses the homology within a fixed distance \( r \) of \( S \). We can compute all such diagrams in time quartic in the number of simplices in the triangulation of \( X \).

### 5.1 Local Homology

We begin with the application of kernels to measuring the local homology of a space in \( \mathbb{R}^n \) at a point not necessarily in the space. Following earlier work, we assume that the space is not known other then indirectly through a finite set of points sampled in or near the space.

**Measuring local homology.** Bendich et al. study the reconstruction of a stratified space \( \mathcal{S} \) from a finite point sample \( U \) in \( \mathbb{R}^n \) [2]. Specifically, they use persistence to define a multi-scale version of the local homology of \( \mathcal{S} \) at a point \( z \in \mathbb{R}^n \). Let \( S_\alpha \) be the sets of points at Euclidean distance at most \( \alpha \) from \( S \), \( S_0 \) the set of points at distance at least \( \alpha \) from \( S \), and \( B_r \) the closed ball of radius \( r \) centered at \( z \). They express the homology within a fixed distance \( r \) of \( z \) in terms of the persistence diagram of the sequence

\[
0 \rightarrow H(S_\alpha \cap B_r) \rightarrow \cdots \rightarrow H(B_r) \\
\rightarrow H(B_r, S_0 \cap B_r) \rightarrow \cdots \rightarrow 0,
\]

(1)

where \( \alpha \) first goes up, from 0 to \( \infty \), and then down, from \( \infty \) to 0. The first half of the sequence captures the development of the absolute homology of \( S_\alpha \) within \( B_r \), and it can be shown that the second half captures the development of the relative homology of \( S_\alpha \) within the pair \( (B_r, \partial B_r) \). We note that cycles that lie entirely inside the ball are captured twice, once in each half. Finally, Bendich et al. vary \( r \) from 0 to \( \infty \) and this way obtain a vineyard that expresses the local homology of \( S_\alpha \) within the neighborhood of the point \( z \); see Figure 6.

It is thus no longer necessary to include absolute homology classes in the measurement and the vineyard simplifies into a form that more closely reflects the shape of the space in the neighborhood of the point \( z \); see Figure 6. Relationships between the vineyard of \( S \) and that of a finite point sample \( U \subseteq \mathbb{R}^n \) similar to the Local Homology Inference and the Inverse LHI Theorems in [2] can be proved using the same methods. Details are omitted.

**Computing local homology.** Following the prior work, we use the Delaunay triangulations of \( U \) restricted to the ball \( B_r \) and to the ball without the interior of the power cell of the point \( z \), \( Z_0(r) = B_r - \operatorname{int} Z(r) \); see [2, Section 6] for details. We use distance functions and Delaunay triangulations for clear and solid theoretical footing. For practical high-dimensional implementations, like [2], these results need to be extended to Vietoris-Rips [12] and Witness complexes [9]. Let \( K_\alpha = \operatorname{Del}(U|U_\alpha \cap B_r) \) and \( L_\alpha = \operatorname{Del}(U|U_\alpha \cap Z_0(r)) \) be the Delaunay triangulations that are further restricted to the set of points at distance at most \( \alpha \) from \( U \). Let furthermore

\[
\begin{align*}
\iota_\alpha & : H(U_\alpha \cap Z_0(r)) \rightarrow H(U_\alpha \cap B_r) \\
\iota_\alpha & : H(L_\alpha) \rightarrow H(K_\alpha)
\end{align*}
\]

be the maps between the homology groups induced by inclusion. To justify the use of the restricted Delaunay triangulations, we need to show that the persistence diagrams of the kernels of these maps are the same. This requires that the following diagram is well-defined, commutative, and its
vertical maps are isomorphisms whenever $\alpha \leq \alpha'$:
\[
\begin{align*}
\ker j_\alpha & \to \ker j_{\alpha'} \\
\downarrow i_\alpha & \quad \downarrow i_{\alpha'} \\
\ker i_\alpha & \to \ker i_{\alpha'} \\
\uparrow h_\alpha & \quad \uparrow h_{\alpha'} \\
\ker l_\alpha & \to \ker l_{\alpha'}.
\end{align*}
\]

Consider the following diagram whose maps are all induced by inclusions except for the lower vertical maps which are motivated by the Nerve Subdivision Lemma [2, Section 8].
\[
\begin{align*}
H(U_\alpha \cap \partial B_r) & \xrightarrow{\rho} H(U_\alpha \cap B_r) \\
\downarrow & \quad \downarrow \\
H(U_\alpha \cap Z_0(r)) & \xrightarrow{\rho} H(U_\alpha \cap B_r) \\
\uparrow & \quad \uparrow \\
H(L_\alpha) & \to H(K_\alpha).
\end{align*}
\]
The top square commutes. The bottom square also commutes since the horizontal maps are induced by inclusion and the vertical map for $L_\alpha$ is the restriction of the one for $K_\alpha$. Therefore the kernel diagram is well-defined. Since the kernels are subgroups of the domains of their defining maps, the kernel diagram is a restriction of a diagram considered in [2, Section 8]. The analysis there implies that it commutes and its vertical maps are isomorphisms, as required.

The Stability Theorem of Section 2 implies that the kernel persistence diagrams change continuously with the radius of the restricted ball $B_r$. We construct the implied vineyard by maintaining the ordering of the simplices and the reduced matrices. However, unlike with relative homology in [2] we cannot use excision to maintain different orderings of a static complex. We need to be able to handle insertion of simplices into the ordering when the power cell of the point $z$ expands with $r$ to include new simplices. Fortunately, the new simplices are paired amongst each other so that these updates can be done in linear time per insertion. Once a simplex is inserted, its position in the ordering can be described by a continuous function; see Appendices A and B of [2]. It therefore suffices to maintain the decompositions in the four steps of the algorithm under transpositions of contiguous simplices. Details on how to perform these operations are given in Appendix A.

### 5.2 Noisy Domains

Persistent homology has proven to be well-suited for dealing with noisy functions. Indeed, the stability of persistence diagrams implies that the topological features of an unknown ideal function $\tilde{f} : X \to \mathbb{R}$ can be approximately recovered knowing only a noisy approximation $f$ of $\tilde{f}$. We claim that the persistence for images can be used to furthermore filter out the topological noise induced by the domain itself.

**Stability.** We assume an unknown ideal domain given as the zero sublevel set of the unknown ideal function $\tilde{h} : \mathbb{R}^n \to \mathbb{R}$, that is, $X = \tilde{h}^{-1}(-\infty, 0]$. On this domain we consider another unknown ideal function but because we will vary the domain we assume it is defined on the entire ambient space, $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$. Can we estimate the persistence diagram of the restriction $\tilde{f}|_{\tilde{X}} : \tilde{X} \to \mathbb{R}$ knowing only noisy approximations $h, f$ of $\tilde{h}, \tilde{f}$? We give an affirmative answer under mild requirements on the functions. To describe these requirements we use superscripts for sublevel sets of $h$ and $f$ and subscripts for sublevel sets of $\tilde{h}$ and $\tilde{f}$, that is, $X^u = h^{-1}(-\infty, a]$, $\tilde{X}^u = \tilde{h}^{-1}(-\infty, a]$, $X_a^u = X_a \cap X^u$ and similarly for $\tilde{X}$, $\tilde{X}_a$. Writing $\epsilon = \|\tilde{h} - h\|_{\infty}$ we require that $\tilde{h}$ is smooth and the norm of its gradient is bounded away from 0 where this is relevant, that is, $\|
abla \tilde{h}\| \geq \mu > 0$ on $X^{2\epsilon} - X^{-2\epsilon}$. Furthermore, we write $\delta = \|f - \tilde{f}\|_{\infty}$ and require that $f$ is Lipschitz with constant $\kappa$, that is, $|f(x) - f(y)| \leq \kappa \|x - y\|$ for all $x, y \in \mathbb{R}^n$.

We note that the requirement on $\tilde{h}$ implies a homotopy between the identity on $\tilde{X}^{2\epsilon}$ and a retraction $\rho$ from $\tilde{X}^{2\epsilon}$ to $\tilde{X}^\epsilon$. To construct the homotopy we consider the integral lines of the vector field $-\nabla h$ starting at points $x \in \tilde{X}^{2\epsilon}$. Let $g(x)$ be the first point on the curve starting at $x$ that satisfies $h(g(x)) = \tilde{h}(x) - 2\epsilon$ or $\tilde{h}(g(x)) \leq -2\epsilon$. In words, $g$ moves the fringe outside the boundary of $X$ to the fringe inside that boundary and it moves the fringe inside that boundary to the boundary of $\tilde{X}^{2\epsilon}$. Since the gradient has norm no smaller than $\mu > 0$ along the integral line, the retraction $\rho : \tilde{X}^{2\epsilon} \to \tilde{X}^{2\epsilon}$ is well defined. By construction, there is a homotopy between the identity on $\tilde{X}^{2\epsilon}$ and $g$ that moves points by at most $2\epsilon/\mu$. We are now ready to state our result.

**Stability Theorem for Noisy Domains.** Let $h, \tilde{h} : \mathbb{R}^n \to \mathbb{R}$ with $\epsilon = \|h - \tilde{h}\|_{\infty}$, $\tilde{h}$ smooth, and the norm of the gradient satisfying $\|
abla \tilde{h}\| \geq \mu > 0$ on $X^{2\epsilon} - X^{-2\epsilon}$. Furthermore, let $f, \tilde{f} : \mathbb{R}^n \to \mathbb{R}$ with $\delta = \|f - \tilde{f}\|_{\infty}$ and $f$ Lipschitz with constant $\kappa$. Then
\[
d_B(Dgm(\tilde{f}|_{\tilde{X}}), Dgm(im f_{|X^\epsilon \to \tilde{f}|_{\tilde{X}}}) \leq 2\kappa \epsilon/\mu + \delta
\]
provided the restrictions of $\tilde{f}$ to $\tilde{X}$ and of $f$ to $\tilde{X}$, $\tilde{X}^\epsilon$, $X^{-\epsilon}$ are continuous and tame and there exists a triangulation of $X^\epsilon$ in which $X^{-\epsilon}$ and $\tilde{X}$ arise as subcomplexes.

**Proof.** By the Stability Theorem for ordinary persistence we have $d_B(Dgm(\tilde{f}|_{\tilde{X}}), Dgm(\tilde{f}|_{\tilde{X}})) \leq \delta$. It remains to show that the bottleneck distance between $Dgm(f|_{X^\epsilon})$ and $Dgm(im f_{|X^\epsilon \to \tilde{f}|_{\tilde{X}}})$ is bounded from above by $c = 2\kappa \epsilon/\mu$. 

Writing \( F_a \) for \( H(\tilde{X}_a^0) \), the diagrams \( \text{Dgm}(f|_Y) \) is obtained from the sequence formed by the maps \( F_a \to F_b \) induced by the inclusion \( \tilde{X}_a^0 \subseteq \tilde{X}_b^0 \) for all \( a \leq b \). Similarly, writing \( J_a \) for the image of the map \( H(\tilde{X}_a^{-\varepsilon}) \to H(\tilde{X}_a^0) \), the diagram \( \text{Dgm}(\text{im } f|_{\tilde{X}_a^-} \to f|_{\tilde{X}_a^0}) \) is obtained from the sequence of maps \( J_a \to J_b \) again induced by inclusion and for all \( a \leq b \). To adapt the proof of stability given in [6], we need to connect these two sequences by maps \( \phi_a : F_{a-c} \to J_a \) and \( \psi_a : J_{a-c} \to F_a \), for all \( a \in \mathbb{R} \), in such a way that the diagram formed by the two sequences together with the new maps commutes. We construct the new maps using the homotopy between the identity and the retraction \( \varrho : \tilde{X}^2 \to \tilde{X}^2 \). As mentioned earlier, the homotopy moves a point by at most \( 2\varepsilon/\mu \) and since \( f \) is Lipschitz with constant \( \kappa \), \( \varrho \) maps \( \tilde{X}_a^{-2\varepsilon} \) to \( \tilde{X}_a^{-\varepsilon} \) which is included in \( \tilde{X}_a^{-\varepsilon} \). The induced homomorphism from \( H(\tilde{X}_a^{-c}) \) to \( H(\tilde{X}_a^{-2\varepsilon}) \) composed with the induced homomorphism from \( H(\tilde{X}_a^{-\varepsilon}) \) to \( H(\tilde{X}_a^0) \) gives the map \( \phi_a \). As shown in Figure 7, \( \phi_a \) connects the two sequences with a shift of \( c = 2\varepsilon/\mu \). By a similar process, we construct the map \( \psi_a \) connecting the two sequences of vector spaces in the other direction and again with a shift of \( c \). Because \( \varrho \) is homotopic to the identity, the diagram formed by the two sequences and the maps \( \phi_a \) and \( \psi_a \) commutes. The remainder of the proof can be adapted directly from [6].

**Diagram approximation.** An interesting case of the above theorem arises when we consider a finite set of points, \( U \), sampling an unknown shape, \( S \subseteq \mathbb{R}^n \). Let \( h : \mathbb{R}^n \to \mathbb{R} \) be the distance function of \( S \), that is, \( h(x) = \inf_{y \in S} \| x - y \| \). Similarly, let \( h : \mathbb{R}^n \to \mathbb{R} \) be the distance function of \( U \) and set \( \varepsilon \) to the Hausdorff distance between \( S \) and \( U \). For technical reasons we may have to replace \( \tilde{h} \) by a smooth approximation, for example obtained by convolution with an infinitesimally narrow Gaussian so that the assumptions in the theorem are satisfied. In this setting, the requirement that the norm of the gradient of \( \tilde{h} \) is bounded from below by \( \mu \) is equivalent to the \( \mu \)-reach of \( S \) exceeding \( 4\varepsilon \). Here we recall that the \( \mu \)-reach as recently introduced in [4] is a notion of feature size that permits the treatment of non-smooth objects. Under this assumption, the Stability Theorem for Noisy Domains implies that it is possible to estimate the persistence diagram of a function \( f \) restricted to the \( 2\varepsilon \)-offset of \( S \) knowing only a Lipschitz function \( f \) that approximates \( f \) and the point set \( U \) that samples \( S \).

To approximate the persistence diagram of \( f \) restricted to \( S \) itself, we exploit the existence of an isotopy \( \tilde{f} \) from \( \mathbb{S}^{2\varepsilon} \) to an arbitrarily small offset \( \mathbb{S}^{0} \) of the shape such that the points move by less than \( 2\varepsilon/\mu \) during the deformation. The construction of \( \tilde{f} \) is similar to the construction of the homotopy between the identity and \( g \) described above. The isotopy implies that \( \text{Dgm}(f|_{\mathbb{S}^{0}}) \) equals \( \text{Dgm}(f \circ \iota^{-1}|_{\mathbb{S}^{0}}) \) which in turn is \( \varepsilon \)-close to \( \text{Dgm}(f|_{\mathbb{S}^{0}}) \) by stability. Hence, the latter diagram can also be estimated from \( f \) and \( U \) with an accuracy of \( 2\varepsilon \). Perhaps surprisingly, \( \text{Dgm}(f|_{\mathbb{S}^{0}}) \) may not converge to \( \text{Dgm}(f|_{\mathbb{S}^{0}}) \) as \( \eta \) goes to 0; see [5] for examples of shapes \( S \) that lack this convergence property. However, convergence holds for sufficiently regular spaces \( S \), such as smooth submanifolds or geometrically realized simplicial complexes. In these cases we can estimate \( \text{Dgm}(f|_{\mathbb{S}^{0}}) \) with precision \( 2\varepsilon \).

In the more realistic case in which \( f \) is only known at a finite set of points, \( U \), a valid approach replaces \( f \) by the function \( \tilde{f} \) that is constant on the Voronoi cells of the points and coincides with \( f \) on \( U \). While \( f \) is not continuous, it is almost everywhere continuous in a way that does not disrupt the proof of stability. Furthermore, \( f \) and \( \tilde{f} \) differ by at most \( 4\varepsilon \) on \( \tilde{h}^{-1}(-\infty, 4\varepsilon] \). By the Stability Theorem for Noisy Domains, the persistence diagrams of the images defined by \( f \) and \( \tilde{f} \) are close. The diagram for \( \tilde{f} \) can be computed using the alpha shape filtration of \( U \). We thus get a practical algorithm for estimating the persistence diagram of functions given only at a finite set of points.

**6 Discussion**

In this paper, we consider persistent homology for sequences of kernels, images, and cokernels defined by a pair of topological spaces, \( Y \subseteq X \), and two functions, \( f : X \to \mathbb{R} \) and \( g : Y \to \mathbb{R} \), with \( f(y) \leq g(y) \) for every \( y \in Y \). Since \( g \) majorizes the restriction of \( f \) to \( Y \), its sublevel sets are contained in those of \( f \), \( [g^{-1}(-\infty, a)] \subseteq [f^{-1}(-\infty, a)] \), and we have homomorphisms \( j_a : H(Y_a) \to H(X_a) \) induced by the inclusions. To see that persistent homology is well defined we just need to note that the diagrams

\[
\begin{align*}
H(X_a) &\to H(X_b) \\
\uparrow j_a &\downarrow j_b \\
H(Y_a) &\to H(Y_b)
\end{align*}
\]

commute for any \( a \leq b \) and thus induce homomorphic maps \( \ker j_a \to \ker j_b \), \( \text{im} j_a \to \text{im} j_b \), and \( \text{cok} j_a \to \text{cok} j_b \).
It is worth noting that the mapping cylinder construction described in Section 2 can be used to extend the framework from inclusion $Y \subseteq X$ to an arbitrary continuous map $j : Y \rightarrow X$. If we have two functions $f : X \rightarrow R$, and $g : Y \rightarrow R$, such that $j(Y_a) \subseteq X_a$ with the sublevel set of each space taken with respect to its own map, then the maps $j_a : Y_a \rightarrow X_a$ between the sublevel sets induce homomorphisms on homology groups just the same. Constructing a mapping cylinder $X' = X \cup Y \times [0,1]$ by identifying $(y,0) \in Y \times \{0\}$ with $j(y) \in X$, it is easy to verify that the sequences of kernels, images, and cokernels induced by inclusion $Y_a = Y_a \times \{1\} \subseteq X_a$ give the same persistence pairing as the three sequences induced by the continuous maps $j_a$ defined above.

One can also extend the framework described in this paper from spaces $Y_a \subseteq X_a$ to pairs of spaces $(Y_a, Y_{a_0}) \subseteq (X_a, X_{a_0})$ using the cone construction exploited for the computation of extended persistence [7]. Indeed, observing that the relative homology groups $H(Y_a, Y_{a_0})$ are isomorphic to the homology groups of $Y_a$ with a cone on $Y_{a_0}$ rel the cone point, i.e. $H(Y_a, Y_{a_0}) \cong H(Y_a \cup CY_{a_0}, \omega)$, we can compute the persistence of kernels, images, and cokernels induced by the inclusions of pairs of spaces by using the algorithms of this paper on corresponding cones.

The mapping cylinder construction and the cone construction can be combined to cope with arbitrary continuous maps between pairs of spaces rather than only inclusions.

The algorithms for computing the persistence diagrams of the sequences of kernels, images, and cokernels are variants of the classic Smith normal form algorithm; see [11, Chapter 1.11]. More directly, we build on the algorithm in [8] which reduces the entire incidence matrix at once, paying careful attention to the orderings of the rows and the columns which are the same and consistent with the ordering of the sublevel sets. The main difficulty in the new setting is that we have two orderings, one for $K$ triangulating $X$ and the other for $L \subseteq K$ triangulating $Y$. We cope using the mapping cylinder construction and matrices in which the rows are reordered so that $L$ precedes $K - L$. The resulting algorithms run in time at most cubic in the size of $K$, same as the reduction algorithm in [8] as well as the classic Smith normal form algorithm for modulo-2 arithmetic. The fact that a simple reordering of the rows does the trick suggests that there may be other interesting pieces of information that can be extracted from reduced reordered matrices. Does the pairing defined by the lowest ones in a reduced incidence matrix in which columns and rows are ordered independently and arbitrarily have an intuitive interpretation that carries topological meaning?

References


Appendix A

We consider the actions necessary to maintain the matrix decompositions under the transposition of two contiguous simplices $\sigma_i$ and $\sigma_{i+1}$. The maintenance of the $R = DU$ decomposition under such transpositions has been considered in [8] and their algorithm applies directly to the maintenance of $R_f = D_f V_f$ and $R_g = D_g V_g$. However, maintaining the other three decompositions is more difficult because $D_{bet}$, $D_{mor}$, and $D_{vuk}$ are not ordinary incidence matrices. The algorithm in [8] expresses an update in terms of pre- and post-multiplications by idempotent matrices. We observe that multiplying $V$ by the same matrices as $R$ maintains the equality $R = DV$ as well as $VU = I$. 
Step 2, the image. Consider first the maintenance of \( R_{\text{im}} = D_{\text{im}} V_{\text{im}} \) which is made difficult by ordering the rows and columns differently. We distinguish the case in which \( \sigma_i \) and \( \sigma_{i+1} \) both belong to \( L \) or to \( K - L \) from the case in which one belongs to \( L \) and the other to \( K - L \). In the former case, we let \( P' \) be the transposition matrix of the rows \( \sigma_i \) and \( \sigma_{i+1} \), and in the latter case we set \( P' = I \). Now we can follow the case analysis in [8] replacing the pre-multiplication of \( D_{\text{im}} \) and \( R_{\text{im}} \) by \( P \) with \( P'' \). The only other adjustment that we must make is in Case 1 in [8], in which both simplices \( \sigma_i \) and \( \sigma_{i+1} \) do not belong to the same group, since in this case \( P' = I \) and no rows transpose.

Step 3, the kernel. Consider second the maintenance of \( R_{\text{ker}} = D_{\text{ker}} V_{\text{ker}} \). Recall that the matrix \( D_{\text{ker}} \) is obtained from \( V_{\text{im}} \) by keeping only the columns of positive simplices and reordering the rows. If both transposing simplices are positive in \( R_{\text{ker}} \) then their columns in \( V_{\text{ker}} \) do not change. Therefore, their columns in \( D_{\text{ker}} \) and in \( R_{\text{ker}} \) transpose as well as both their rows and columns in \( V_{\text{ker}} \). If as a result \( V_{\text{ker}} \) ceases to be upper-triangular then we can perform an update similar to Case 2.1 of [8] by adding the column of \( \sigma_i \) to the column of \( \sigma_{i+1} \) before the transposition. Using the matrix \( S_i^{i+1} \) for this operation, we get

\[
P' R_{\text{ker}} S_i^{i+1} P = (P' D_{\text{ker}} P)(PV_{\text{ker}} S_i^{i+1} P).
\]

Observe that this operation may render \( R_{\text{ker}} \) non-reduced which can also happen as a result of a row transposition if \( \sigma_i \) and \( \sigma_{i+1} \) both belong to \( L \) or to \( K - L \). If \( R_{\text{ker}} \) becomes non-reduced we can fix it by multiplying again by \( S_i^{i+1} \) which gives

\[
P' R_{\text{ker}} S_i^{i+1} P S_i^{i+1} = (P' D_{\text{ker}} P)(PV_{\text{ker}} S_i^{i+1} P S_i^{i+1}).
\]

This update may result in a Type 2 switch in the pairing [8], that is, both transposing simplices are responsible for deaths of homology classes. If the two transposing simplices are both negative then there are no columns exchanges in \( D_{\text{ker}} \), \( R_{\text{ker}} \), and \( V_{\text{ker}} \). However, the rows of \( D_{\text{ker}} \) and \( R_{\text{ker}} \) exchange if both \( \sigma_i \) and \( \sigma_{i+1} \) are in \( L \) or in \( K - L \). The former case is easy because the transposing simplices do not contain the lowest one in their rows. However, in the latter case, similarly to Case 1.1 of [8], the matrix \( R_{\text{ker}} \) may become non-reduced. Denote by \( \sigma_i \) and \( \sigma_k \) (both in \( L \)) the simplices in \( R_{\text{ker}} \) that have in their columns the lowest ones in the positions of simplices \( \sigma_i \) and \( \sigma_{i+1} \), respectively. We can reduce \( R_{\text{ker}} \) by adding the preceding column to the succeeding column. We thus obtain

\[
P' R_{\text{ker}} S_i^k = (P' D_{\text{ker}})(V_{\text{ker}} S_i^k) \quad \text{or}
\]

\[
P' R_{\text{ker}} S_k^i = (P' D_{\text{ker}})(V_{\text{ker}} S_k^i).
\]

If this update is necessary and \( k \) precedes \( l \) then we have a Type 1 switch in the pairing [8], that is, both transposing simplices are responsible for births of homology classes. If \( \sigma_i \) is negative and \( \sigma_{i+1} \) is positive in \( R_{\text{ker}} \), that is, we are in Case 3 of [8], then regardless of whether the pairing switches or not, the columns of \( D_{\text{ker}} \) remain the same. This observation follows from the update rule in case the pairing switches. This leaves the column of the negative simplex to be the sum of the transposing columns while leaving the column of the positive simplex intact. The rows of \( D_{\text{ker}} \) transpose only if \( \sigma_i \) and \( \sigma_{i+1} \) are both negative and \( S_{\text{ker}} \) is expected since the pairs in the kernel (image persistence is unaffected by Step 3) are always between simplices of the same dimension.

Step 4, the cokernel. Consider finally the maintenance of \( R_{\text{cok}} = D_{\text{cok}} V_{\text{cok}} \). For reasons that will become apparent shortly, we switch to maintaining the decomposition

\[
D_{\text{cok}} = R_{\text{cok}} U_{\text{cok}}, \quad \text{where } U_{\text{cok}} = V_{\text{cok}}^{-1}.
\]

We might as well since the matrix \( V_{\text{cok}} \) plays no role in the construction of the persistence diagrams. It is a curious property of the algorithm in [8] maintaining \( D = RU \) (or equivalently \( R = DV \)) that the only time the columns of \( V \) corresponding to positive simplices can change is in the preprocessing step of Case 1, namely when \( \sigma_i \) and \( \sigma_{i+1} \) are both positive in \( R \) and \( V[i, i+1] = 1 \). Applied to the decomposition of \( D_{\text{ker}} \), we add column \( V_{\text{cok}}[i] \) to \( V_{\text{cok}}[i+1] \) whenever this happens. To offset the resulting change in \( D_{\text{cok}} \), we add the row of \( \sigma_{i+1} \) to the row of \( \sigma_i \) in \( U_{\text{cok}} \), noting that the two rows are not necessarily adjacent. We can update \( U_{\text{cok}} \) in linear time while the equivalent fix to \( V_{\text{cok}} \) would require a quadratic number of operations. The only remaining changes to \( D_{\text{cok}} \) are
transpositions of rows and columns which we handle directly using the algorithm in [8]. It follows that the maintenance of the decomposition $D_{cok} = R_{cok} U_{cok}$ takes linear time per operation.