Discrete Comput Geom (2013) 50:330–353
DOI 10.1007/s00454-013-9529-6

Morse Theory for Filtrations and Efficient Computation of Persistent Homology

Konstantin Mischaikow · Vidit Nanda

Received: 21 December 2011 / Revised: 11 March 2013 / Accepted: 27 June 2013 /
Published online: 27 July 2013

Abstract We introduce an efficient preprocessing algorithm to reduce the number of cells in a filtered cell complex while preserving its persistent homology groups. The technique is based on an extension of combinatorial Morse theory from complexes to filtrations.

Keywords Computational topology · Discrete Morse theory · Persistent homology

1 Introduction

The use of topological methods for data analysis is rapidly growing and persistent homology is proving to be one of the more successful techniques [1,6,7,10]. Three fundamental properties account for the importance of persistent homology: (i) being based on algebraic topology, it provides a well understood codification of potentially complicated and/or high dimensional geometric information, (ii) the information it provides is stable with respect to perturbations [3], and (iii) it is readily computable [4,8,24,34,35]. Our focus is on this last point.

The most common algorithm used for computing persistent homology is presented in [35] wherein it is remarked that the worst case complexity is of the same order in time and space as that of computing homology. Subsequent work [21] has reduced

K. Mischaikow
Department of Mathematics and BioMaPS, Rutgers, The State University of New Jersey,
110 Frelinghuysen Rd, Piscataway, NJ 08854-8019, USA
e-mail: mischaik@math.rutgers.edu

V. Nanda (✉)
Department of Mathematics, The University of Pennsylvania, 209 South 33rd Street,
Philadelphia, PA 19104, USA
e-mail: vnanda@sas.upenn.edu

this complexity to $\mathrm{O}(n^{\omega})$, where ω is the matrix multiplication exponent. To the best of our knowledge, there are no known implementations of fast matrix multiplication besides Strassen's algorithm [31], which has an exponent $\omega_S \sim 2.8$.

While our work is strongly motivated by the usefulness of persistent homology, there are instances in which one is interested in computing general homology groups of filtered complexes. The optimal worst case analysis for current homology algorithms appears to be super-quartic over general principal ideal domains [12] and roughly cubical in practice over the integers [5,28,30] with respect to the size of the input complex. For massive datasets, this can be a severe limitation. Since we know of no way to improve the worst case complexity of the problem, the strategy we propose in this paper is to use ideas from combinatorial Morse theory [9] to reduce the initial complex using geometric and combinatorial methods before applying the algorithms of [35]. This reduction preserves all homological information in general and persistent homology groups in particular.

For a heuristic understanding of our approach, consider a *complex* $\mathcal{X}$ (a precise definition is given in Sect. 2) with a finite nested sequence of subcomplexes $\mathcal{X}^0 \subset \mathcal{X}^1 \subset \cdots \subset \mathcal{X}^K = \mathcal{X}$. We refer to this structure as a *filtration* $\mathcal{F}$ of $\mathcal{X}$. The inclusions canonically induce maps $i_*^k\colon\ \mathrm{H}_*(\mathcal{X}^k) \to \mathrm{H}_*(\mathcal{X}^{k+1})$ on the homology groups. For each number $p \geq k$, let $i_*^{k,p}: \mathrm{H}_*(\mathcal{X}^k) \to \mathrm{H}_*(\mathcal{X}^p)$ denote the composition $i_*^{p-1} \circ \cdots \circ i_*^{k+1} \circ i_*^k$.

When working over field coefficients, it is possible to simultaneously choose bases of $\mathrm{H}_*(\mathcal{X}^k)$, $k = 0, \ldots, K$ such that for each basis element $\alpha \in \mathrm{H}_q(\mathcal{X}^k)$ there exists a well-defined pair of integers $b_\alpha \leq k$ and $d_\alpha \geq k+1$ satisfying the following properties: b_α is the smallest integer ℓ so that $\alpha \in i_q^{\ell,k}(\mathrm{H}_q(\mathcal{X}^\ell))$ and d_α is the largest integer ℓ with $i_q^{k,\ell-1}(\alpha) \neq 0$. The pair (b_α, d_α) indicates the "birth" and "death" of the topological feature identified by α. Observe that if β is the element of $\mathrm{H}_q(\mathcal{X}^p)$ satisfying $\beta = i^{k,p}(\alpha)$, then $(b_\alpha, d_\alpha) = (b_\beta, d_\beta)$ and hence we identify these intervals. The collection of these equivalence classes of pairs ranging over all $\alpha \in \mathrm{H}_q(\mathcal{X}^k)$ for $k = 0, \ldots, K$ produces the qth *persistence diagram* for the filtration $\mathcal{F}$.

The complexity of computing the persistence diagram of a filtration $\mathcal{F}$ is essentially determined by the complexes $\{\mathcal{X}^k\}$. Thus, a natural approach for reducing the computational cost is to perform an efficient preprocessing step that constructs an alternate filtration $\mathcal{F}'$ consisting of significantly smaller complexes $\{\mathcal{X}'^k\}$ which has the same persistence diagram as $\mathcal{F}$. The same strategy has been adopted to compute homology of an unfiltered complex [14,19]. Again, we provide a heuristic description of this technique.

In the classical setting, the Morse homology of smooth manifolds is defined in terms of a complex where the chains are generated by critical points of a smooth functional and the boundary operator is determined by heteroclinic orbits generated by the gradient flow of the functional. In the combinatorial setting, the gradient flow is replaced by a partial pairing on cells in the complex. The unpaired cells generate the chains in the Morse complex, while the boundary operator is defined via paths in the cell complex generated by the pairing. The preprocessing algorithm of [14] is based in part on the coreduction algorithm of [23] and provides an efficient means for constructing a partial pairing on a given cell complex. As is demonstrated in [13], for many complexes the resulting Morse complex is many orders of magnitude smaller than the original. A contribution of this paper is an algorithm that takes a filtration

$\mathcal{F}$ and produces a new, typically much smaller, filtration $\mathcal{M}$ such that the persistence diagram of $\mathcal{M}$ agrees with that of $\mathcal{F}$. In Sect. 6 we show that for many examples this preprocessing step has the advantage of significantly reducing both the computational time and the required memory for running the persistence algorithm [35].

An outline of this paper is as follows. Section 2 recalls the fundamental ideas and constructions related to complexes, persistent homology and combinatorial Morse theory. Section 3 provides a categorical construction that allows us to relate the persistence homology groups of different filtrations. Using this language we prove Theorem 4.3 in Sect. 4, which establishes that the persistent homology of a given filtration is equivalent to the persistent homology of an associated Morse filtration. Section 5 contains the preprocessing algorithm `MorseReduce` along with our main result which demonstrates that the output of `MorseReduce` is a filtration with the same persistent homology as the input filtration. Finally, Sect. 6 presents experimental results derived from applying our preprocessing algorithm to a variety of filtrations based on different types of complexes.

2 Background

In this section we provide a brief review, primarily to establish notation of complexes, persistent homology and combinatorial Morse theory.

2.1 Complexes

As indicated in the Introduction, our interest in computing persistent homology is motivated by data analysis. For these problems typically one does not have an a priori understanding of the structure of the underlying space and thus one works with abstract complexes which may or may not correspond to a geometric or topological realization. With this in mind, we use a rather general notion of complex that dates back to Tucker [32] and Lefschetz [18]. Throughout this paper, **R** denotes a principal ideal domain (PID) whose invertible elements will be called *units*.

Definition 2.1 Consider a finite graded set $\mathcal{X} = \bigsqcup_{q\in\mathbb{Z}} \mathcal{X}_q$ along with a function $\kappa\colon \mathcal{X} \times \mathcal{X} \to \mathbf{R}$ and denote $\xi \in \mathcal{X}_q$ by $\dim \xi = q$. Then $(\mathcal{X}, \kappa)$ is a *complex* if the following properties hold.

(i) For each ξ and ξ' in $\mathcal{X}$,

$$\kappa(\xi, \xi') \neq 0 \quad \text{implies} \quad \dim \xi = \dim \xi' + 1. \tag{1}$$

(ii) For each ξ and ξ'' in $\mathcal{X}$,

$$\sum_{\xi'\in\mathcal{X}} \kappa(\xi, \xi') \cdot \kappa(\xi', \xi'') = 0. \tag{2}$$

An element $\xi \in \mathcal{X}$ is called a *cell* and $\dim \xi$ is called the *dimension* of ξ. The function κ is called the *incidence function* for the complex $(\mathcal{X}, \kappa)$. We denote $(\mathcal{X}, \kappa)$

simply by $\mathcal{X}$ when there is no possible confusion about the underlying incidence function. The *face partial order* $\preccurlyeq$ is induced on the elements of $\mathcal{X}$ by the transitive closure of the generating relation $\prec$ given as follows. For $\xi, \xi' \in \mathcal{X}$

$$\xi' \prec \xi \quad \text{if} \quad \kappa(\xi, \xi') \neq 0.$$

By (1), the function $\dim : \mathcal{X} \to \mathbb{Z}$ is an order-preserving map.

Consider $\mathcal{X}' \subset \mathcal{X}$ and note that the restriction of κ to $\mathcal{X}' \times \mathcal{X}'$ satisfies (1). If for each $\eta \in \mathcal{X}'$ the set $\{\xi \in \mathcal{X} \mid \xi \prec \eta\}$ is contained in $\mathcal{X}'$, then we say that $\mathcal{X}'$ satisfies the *subcomplex property* and call $\mathcal{X}'$ a *subcomplex* of $(\mathcal{X}, \kappa)$. Note that Eq. (2) is automatically satisfied for a subcomplex $\mathcal{X}'$, and so $(\mathcal{X}', \kappa)$ is a complex in its own right.

Given a complex $(\mathcal{X}, \kappa)$ the associated *chain complex* consists of the free modules $C_q(\mathcal{X}) := \mathbf{R}(\mathcal{X}_q)$, where the basis elements are identified with the q-dimensional cells $\xi \in \mathcal{X}_q$, and the *boundary operator* $\partial_q : C_q(\mathcal{X}) \to C_{q-1}(\mathcal{X})$ is generated by

$$\partial_q \xi := \sum_{\xi' \in \mathcal{X}} \kappa(\xi, \xi')\xi'.$$

The *q-cycles* and *q-boundaries* are defined to be the submodules $Z_q(\mathcal{X}) := \ker \partial_q$ and $B_q(\mathcal{X}) := \operatorname{im} \partial_{q+1}$ of $C_q(\mathcal{X})$, respectively and the *homology groups* are given by the quotient module

$$\mathrm{H}_q(\mathcal{X}) := \frac{Z_q(\mathcal{X})}{B_q(\mathcal{X})}.$$

Let $\phi_*, \psi_* : C_*(\mathcal{X}) \to C_*(\mathcal{X}')$ be chain maps. Recall that a collection of module morphisms $\Theta = \{\Theta_q : C_q(\mathcal{X}) \to C_{q+1}(\mathcal{X}')\}$ is a *chain homotopy between ϕ_* and ψ_** if

$$\Theta_{q-1} \circ \partial_q + \partial'_{q+1} \circ \Theta_q \equiv \phi_q - \psi_q$$

on $C_q(\mathcal{X})$. It is a standard result that if ϕ_* and ψ_* are chain homotopic—that is, if there exists a chain homotopy between them—then they induce the same homomorphism on homology groups. Two chain maps $\phi_* : C_*(\mathcal{X}) \to C_*(\mathcal{X}')$ and $\psi_* : C_*(\mathcal{X}') \to C_*(\mathcal{X})$ are *chain equivalences* if $\phi_* \circ \psi_*$ is chain homotopic to the identity map on $C_*(\mathcal{X}')$ and $\psi_* \circ \phi_*$ is chain homotopic to the identity map on $C_*(\mathcal{X})$. Observe that this implies that the induced maps $\phi_* : \mathrm{H}_*(\mathcal{X}) \to \mathrm{H}_*(\mathcal{X}')$ and $\psi_* : \mathrm{H}_*(\mathcal{X}') \to \mathrm{H}_*(\mathcal{X})$ are mutual inverses and hence isomorphisms. For a detailed discussion and proofs see [29, Chap. 4] and [25, Chap. 1.13].

2.2 Persistent Homology

Recall that $\mathcal{F} = \{\mathcal{X}^k \mid k = 1, \ldots K\}$ is a *filtration* of the complex $\mathcal{X}$ if for all k, $\mathcal{X}^k$ is a subcomplex of $\mathcal{X}$ and $\mathcal{X}^k \subset \mathcal{X}^{k+1}$. The individual complex $\mathcal{X}^k$ is referred to as the kth *frame* of the filtration. Let $i^{p,k}\colon \mathcal{X}^k \hookrightarrow \mathcal{X}^{k+p}$ denote the inclusion map. This induces a natural map on the chain complex, which we also denote by $i_*^{p,k}\colon C_*(\mathcal{X}^k) \to C_*(\mathcal{X}^{k+p})$.

The p-persistent qth homology group of the kth frame $\mathcal{X}^k$ is defined to be

$$\mathrm{H}_q^p(\mathcal{X}^k) = \frac{i_q^{p,k}(Z_q(\mathcal{X}^k))}{i_q^{p,k}(Z_q(\mathcal{X}^k)) \cap B_q(\mathcal{X}^{k+p})} \tag{3}$$

A discussion of the direct computation of these groups for all p, k, and q can be found in [35] along with an explicit algorithm based on the Smith normal form in the case when $\mathbf{R}$ is a field.

2.3 Combinatorial Morse Theory

Let $(\mathcal{X}, \kappa)$ be a complex over the PID $\mathbf{R}$ and denote by $\prec$ the generating relation of the face partial order $\preccurlyeq$ on $\mathcal{X}$.

Definition 2.2 A *partial matching* of $(\mathcal{X}, \kappa)$ consists of a partition of $\mathcal{X}$ into three sets $\mathcal{A}$, $\mathcal{K}$ and $\mathcal{Q}$ along with a bijection $w\colon \mathcal{Q} \to \mathcal{K}$, such that for each $Q \in \mathcal{Q}$ the incidence $\kappa(w(Q), Q)$ is a unit in $\mathbf{R}$. We denote this matching by $(\mathcal{A}, w\colon \mathcal{Q} \to \mathcal{K})$.

Observe that by Definition 2.1(i) and the unit incidence requirement, we have $\dim w(Q) = \dim Q + 1$ and $Q \prec w(Q)$ for each $Q \in \mathcal{Q}$. Given a partial matching of $\mathcal{X}$, define a relation $\ll$ on $\mathcal{Q}$ by the transitive closure of the generating relation $\lhd$ given below. For distinct elements $Q, Q' \in \mathcal{Q}$,

$$Q' \lhd Q \quad \text{if} \quad Q' \prec w(Q). \tag{4}$$

If $\ll$ is a partial order on $\mathcal{Q}$ then $(\mathcal{A}, w\colon \mathcal{Q} \to \mathcal{K})$ is called an *acyclic* matching of $\mathcal{X}$. In this paper, we are only interested in those matchings which are acyclic.

Remark 2.3 The definition of an acyclic matching $(\mathcal{A}, w\colon \mathcal{Q} \to \mathcal{K})$ is clearly related to earlier presentations of combinatorial Morse theory. See for example the work of Forman [9], Chari [2], and in particular Kozlov [16,17]. Elements of $\mathcal{A}$ are typically referred to as *critical cells* in analogy to classical Morse theory. The paired elements in $\mathcal{K}$ and $\mathcal{Q}$ are often not explicitly labelled since from a purely Morse theoretic perspective they are unimportant objects; it is only the pairing w that plays an essential role. However, our interest is in using combinatorial Morse theory to develop algorithms that are designed to be applied to complexes arising from experimental or numerical data sets. In particular, as is explained in Sect. 5 we often iteratively apply the preprocessing algorithm of this paper to the resulting Morse complex. This has no analogue in the classical Morse theory and in particular the critical cells of one complex cease to be critical cells in the next iterate of the algorithm. Similarly, in some applications (e.g. computing induced maps on homology [14]) it is essential to be able to recover homology generators in the original complex. For this we need to keep track of the paired cells and find it useful to have different labels for the different elements of the pairing.

Given an acyclic matching $(\mathcal{A}, w : \mathcal{Q} \to \mathcal{K})$ of $\mathcal{X}$, a *gradient path* is a non-empty sequence of cells $\rho = \big(Q_1, w(Q_1), \ldots, Q_M, w(Q_M)\big)$ with $Q_m \in \mathcal{Q}$ such that $Q_m \neq Q_{m+1} \lhd Q_m$ for each m. Thus, successive elements from $\mathcal{Q}$ in a given gradient path are strictly monotonically decreasing with respect to the partial order $\ll$ and consequently no such path can be a cycle. The initial cell Q_1 of ρ is denoted by $\mathbf{q}_\rho \in \mathcal{Q}$ and the final cell $w(Q_M)$ by $\mathbf{k}_\rho \in \mathcal{K}$. The *index* $\nu(\rho)$ of ρ is defined as the following element of $\mathbf{R}$:

$$\nu(\rho) := \frac{\prod_{m=1}^{M-1} \kappa(w(Q_m), Q_{m+1})}{\prod_{m=1}^{M} -\kappa(w(Q_m), Q_m)}. \tag{5}$$

Given cells A and A' in $\mathcal{A}$, a gradient path ρ is a *connection* from A to A' if $\mathbf{q}_\rho \prec A$ and $A' \prec \mathbf{k}_\rho$. This relationship is denoted by $A \overset{\rho}{\rightsquigarrow} A'$. The *multiplicity* of this connection ρ is defined to be

$$m(\rho) := \kappa(A, \mathbf{q}_\rho) \cdot \nu(\rho) \cdot \kappa(\mathbf{k}_\rho, A'). \tag{6}$$

Define a new map $\widetilde{\kappa} : \mathcal{A} \times \mathcal{A} \to \mathbf{R}$ by the relation

$$\widetilde{\kappa}(A, A') = \kappa(A, A') + \sum_{A \overset{\rho}{\rightsquigarrow} A'} m(\rho), \tag{7}$$

where the sum is taken over all connections ρ from A to A' and equals 0 if no such connections exist.

Theorem 2.4 *Let $(\mathcal{X}, \kappa)$ be a complex over a principal ideal domain* $\mathbf{R}$. *Consider a fixed acyclic matching $(\mathcal{A}, w\colon \mathcal{Q} \to \mathcal{K})$ of $\mathcal{X}$ and let $\mathcal{A}_q := \mathcal{A} \cap \mathcal{X}_q$. Then $(\mathcal{A}, \widetilde{\kappa})$ is also a complex over* $\mathbf{R}$, *where $\mathcal{A} = \bigsqcup_{q \in \mathbb{Z}} \mathcal{A}_q$ and $\widetilde{\kappa} : \mathcal{A} \times \mathcal{A} \to \mathbf{R}$ is defined by (7). Furthermore,*

$$\mathrm{H}_*(\mathcal{X}) \cong \mathrm{H}_*(\mathcal{A}).$$

The complex $(\mathcal{A}, \widetilde{\kappa})$ is called the *Morse complex* associated to the acyclic matching $(\mathcal{A}, w\colon \mathcal{Q} \to \mathcal{K})$ of $\mathcal{X}$ and $\widetilde{\kappa}$ is called the associated *Morse incidence function*. Theorem 2.4 follows from the work of Forman [9] and has been re-proven in a variety of contexts [2,14,16,17]. For the purpose of obtaining Theorem 4.3 we have adopted a slightly different presentation. Thus, to introduce the necessary notation we conclude this section with a terse outline of the proof which is obtained inductively via the following reduction step.

Let $(\mathcal{X}, \kappa)$ be a complex with an acyclic matching $(\mathcal{A}, w\colon \mathcal{Q} \to \mathcal{K})$. Given $Q \in \mathcal{Q}$, define $\mathcal{X}_Q \subset \mathcal{X}$ by

$$\mathcal{X}_Q := \mathcal{X} \setminus \{Q, w(Q)\}$$

and the function $\kappa_Q \colon \mathcal{X}_Q \times \mathcal{X}_Q \to \mathbf{R}$ by

$$\kappa_Q(\eta, \xi) = \kappa(\eta, \xi) - \frac{\kappa(\eta, Q) \cdot \kappa(w(Q), \xi)}{\kappa(w(Q), Q)} \tag{8}$$

A direct computation shows that κ_Q is an incidence function, and so $(\mathcal{X}_Q, \kappa_Q)$ is a complex. In fact, we may view the construction of κ_Q from κ as a sequence of row and column operations on the matrix representation of the boundary operator ∂ which make the unit incidence of Q and $w(Q)$ a pivot.

Observe that the acyclic matching $(\mathcal{A}, w \colon \mathcal{Q} \to \mathcal{K})$ on $\mathcal{X}$ induces an acyclic matching on $\mathcal{X}_Q$ of the form $(\mathcal{A}, w_Q \colon \mathcal{Q}_Q \to \mathcal{K}_Q)$ where $\mathcal{Q}_Q = \mathcal{Q} \setminus \{Q\}$, $\mathcal{K}_Q = \mathcal{K} \setminus \{w(Q)\}$, and $w_Q = w\,|_{\mathcal{Q}_Q}$.

Define $\psi_{Q*} : C_*(\mathcal{X}) \to C_*(\mathcal{X}_Q)$ by

$$\psi_Q(\eta) = \begin{cases} 0 & \text{if} \quad \eta = w(Q), \\ -\sum_{\xi \in \mathcal{X}_Q} \dfrac{\kappa(w(Q), \xi)}{\kappa(w(Q), Q)} \xi & \text{if} \quad \eta = Q, \\ \eta & \text{otherwise}, \end{cases} \tag{9}$$

and $\phi_{Q*} : C_*(\mathcal{X}_Q) \to C_*(\mathcal{X})$ by

$$\phi_Q(\eta) = \eta - \frac{\kappa(\eta, Q)}{\kappa\big(w(Q), Q\big)} w(Q). \tag{10}$$

It is left to the reader to check that ψ_{Q*} and ϕ_{Q*} are chain maps.

Lemma 2.5 *The maps ψ_{Q*} and ϕ_{Q*} are chain equivalences.*

Proof A direct computation shows that the composition $\psi_{Q*} \circ \phi_{Q*}$ is the identity map on $\mathcal{X}_Q$. It remains to be shown that $\phi_{Q*} \circ \psi_{Q*}$ is chain homotopic to the identity on $\mathcal{X}$. Define $\Theta \colon C_*(\mathcal{X}) \to C_*(\mathcal{X})$ to be the collection of maps $\big\{\Theta_q \colon C_q(\mathcal{X}) \to C_{q+1}(\mathcal{X})\big\}$ where

$$\Theta_q(\eta) = \begin{cases} \dfrac{1}{\kappa\big(w(Q), Q\big)} w(Q) & \text{if} \quad \eta = Q, \\ 0 & \text{otherwise}. \end{cases}$$

Again, direct computations show that

$$\Theta_{q-1} \circ \partial_q + \partial^Q_{q+1} \circ \Theta_q = \mathbf{I}_{C_q(\mathcal{X})} - \phi_q \circ \psi_q,$$

where ∂^Q is the boundary operator generated by κ_Q and $\mathbf{I}_{C_q(\mathcal{X})}$ is the identity map on the q-chains of $\mathcal{X}$. □

An immediate consequence of Lemma 2.5 is that $\mathrm{H}_*(\mathcal{X}) \cong \mathrm{H}_*(\mathcal{X}_Q)$. Before concluding the proof of Theorem 2.4 we need the following result which guarantees that the Morse incidence function remains unaffected by the reduction step.

Proposition 2.6 *Let $\widetilde{\kappa}_Q$ denote the Morse incidence function of the reduced complex $\mathcal{X}_Q$ with the induced acyclic matching $(\mathcal{A}, w_Q : \mathcal{Q}_Q \to \mathcal{K}_Q)$. Then, $\widetilde{\kappa}_Q \equiv \widetilde{\kappa}$ on $\mathcal{A} \times \mathcal{A}$.*

Proof Note that $\kappa_Q(w(Q'), Q') = \kappa(w(Q'), Q')$ for any $Q' \in \mathcal{Q}_Q$ by the following contradiction. From (8), if the difference

$$\frac{\kappa(w(Q), Q') \cdot \kappa(w(Q'), Q)}{\kappa(w(Q), Q)}$$

does not vanish then we have $Q' \lhd Q \lhd Q'$ by the non-zeroness of the numerator, which violates the standing assumption that the matching on $\mathcal{X}$ is acyclic.

Fix cells A and A' in $\mathcal{A}$ and let $\rho = (Q_1, \ldots, w(Q_M))$ be a connection in $\mathcal{X}_Q$ from A to A'. We make the simplifying assumptions that $Q \not\lhd A$ and $A' \not\lhd w(Q)$, because the argument is very similar to the sequel when one or both of these assumptions is revoked. Now, we have $\kappa_Q(A, A') = \kappa(A, A')$, so we only need to show that the sum-over-connections term of (7) is the same for $\mathcal{X}$ and $\mathcal{X}_Q$.

Since successive elements of $\mathcal{Q}$ in a gradient path are $\ll$-decreasing, there is at most one $m \in 1, \ldots, (M-1)$ with $\kappa_Q(w(Q_m), Q_{m+1}) \neq \kappa(w(Q_m), Q_{m+1})$. If there is no such m, then the index of ρ—and hence its multiplicity—is the same in both $\mathcal{X}$ and $\mathcal{X}'$. On the other hand, if there is such an m then there exists a unique augmented connection ρ^+ from A to A' in $\mathcal{X}$ given by

$$\rho^+ = \big(Q_1, \ldots, w(Q_m), Q, w(Q), Q_{m+1}, \ldots, w(Q_M)\big)$$

and it is readily verified from (5) that the index of ρ in $\mathcal{X}_Q$ equals the sum of the indices of ρ and ρ^+ in $\mathcal{X}$. Thus, the sum over all connections of the multiplicities is preserved in the reduced complex $(\mathcal{X}_Q, \kappa_Q)$. □

Finally, we provide a brief proof of the central theorem of combinatorial Morse theory.

Proof of Theorem 2.4 Let $\{(w(Q_i), Q_i) \mid i = 1, \ldots, I\}$ denote the set of all paired cells. Define the maps $\psi_* \colon C_*(\mathcal{X}) \to C_*(\mathcal{A})$ and $\phi_* \colon C_*(\mathcal{A}) \to C_*(\mathcal{X})$ by the compositions

$$\psi_* := \prod_{i=1}^{I} \psi_{Q_i *} \quad \text{and} \quad \phi_* := \prod_{i=I}^{1} \phi_{Q_i *}.$$

It follows from the reduction step and Proposition 2.6 that the domain and range of ψ and ϕ are as indicated. This argument also guarantees that $(\mathcal{A}, \widetilde{\kappa})$ is a complex by induction: we have already assumed that $(\mathcal{X}, \kappa)$ is a complex as the base case, and demonstrated that removing a cell pair $(Q, w(Q))$ does not alter the cells in $\mathcal{A}$ and that the resulting sequence of incidence functions converges to the Morse incidence function $\widetilde{\kappa}$. Finally, Lemma 2.5 guarantees that the maps ψ and ϕ are chain equivalences as desired. □

3 Filtered Chain Maps

As Theorem 2.4 indicates, combinatorial Morse theory provides a means by which the homology of a complex can be computed using a potentially smaller complex. The main result of this paper is a corresponding result for computing persistent homology of a filtration. There are two issues that need to be resolved: construction of the new filtration and demonstrating that the persistent homology is the same for both filtrations. To clarify the proof of the second issue, we provide a short description of an obvious categorical structure on the set of filtrations.

Let $(\mathcal{X}, \kappa)$ and $(\mathcal{X}', \kappa')$ be complexes and let $\mathcal{F}$ and $\mathcal{F}'$ be filtrations of $\mathcal{X}$ and $\mathcal{X}'$ respectively. We are interested in comparing the persistent homology between these filtrations and thus we turn our attention to chain maps.

Definition 3.1 A *filtered chain map* $\Phi\colon \mathcal{F} \to \mathcal{F}'$ is a sequence $\{\phi_*^k\colon C_*(\mathcal{X}^k) \to C_*(\mathcal{X}'^k)\}$ of chain maps so that for each k the following diagram commutes.

$$\begin{tikzcd}
C_*(\mathcal{X}^k) \arrow[r, hook] \arrow[d, "\phi_*^k"] & C_*(\mathcal{X}^{k+1}) \arrow[d, "\phi_*^{k+1}"] \\
C_*(\mathcal{X}'^k) \arrow[r, hook] & C_*(\mathcal{X}'^{k+1})
\end{tikzcd}$$

Here the horizontal chain maps are induced by inclusions of cells.

A filtered chain map induces a family of maps on homology from $\mathrm{H}_*(\mathcal{X}^k)$ to $\mathrm{H}_*(\mathcal{X}'^k)$ for each k. More interesting for the purposes of this paper is the following.

Proposition 3.2 *Given a filtered chain map $\Phi : \mathcal{F} \to \mathcal{F}'$, there exist well-defined morphisms $\phi_q^{p,k}\colon \mathrm{H}_q^p(\mathcal{X}^k) \to \mathrm{H}_q^p(\mathcal{X}'^k)$ of persistent homology groups given by*

$$\phi_q^{p,k}(z) = \phi_q^{k+p} \circ i_q^{p,k}(z), \quad z \in Z_q(\mathcal{X}^k),$$

where $i_q^{p,k}$ is induced by the inclusion $\mathcal{X}_q^k \hookrightarrow \mathcal{X}_q^{k+p}$

The proof follows directly from the commuting diagram of the preceding definition and the fact that every map in sight is a chain map. Next, we extend the concept of a chain homotopy to filtrations as follows.

Definition 3.3 Let $\Phi, \Psi\colon \mathcal{F} \to \mathcal{F}'$ be filtered chain maps. A *filtered chain homotopy between* Φ *and* Ψ consists of a collection of chain homotopies $\{\Theta^k \mid k = 1, \dots, K\}$ between each ϕ_*^k and ψ_*^k.

If Φ and Ψ are filtered chain homotopic maps from $\mathcal{F}$ to $\mathcal{F}'$ then the induced maps $\phi_*^{p,k}\colon \mathrm{H}_*^p(\mathcal{X}^k) \to \mathrm{H}_*^p(\mathcal{X}'^k)$ and $\psi_*^{p,k}\colon \mathrm{H}_*^p(\mathcal{X}^k) \to \mathrm{H}_*^p(\mathcal{X}'^k)$ are identical on the persistent homology groups. This follows from the commuting diagram of Definition 3.1. Finally, filtered chain maps $\Phi : \mathcal{F} \to \mathcal{F}'$ and $\Psi : \mathcal{F}' \to \mathcal{F}$ are *filtered chain equivalences* if $\Phi \circ \Psi$ and $\Psi \circ \Phi$ are filtered chain homotopic to the identity. In particular, we have the following simple proposition.

Proposition 3.4 *If $\Phi : \mathcal{F} \to \mathcal{F}'$ and $\Psi : \mathcal{F}' \to \mathcal{F}$ are filtered chain equivalences, then their induced maps on the persistent homology groups are inverses, and hence*

$$\mathrm{H}_q^p(\mathcal{X}^k) \cong \mathrm{H}_q^p(\mathcal{X}'^k)$$

for all p, q, k.

The proof of this proposition for a fixed p and k is a straightforward calculation which only requires the existence of a chain homotopy Θ_*^{k+p} between ϕ_*^{k+p} and ψ_*^{k+p}.

4 Filtered Morse Complexes

We begin by extending an acyclic matching of a complex to a filtration. Consider a filtration $\mathcal{F} = \{\mathcal{X}^k \mid k = 1, \dots, K\}$ of a complex $(\mathcal{X}, \kappa)$.

Definition 4.1 A *filtered acyclic matching* of $\mathcal{F}$ comprises an acyclic matching $(\mathcal{A}^k, w^k \colon \mathcal{Q}^k \to \mathcal{K}^k)$ for each frame $\mathcal{X}^k$ with the following additional structure: $\mathcal{A}^k \subset \mathcal{A}^{k+1}$, $\mathcal{K}^k \subset \mathcal{K}^{k+1}$, $\mathcal{Q}^k \subset \mathcal{Q}^{k+1}$ and

$$w^k \equiv w^{k+1} \mid_{\mathcal{Q}^k} \colon \mathcal{Q}^k \to \mathcal{K}^k$$

for each $k \in \{1, \dots, K-1\}$.

Assume we have a filtered acyclic matching $(\mathcal{A}^k, w^k \colon \mathcal{Q}^k \to \mathcal{K}^k)$ of the filtration $\mathcal{F}$. By convention, we write $(\mathcal{A}, w \colon \mathcal{Q} \to \mathcal{K})$ for the matching $(\mathcal{A}^K, w^K \colon \mathcal{Q}^K \to \mathcal{K}^K)$ of the final frame $\mathcal{X}^K = \mathcal{X}$. In particular, $(\mathcal{A}, \widetilde{\kappa})$ is the Morse complex corresponding to the acyclic matching $(\mathcal{A}, w : \mathcal{Q} \to \mathcal{K})$ of $(\mathcal{X}, \kappa)$.

Proposition 4.2 *$\mathcal{M}:=\{\mathcal{A}^k \mid k = 1, \dots, K\}$ is a filtration of the Morse complex $(\mathcal{A}, \widetilde{\kappa})$.*

Proof First we show that $\widetilde{\kappa} \mid_{\mathcal{A}^k \times \mathcal{A}^k} \equiv \widetilde{\kappa}^k$ for each k. Given A in $\mathcal{A}^k$ and an arbitrary $A' \in \mathcal{A}$, it suffices to check by (7) that there are no connections in $\mathcal{X} \setminus \mathcal{X}^k$ from A to A'. To see this, observe that any connection $\rho = (Q_1, w(Q_1), \dots, Q_M, w(Q_M))$ from A must have its initial element $\mathbf{q}_\rho = Q_1$ satisfy $Q_1 \prec A$ where $\prec$ generates the face partial order $\preccurlyeq$ on $\mathcal{X}$. Since $\mathcal{F}$ is a filtration of $\mathcal{X}$, it is known that $\mathcal{X}^k$ is a subcomplex of $\mathcal{X}$ and so $Q_1 \in \mathcal{X}^k$. By the definition of a filtered acyclic matching, $w(Q_1)$ also lies in $\mathcal{K}^k \subset \mathcal{X}^k$. By definition of a gradient path, $Q_2 \lhd Q_1$, i.e., $Q_2 \prec w(Q_1)$ and hence $Q_2 \in \mathcal{X}^k$ by the subcomplex property. Proceeding in this way, we see that every cell in ρ lies in $\mathcal{X}^k$. Thus, we observe that

(1) $\widetilde{\kappa}^k(A, A') = \widetilde{\kappa}(A, A')$ as desired, and more importantly,
(2) given $A \in \mathcal{X}^k$, any connection ρ from A lies entirely in $\mathcal{X}^k$.

Now assume that $\widetilde{\kappa}(A, A') \neq 0$ for some $A \in \mathcal{A}^k$. We will show that $A' \in \mathcal{A}^k$, thus proving the desired subcomplex property for $\mathcal{A}^k \subset \mathcal{A}$. From (7) we see that either $\kappa(A, A') \neq 0$ or there exists at least one connection ρ from A to A' with $m(\rho) \neq 0$. In the first case we have $A' \in \mathcal{X}^k$ by the fact that $\mathcal{X}^k$ is a subcomplex of $\mathcal{X}$, so without loss of generality we assume that there exists some connection ρ from A to A'. By the second observation above, we know that each cell of ρ lies in $\mathcal{X}^k$, and in particular we have the last element $\mathbf{k}_\rho$ in $\mathcal{X}^k$. Since $A' \prec \mathbf{k}_\rho$ by definition of a connection, we see that $A' \in \mathcal{X}^k$ by the subcomplex property. □

We call $\mathcal{M}$ the *Morse filtration* associated to the filtered acyclic matching $(\mathcal{A}^k, w^k \colon \mathcal{K}^k \to \mathcal{Q}^k)$. By Theorem 2.4, the associated Morse complex $(\mathcal{A}, \widetilde{\kappa})$ has the same homology as the complex $(\mathcal{X}, \kappa)$. We now extend this result to the level of persistent homology.

Theorem 4.3 *Let $\mathcal{F} = \{\mathcal{X}^k \mid k = 1, \dots, K\}$ be a filtration of a complex $(\mathcal{X}, \kappa)$ with filtered acyclic matching $(\mathcal{A}^k, w^k \colon \mathcal{Q}^k \to \mathcal{K}^k)$. Let $(\mathcal{A}, \widetilde{\kappa})$ be the associated Morse complex with Morse filtration $\mathcal{M} = \{\mathcal{A}^k \mid k = 1, \dots, K\}$. Then for all k, q, and p,*

$$\mathrm{H}_q^p(\mathcal{X}^k) \cong \mathrm{H}_q^p(\mathcal{A}^k). \tag{11}$$

In order to make the proof of this theorem transparent, consider the function $b \colon \mathcal{X} \to \mathbb{Z}$ given by

$$b(\xi) := \min\{k \mid \xi \in \mathcal{X}^k\}. \tag{12}$$

The two important properties of b are:

(1) by Definition 4.1, $b(Q) = b(w(Q))$ for each $Q \in \mathcal{Q}$, and
(2) by the subcomplex property, $b(\xi) \le b(\eta)$ whenever $\xi \prec \eta$.

Let $\{(w(Q_i), Q_i) \mid i = 1, \dots, I\}$ denote the set of all paired cells in $\mathcal{K} \times \mathcal{Q}$ with the following additional constraint:

$$\text{if } b(Q_j) > b(Q_i) \quad \text{then} \quad j > i.$$

This gives us positive integers $I_1 \le I_2 \le \dots \le I_K = I$ such that $Q_i \in \mathcal{X}^k$ if and only if $i \le I_k$. Define chain maps $\psi_*^k \colon C_*(\mathcal{X}^k) \to C_*(\mathcal{A}^k)$ and $\phi_*^k \colon C_*(\mathcal{A}^k) \to C_*(\mathcal{X}^k)$ by the compositions

$$\psi_*^k := \prod_{i=1}^{I_k} \psi_{Q_i *} \quad \text{and} \quad \phi_*^k := \prod_{i=I_k}^{1} \phi_{Q_i *}. \tag{13}$$

Denote by Ψ and Φ the collections of chain maps $\{\psi^k\}$ and $\{\phi^k\}$, respectively. By Proposition 3.4, the proof of Theorem 4.3 concludes with the following result.

Proposition 4.4 *The maps $\Psi : \mathcal{F} \to \mathcal{M}$ and $\Phi : \mathcal{M} \to \mathcal{F}$ are filtered chain equivalences.*

Proof Lemma 2.5 implies that ψ_*^k and ϕ_*^k are chain equivalences for all $k = 1, \dots, K$, and so by the requirements of Definition 3.3 we are only required to show that Ψ and

Φ are filtered chain maps. That is, given any $k \in \{1, \dots, K-1\}$ we will show that the following diagrams commute:

$$\begin{tikzcd} C_*(\mathcal{X}^k) \arrow[r, hook] \arrow[d, "\psi_*^k"'] & C_*(\mathcal{X}^{k+1}) \arrow[d, "\psi_*^{k+1}"] \\ C_*(\mathcal{A}^k) \arrow[r, hook] & C_*(\mathcal{A}^{k+1}) \end{tikzcd} \qquad \begin{tikzcd} C_*(\mathcal{X}^k) \arrow[r, hook] & C_*(\mathcal{X}^{k+1}) \\ C_*(\mathcal{A}^k) \arrow[r, hook] \arrow[u, "\phi_*^k"] & C_*(\mathcal{A}^{k+1}) \arrow[u, "\phi_*^{k+1}"'] \end{tikzcd}$$

Fix some $Q \in \mathcal{Q}$ with $b(Q) \geq k+1$. Note from the defining formula (9) that ψ_Q differs from the identity only on Q and $w(Q)$, both of which have b values exceeding $k+1$ by our explicit assumption on Q and the first observed property of the function b. Therefore, $\psi_{Q*}\mid_{C(\mathcal{X}^k)}$ is the identity map. Thus, Ψ is a filtered chain map by (13).

Similarly, we show that ϕ_{Q*} is the identity map on $C(\mathcal{X}^k)$ whenever $b(Q) \geq k+1$. From (10), $\phi_{Q*}(\eta)$ may differ from η only when $\kappa(\eta, Q) \neq 0$, i.e., when $Q \prec \eta$. By the second observed property of the function b, we must have $b(\eta) \geq b(Q) = k+1$ and so $\eta \in \mathcal{X} \setminus \mathcal{X}^k$ as desired. Thus, Φ is a filtered chain map as well. □

5 Algorithms

Throughout this section $\mathcal{F} = \{\mathcal{X}^k \mid \mathcal{X}^k \subset \mathcal{X}^{k+1}, k = 1, \dots, K\}$ denotes a fixed filtration of a complex $(\mathcal{X}, \kappa)$ with boundary operator ∂ and face partial order $\preccurlyeq$ generated by $\prec$. We make use of the following notation in the algorithms: given $\xi \in \mathcal{X}$, the coboundary cells of ξ are given by

$$\mathsf{cb}(\xi) := \{\eta \in \mathcal{X} \mid \xi \prec \eta\}.$$

It is not necessary to impose any specific order on the cells in $\mathsf{cb}(\xi)$.

5.1 Description

Theorem 4.3 implies that it is possible to compute the persistent homology groups of $\mathcal{F}$ by applying the algorithm of [35] to a smaller filtration $\mathcal{M} = \{\mathcal{A}^k \mid k = 1, \dots, K\}$ associated with a Morse complex $(\mathcal{A}, \widetilde{\kappa})$ for $\mathcal{X}$. The usefulness of this approach depends upon having an efficient algorithm for constructing the filtration $\mathcal{M}$ and the Morse incidence function $\widetilde{\kappa}$, or equivalently the boundary operator $\partial^{\mathcal{A}}$ on $\mathcal{A}$.

The filtration $\mathcal{M}$ and incidence function $\widetilde{\kappa}$ depend on the acyclic matching $(\mathcal{A}, w\colon \mathcal{Q} \to \mathcal{K})$. The trivial matching given by $\mathcal{A} = \mathcal{X}$ and $\mathcal{Q} = \mathcal{K} = \varnothing$ always exists, but results in the same filtration and thus provides no savings in computational cost. Clearly, the desired goal is to choose an acyclic matching which minimizes the number of cells in $\mathcal{A}$, or equivalently maximizes the number of cells paired by $w\colon \mathcal{Q} \to \mathcal{K}$. It is known that in general the problem of constructing an optimal acyclic matching is NP hard (see [15] and [20, Sect. 4.5]).

Our approach to producing an acyclic matching is based on the coreduction homology algorithm of Mrozek and Batko [23] which has proven effective in computing homology of complexes [14,13]. This algorithm is based on the following idea. Let $\mathcal{X}' \subset \mathcal{X}$. A pair of cells $\xi, \eta \in \mathcal{X}'$ form a *coreduction pair* in $\mathcal{X}'$ if restricted to $C_*(\mathcal{X}')$

$$\partial \xi = u \cdot \eta,$$

where $u \in \mathbf{R}$ is a unit. In this case we make the identifications $\xi \in \mathcal{K}$, $\eta \in \mathcal{Q}$ and $w(\eta) = \xi$, and remove both ξ and η from $\mathcal{X}'$.

We differ from [14,13] in the construction of $\partial^{\mathcal{A}}$. From (7) it is clear that $\widetilde{\kappa}$—and hence $\partial^{\mathcal{A}}$—is defined by summing over all connections between cells in $\mathcal{A}$. A naïve attempt to enumerate all the connections between two such cells can lead to a combinatorial explosion. To circumvent this summation, we make use of the observation that the coreduction-based construction of the pairing $w\colon \mathcal{Q} \to \mathcal{K}$ is done by building gradient paths in reverse order with respect to $\ll$. We therefore proceed by assigning to each cell $\zeta \in \mathcal{X}$ a chain $g(\zeta) \in C_*(\mathcal{A})$ such that if $A \in \mathcal{A}$, then $g(A) = \partial^{\mathcal{A}} A$.

Initially we set $g(\zeta) := 0$ for every cell. However, as the coreduction algorithm is used to construct the acyclic matching—that is, as the gradient paths are constructed—the value of $g(\zeta)$ is suitably modified. Thus, the computation of $\partial^{\mathcal{A}}$ can be done during the construction of $w\colon \mathcal{Q} \to \mathcal{K}$ using the subroutine `UpdateGradientChain` presented below.

Algorithm: `UpdateGradientChain`
In: $\xi \in \mathcal{X}$;
Out: Updates $g(\zeta)$ for each $\zeta \in \mathsf{cb}_{\mathcal{N}}(\xi)$

```
for each $\zeta \in \mathsf{cb}_{\mathcal{N}}(\xi)$
    if $\xi \in \mathcal{N}^k_{\mathcal{A}}$ for some $k$
        $g(\zeta) \leftarrow g(\zeta) + \kappa(\zeta, \xi) \cdot \xi$
    else
        $g(\zeta) \leftarrow g(\zeta) + \kappa(\zeta, \xi) \cdot g(\xi)$
    end if
end for
```

To emphasize that we only need to store each cell once rather than save a copy for each subcomplex $\mathcal{X}^k$ containing that cell, we partition the cells in $\mathcal{X}$ by setting

$$\mathcal{N}^k = \mathcal{X}^k \setminus \mathcal{X}^{k-1}, \quad k = 1, \dots, K,$$

where $\mathcal{X}^0 = \varnothing$. Note that each cell $\xi \in \mathcal{X}$ lies uniquely in $\mathcal{N}^{b(\xi)}$ where $b : \mathcal{X} \to \mathbb{Z}$ is as defined in (12). The partition $\{\mathcal{N}^k \mid k = 1, \dots, K\}$ defines the input to our algorithms; each cell ξ is eventually excised from $\mathcal{N}^{b(\xi)}$ either as an element of $\mathcal{A}$ or in a coreduction pair. Given a cell $\xi \in \mathcal{N}^k$, we denote by $\mathsf{cb}_{\mathcal{N}}(\xi)$ and $\partial^{\mathcal{N}}(\xi)$ the coboundary cells and the boundary chain when restricted to $\{\mathcal{N}^*\}$. Once a cell is removed from $\mathcal{N}^k$, it is also removed from the corresponding $\mathsf{cb}_{\mathcal{N}}$ and boundary $\partial^{\mathcal{N}}$ of the remaining cells. Similarly, the cells of the output Morse complex $(\mathcal{A}, \widetilde{\kappa})$ are also partitioned via $\mathcal{N}^k_{\mathcal{A}} = \mathcal{A}^k \setminus \mathcal{A}^{k-1}$.

The next two subroutines perform tasks pertaining to removing cells. The first subroutine—called `MakeCritical`—chooses an arbitrary A' of minimal dimension in a non-empty $\mathcal{N}^k$ and excises it as an element of $\mathcal{A}$.

```
Algorithm: MakeCritical
In: $k \in \{1, \dots, K\}$ so that $\mathcal{N}^k \neq \varnothing$
Out: $A' \in \mathcal{A}^k$

choose $A' \in \mathcal{N}^k$ of min dimension
add $A'$ to $\mathcal{N}^k_{\mathcal{A}}$;
updateGradientChain($A'$)
remove $A'$ from $\mathcal{N}^k$
$\partial^{\mathcal{A}} A' \leftarrow g(A')$
return $A'$
```

Thus, the output A' of MakeCritical becomes a generator of $C_*(\mathcal{A}^k)$. The gradient chains of remaining coboundary cells $\mathsf{cb}_{\mathcal{N}}(A')$ are then updated to reflect their incidence with A'. In this manner, the construction of gradient chains is from the "bottom-up". Finally, the action of the Morse boundary operator $\partial^{\mathcal{A}}$ on A' is recovered from the corresponding gradient chain $g(A')$.

The obvious operation of the second subroutine, called RemovePair, is to perform the reduction step from Sect. 2.3 with respect to a single coreduction pair (K, Q) from the complex.

```
Algorithm: RemovePair
In: $K, Q \in \mathcal{N}^{k*}$ with $\partial^{\mathcal{N}} K = u \cdot Q$, Queue of cells Que, $d \in \mathbb{N}$
Out: Removes $(K, Q)$ as a cell pair from $\mathcal{N}^{k*}$

remove $K$ from $\mathcal{N}^{k*}$
enqueue $\mathsf{cb}_{\mathcal{N}}(Q)$ in Que
if dim $Q = d$
    $g(Q) \leftarrow -\frac{g(K)}{u}$
    updateGradientChain($Q$)
end if
remove $Q$ from $\mathcal{N}^{k*}$
```

Recall that on the theoretical level coreduction pairs are identified as w-paired cells and hence they define steps in gradient paths. Thus, before the coreduction pair can be removed two additional steps need to be performed involving the remaining coboundary cells $\mathsf{cb}(Q)$ of Q. First, we check if the removal of Q has created new coreduction pairs. For this, it suffices to check cells in the coboundary of Q and so we enqueue those cells in a queue structure. Secondly, if the pair (K, Q) potentially lies on a gradient path between unpaired cells of adjacent dimension, the gradient chains of Q and hence of its remaining coboundary cells are updated by a call to UpdateGradientChain.

These subroutines are combined to form MorseReduce, which is our main algorithm. The input to this algorithm is a filtration $\mathcal{F}$ of a complex $(\mathcal{X}, \kappa)$ partitioned by $\{\mathcal{N}^k\}$ as described above; the incidence function κ represents knowledge of the boundary operator ∂. The output is a new filtration $\mathcal{M}$ of the Morse complex $(\mathcal{A}, \widetilde{\kappa})$—partitioned by $\{\mathcal{N}^k_{\mathcal{A}}\}$—corresponding to a coreduction-based acyclic matching $(\mathcal{A}^k, w^k : \mathcal{Q}^k \to \mathcal{K}^k)$. The Morse incidence function $\widetilde{\kappa}$ is recovered from the boundary operator $\partial^{\mathcal{A}}$.

Algorithm: `MorseReduce`
In: $\{\mathcal{N}^k\}_1^K$
Out: $\{\mathcal{N}_{\mathcal{A}}^k\}_1^K$

```
01  for each $k = 1, \dots, K$
02      while $\mathcal{N}^k \neq \emptyset$
03          $A' \leftarrow$ MakeCritical($k$)
04          Que := Empty Queue of Cells
05          enqueue $\mathsf{cb}_{\mathcal{N}}(A')$ in Que
06          while Que $\neq \emptyset$
07              dequeue $\xi$ from Que
08              if $\partial^{\mathcal{N}} \xi = 0$
09                  enqueue $\mathsf{cb}_{\mathcal{N}}(\xi)$ in Que
10              else if $\partial^{\mathcal{N}} \xi = u \cdot \eta$ for some $\eta \in \mathcal{N}^{b(\xi)}$ and unit $u$
11                  RemovePair($\xi$, $\eta$, Que, dim $A$)
12              end if
13          end while
14      end while
15  end for
16  return $\{\mathcal{N}_{\mathcal{A}}^k\}_1^K$
```

Note that we use a queue data structure Que which gets re-initialized once for each iteration of the outer **while** loop from Line 02. We keep track of which cells are in Que so that no cell is queued twice per such iteration. This can be achieved in practice either by storing an additional flag for each cell or by mirroring the queue in a separate data structure which has been optimized for search.

5.2 Verification

We use Theorem 4.3 to confirm that the output filtration $\mathcal{M}$ generated by `MorseReduce` has the same persistent homology groups as those of the input filtration $\mathcal{F}$.

Theorem 5.1 *Let $\mathcal{F} = \{\mathcal{X}^k \mid k = 1, \dots, K\}$ be a filtration of a complex $(\mathcal{X}, \kappa)$ over a PID* $\mathbf{R}$ *and define $\mathcal{N}^k := \mathcal{X}^k \setminus \mathcal{X}^{k-1}$ for each k. Then*

(1) `MorseReduce` *terminates when applied to $\{\mathcal{N}^k\}_1^K$ and produces smaller collections of cells $\{\mathcal{N}_{\mathcal{A}}^k\}_1^K$.*
(2) *The output $\{\mathcal{N}_{\mathcal{A}}^k\}$ defines a filtration $\mathcal{M}$ of a complex $(\mathcal{A}, \widetilde{\kappa})$ where each frame $\mathcal{A}^k$ is given by $\bigcup_{\ell=1}^k \mathcal{N}_{\mathcal{A}}^\ell$ and the underlying incidence function $\widetilde{\kappa}$ corresponds to the boundary operator $\partial^{\mathcal{A}}$.*
(3) *For each p, q and k, we have an isomorphism of the corresponding persistent homology group*

$$\mathrm{H}_q^p(\mathcal{X}^k) \cong \mathrm{H}_q^p(\mathcal{A}^k).$$

Proof Each iteration of the outer **while** loop from Line 02 permanently excises at least one cell A via `MakeCritical`.[1] The fact that no cell is queued twice during any iteration of the second **while** loop in Line 07 guarantees the absence of infinite loops. Moreover, it is clear that the final size of each $\mathcal{N}^k_{\mathcal{A}}$ is smaller than the initial size of $\mathcal{N}^k$ because `MakeCritical` is only called once per iteration of the outer **while** loop and each call to `MakeCritical` results in a single cell from $\mathcal{N}^k$ being removed and stored in the corresponding $\mathcal{N}^k_{\mathcal{A}}$. Thus, $\mathcal{N}^k_{\mathcal{A}} \subset \mathcal{N}^k$ for each k.

Observe from Line 10 that if (ξ, η) is sent to `RemovePair`, then $b(\xi) = b(\eta)$ and $\kappa(\xi, \eta)$ equals some unit u in $\mathbf{R}$. Let $k_* = b(\xi)$, and note that defining $w^{k_*}(\eta) = \xi$ for each such pair constructs $w^{k_*} : \mathcal{Q}^{k_*} \to \mathcal{K}^{k_*}$. Combining this pairing information with the output of `MakeCritical` produces a filtered partial matching $(\mathcal{A}^k, w^k : \mathcal{Q}^k \to \mathcal{K}^k)$ of $\mathcal{F}$.

To see that this partial matching is acyclic, observe from Lines 10 and 11 that a pairing $w(\eta) = \xi$ is only made when η is the last remaining face of ξ, i.e., the unique cell in $\{\zeta \in \mathcal{N}^{b(\xi)} \mid \zeta \prec \xi\}$. Recall that $Q \lhd \eta$ for some $Q \in \mathcal{Q}$ if and only if $Q \prec \xi$ by (4). Thus, all elements of $\mathcal{Q}$ satisfying $Q \lhd \eta$ must have already been been excised before the pair (ξ, η) and so the order of pair excision respects the relation $\lhd$ on $\mathcal{Q}$. Therefore, the transitive closure $\ll$ of the generating relation $\lhd$ must be a partial order on $\mathcal{Q}$ as desired.

By Theorem 4.3, in order to show that the output determines a filtration $\mathcal{M}$ with isomorphic persistent homology to $\mathcal{F}$, it suffices to establish that $\mathcal{M}$ is the Morse filtration associated to the acyclic matching $(\mathcal{A}^k, w^k : \mathcal{Q}^k \to \mathcal{K}^k)$. Thus, we must ensure that the stored boundary $\partial^{\mathcal{A}}$ of each cell $A \in \mathcal{A}$ built from the corresponding gradient chain $g(A)$ equals the boundary operator corresponding to the Morse incidence function $\widetilde{\kappa}$ from (7). This is addressed by the subsequent proposition, which concludes the proof. □

The proof of the following proposition employs the usual inner product $\langle\, ,\, \rangle : C(\mathcal{X}) \times C(\mathcal{X}) \to \mathbf{R}$ on chains of the input complex $(\mathcal{X}, \kappa)$ obtained by treating the cells in $\mathcal{X}$ as an orthonormal basis.

Proposition 5.2 *Assume the hypotheses and notation of Theorem 5.1. For cells A and A' in $\mathcal{A}$,*

$$\left\langle g(A), A' \right\rangle = \widetilde{\kappa}(A, A'),$$

where $\widetilde{\kappa}$ is the Morse incidence function (7) corresponding to the acyclic matching from the proof of Theorem 5.1.

Proof We provide a brief summary of how gradient chains are constructed. Assume throughout that A' is excised via `MakeCritical`. Consider the following two cases.

[1] In practice many more cells are also removed, since if (ξ, η) appear as a coreduction pair then they are excised during the `RemovePair` subroutine. Furthermore, observe that their appearance as a coreduction pair is equivalent to their being part of a gradient path that descends to A.

[A] Assume that ζ is an unremoved cell with $A' \prec \zeta$. Then, by Line 02 of `MakeCritical` and the subsequent call to `UpdateGradientChain`, the gradient chain $g(\zeta)$ of ζ is incremented as follows:

$$g(\zeta) \leftarrow g(\zeta) + \kappa(\zeta, A') \cdot A'.$$

Since this is the first instance of A' being added to gradient chains, we are guaranteed to have $\langle g(\zeta), A'\rangle = \kappa(\zeta, A')$ when `MakeCritical` returns A'.
[Q] Assume ζ is an arbitrary unremoved cell. Each cell Q excised as an element of $\mathcal{Q}$ via `RemovePair` inherits its gradient chain from the existing gradient chain of its paired cell $w(Q)$ by the formula

$$g(Q) = \frac{g\big(w(Q)\big)}{-\kappa\big(w(Q), Q\big)}.$$

This follows from Line 04 of `RemovePair`. As `UpdateGradientChain` is called in the next line, each remaining cell ζ satisfying $Q \prec \zeta$ has its gradient chain incremented by $\kappa(\zeta, Q) \cdot g(Q)$. By the preceding formula for $g(Q)$, we have

$$g(\zeta) \leftarrow g(\zeta) + \frac{\kappa(\zeta, Q)}{-\kappa\big(w(Q), Q\big)} \cdot g\big(w(Q)\big).$$

Thus, there are two ways a critical cell A' appears with non-zero coefficient in the gradient chain $g(\zeta)$ of some hitherto unremoved cell ζ: either $A' \prec \zeta$ and we directly apply **[A]**, or $\langle g(w(Q)), A'\rangle \neq 0$ for some previously removed $Q \in \mathcal{Q}$ with $Q \prec \zeta$ and we apply **[Q]**. Combining these contributions, we have the following formula:

$$\big\langle g(\zeta), A'\big\rangle = \kappa(\zeta, A') + \sum_{Q \in \mathcal{Q}} \frac{\kappa(\zeta, Q)}{-\kappa\big(w(Q), Q\big)} \big\langle g\big(w(Q)\big), A'\big\rangle. \tag{14}$$

Now assume that a cell A is eventually removed from the input filtration by `MakeCritical`. Recalling (7), we substitute $\zeta = A$ in (14) to get

$$\big\langle g(A), A'\big\rangle = \kappa(A, A') + \sum_{Q \in \mathcal{Q}} \frac{\kappa(A, Q)}{-\kappa\big(w(Q), Q\big)} \big\langle g\big(w(Q)\big), A'\big\rangle.$$

Applying (14) recursively to each $\langle g(w(Q)), A'\rangle$ in the expression above completes the argument. □

5.3 Complexity

Let $(\mathcal{X}, \kappa)$ be a complex over a PID $\mathbf{R}$ filtered by $\mathcal{F} = \{\mathcal{X}^k\}_1^K$ with face partial order $\preceq$ generated by the usual relation:

$$\xi \prec \xi' \quad \text{if } \kappa(\xi', \xi) \neq 0 \in \mathbf{R}.$$

5.3.1 Parameters and Assumptions

We describe the computational cost of using `MorseReduce` to construct an acyclic matching $(\mathcal{A}, w : \mathcal{Q} \to \mathcal{K})$ on $\mathcal{X}$ as well as the Morse complex $(\mathcal{A}, \widetilde{\kappa})$ in terms of the following *complexity parameters*. A similar analysis for unfiltered complexes may be found in [14].

(1) The *input size*—denoted by n—is the number of cells in $\mathcal{X}$.
(2) The *output size* is the number of cells in the filtered Morse complex $\mathcal{A}$ which we denote by m. Note that m is partitioned by $m = m_0 + \cdots + m_D$ where m_d is the cardinality of d-dimensional cells in $\mathcal{A}$. As we have remarked before, constructing an optimal acyclic matching—that is, a matching which minimizes m—is NP hard [15,20]. Providing sharp bounds on optimal m values relative to n for arbitrary complexes would require major breakthroughs in algebraic topology as well as graph theory. Therefore, we leave m as a parameter.
(3) The *coboundary mass* p of $\mathcal{X}$ is defined as

$$p = \sup_{\xi \in \mathcal{X}} \#\big\{\eta \in \mathcal{X} \mid \kappa(\eta, \xi) \neq 0\big\},$$

where # denotes cardinality. Thus, the coboundary mass bounds the number of cells $\eta \in \mathcal{X}$ which satisfy $\xi \prec \eta$ for a given cell $\xi \in \mathcal{X}$. Even though p may safely be bounded by n, in most situations this is a gross over-estimate. For example, the coboundary mass of a d-dimensional cubical grid is bounded above by $2d$ independent of the total number of cubes present.

For the purposes of complexity analysis, we also make these two *simplifying assumptions*:

(1) we assume that adding, removing or locating a cell $\xi \in \mathcal{X}$ incurs a constant cost, and
(2) we assume that ring operations in **R** may be performed in constant time so that the cost of adding and scaling gradient chains is linear in the length of the chains involved.

5.3.2 Evaluating Complexity

We begin by evaluating the complexity of a single iteration of the outer **while** loop from Line 02 of `MorseReduce`. Assume that in this iteration the call to `MakeCritical` via Line 03 has returned a cell A' of dimension d. Since in each iteration of this **while** loop we add a cell to Que at most once, the maximum size attainable by Que is n. Moreover, each Que insertion involves testing the coboundary of a cell which requires at most p operations. In light of these bounds, we will just assume that the total cost of managing the Que data structure within a single **while** iteration depends linearly on $n \cdot p$ and we will not separately tabulate the cost of each Que operation.

We also require the following observations regarding the cost of the three subroutines in terms of the complexity parameters defined previously.

(1) The cost of calling `UpdateGradientChain` on a d-dimensional cell equals $\mathrm{O}(p \cdot m_d)$. This follows from the fact that we must iterate over each cell ζ in the remaining coboundary of ξ and update the gradient chain $g(\zeta)$ which consists of d-dimensional cells in $\mathcal{A}$.
(2) A call to `MakeCritical` in Line 03 also costs $\mathrm{O}(p \cdot m_d)$, since the only nontrivial operation is the call to `UpdateGradientChain` in Line 03.
(3) In the worst case, the **if** statement from Line 03 of `RemovePair` always evaluates positively and hence `UpdateGradientChain` is called. Thus, each call to `RemovePair` also incurs a worst case cost of $\mathrm{O}(p \cdot m_d)$ since all other nontrivial operations only involve Que insertion.

Since the inner **while** loop from Line 06 of `MorseReduce` depends only on the size of Que, it may run at most n times. Thus, the cost of iterating the outer **while** loop from Line 02 reduces to a single call to `MakeCritical`, the management of the Que structure, and at most n calls to `RemovePair`. Adding these respective contributions, the total cost of a single iteration of this outer **while** loop equals

$$\mathrm{O}(p \cdot m_d) + \mathrm{O}(n \cdot p) + \mathrm{O}(n \cdot p \cdot m_d).$$

The third quantity clearly dominates the first two, so the desired complexity estimate of the outer **while** loop when A' has dimension d is $\mathrm{O}(n \cdot p \cdot m_d)$.

It now suffices to estimate how many iterations of the outer **while** loop are actually executed in a single instance of `MorseReduce`. But this is straightforward: each such iteration corresponds to exactly one cell $A' \in \mathcal{A}$ as returned by `MakeCritical`, so this **while** loop executes precisely m times. Partitioning $m = m_0 + \cdots + m_D$ by dimension as usual, we estimate the following total cost of running `MorseReduce` in terms of our complexity parameters:

$$\mathrm{O}\left(n \cdot p \cdot \sum_{d=0}^{D} m_d^2\right)$$

In light of this expression, it is convenient to define the number $\tilde{m} \leq m^2$ by

$$\tilde{m} = \sum_{d=0}^{D} m_d^2.$$

Thus, we have proved the following result regarding the computational complexity of `MorseReduce`.

Proposition 5.3 *Assume that* `MorseReduce` *is executed on a filtered complex $\mathcal{X}$ of top dimension D, size n and coboundary mass p. If the resulting Morse complex $\mathcal{A}$ has size $m = m_0 + \cdots + m_D$, then the worst-case complexity is bounded by* $\mathrm{O}(n \cdot p \cdot \tilde{m})$, *where $\tilde{m} = m_0^2 + \cdots + m_D^2$.*

Thus, the cost of computing the maps induced on homology by inclusions $\mathcal{X}^k \subset \mathcal{X}^{k+1}$ in the filtered complex $\mathcal{X}$ over an arbitrary PID $\mathbf{R}$ reduces from $\mathrm{O}(n^4)$ [12] to

$\mathrm{O}(n \cdot p \cdot \tilde{m} + m^4)$ if `MorseReduce` is used as a pre-processor. In the special case when $\mathbf{R}$ is a field, computing persistence intervals has complexity $\mathrm{O}(n^{\omega})$ where ω is the matrix multiplication exponent [21]. Therefore, the cost of computing the persistence intervals of $\mathcal{X}$ *after* applying `MorseReduce` to $\mathcal{X}$ equals $\mathrm{O}(n \cdot p \cdot \tilde{m} + m^{\omega})$. In practice, persistence intervals are typically computed using the standard algorithm from [35] which has cubic complexity in the worst case [22]. Thus, using `MorseReduce` as a pre-processor for the standard algorithm lowers the overall complexity from $\mathrm{O}(n^3)$ to $\mathrm{O}(n \cdot p \cdot \tilde{m} + m^3)$. If m is much smaller than n, then the $n \cdot p \cdot \tilde{m}$ term is dominant in each case and one observes essentially linear cost in terms of the input size n.

Remark 5.4 The efficiency of our approach depends crucially on m being much smaller than n. In the worst case, no cells get paired and we are left with $m = n$. Examples of filtered complexes for which this is the case may be easily constructed in one of two ways:

C1 Consider a complex $\mathcal{X}$ such that each non-zero incidence $\kappa(\xi, \xi') \in \mathbf{R}$ is not a unit for any pair of cells $\xi, \xi' \in \mathcal{X}$.

C2 Consider a complex $\mathcal{X}$ so that whenever $\kappa(\xi, \xi') \neq 0$ for cells ξ, ξ' we have $b(\xi) \neq b(\xi')$. Since matched cells are required to have the same b values by Definition 4.1, no non-trivial matching is possible in this case.

It is easy to test the input complex for both conditions **C1** and **C2** in $\mathrm{O}(n \cdot p)$ time by checking each pair of cells $\xi, \xi' \in \mathcal{X}$ with non-trivial incidence $\kappa(\xi, \xi') \neq 0$. Moreover, for a wide variety of practical situations we do not expect the worst case scenario to occur. For example, if one considers simplicial or cubical complexes arising from experimental data, then the following structures are common.

(1) For both cubical and simplicial complexes all non-zero incidences are units ± 1 in any PID $\mathbf{R}$, so **C1** is avoided.
(2) The b values are only prescribed on top-dimensional cells (such as grayscale pixel or voxel values for image data). In these situations, each lower dimensional cell recursively inherits its b value as the minimum b value encountered among its co-boundary cells. This guarantees the existence of at least some cells $\xi \prec \xi'$ with $b(\xi) = b(\xi')$ and avoids **C2**,
(3) The b values are inherited from lower dimensional cells. A prime example is the Vietoris–Rips complex built around point cloud data. Here each simplex inherits the maximum b value encountered in its 1-skeleton. Again, this process ensures the existence of dimensionally adjacent cells which share b values and hence avoids **C2**.

As we demonstrate in Sect. 6, Morse theoretic pre-processing is effective for computing persistent homology of several types of filtered complexes arising from experimental data. We believe that a more nuanced approach to analyzing the effectiveness of combinatorial Morse theory would require imposing reasonable probability measures on the set of all complexes and proving statements regarding the expected fraction of cells reduced. We leave such considerations to future work.

6 Experimental Results

The popularity of persistent homology as a tool for understanding large datasets has led to a variety of highly efficient implementations and preprocessing algorithms. To the best of our knowledge, the first use of combinatorial Morse theory for persistence computation was presented in [27], where the construction of an acyclic matching is done via the algorithm `ProcessLowerStars`. The major differences between `ProcessLowerStars` and `MorseReduce` are as follows:

(1) **Coefficient Rings**: `ProcessLowerStars` constructs the Morse complex over $\mathbb{Z}/2\mathbb{Z}$ coefficients whereas `MorseReduce` may be applied to filtered cell complexes over an arbitrary PID.
(2) **Complex Types**: `ProcessLowerStars` requires a filtered cubical complex as input along with birth times provided only for top dimensional cells. The birth time for a lower dimensional cell is recursively inherited as the minimum birth found among all cells in its coboundary. On the other hand, `MorseReduce` is complex-agnostic and does not impose any such top-down inheritance requirement. Moreover, `ProcessLowerStars` requires perturbing the input filtration so that no two top-cells have the same birth time. This is unnecessary in `MorseReduce` even when dealing with 3D cubical data.
(3) **Dimensions**: `MorseReduce` is dimension-independent whereas `ProcessLowerStars`, as written, requires a top dimension of 3.

The existing frameworks [11,33] for applying combinatorial Morse theory to persistent homology computation rely heavily on the efficient storability of cubical datasets of low dimension over $\mathbb{Z}/2\mathbb{Z}$ coefficients, and we do not see an obvious means of applying similar techniques to other types of complexes. Since our approach with `MorseReduce` only requires face relations on the input complex as encoded by the underlying incidence function, it applies to filtered complexes independent of coefficient ring and dimensionality. The coreduction-based strategy of [24] has similarly wide applicability but it only pairs those cells which have gradient paths descending to unpaired cells of dimension zero, and therefore results in the reduction of fewer cells when compared to `MorseReduce`.

Note that since the output of `MorseReduce` is a filtration in its own right, it is possible to iterate the algorithm until the number of reductions performed becomes essentially negligible. Thus, the cells output by an iteration of `MorseReduce` get further partitioned by the subsequent iteration and may get paired by the associated acyclic matching. We are not aware of any existing technique which allows for such iteration on arbitrary filtrations.

We demonstrate the results of `MorseReduce` on cubical grids, simplicial complexes, Vietoris–Rips complexes and movies. The cubical complexes come from sublevel sets of finite element Cahn–Hilliard simulations and the simplicial complexes arise from brain imaging data. The Vietoris–Rips complexes come from point clouds of experimental granular flow data. Our largest datasets by far, courtesy of M. Schatz, are two black and white movies obtained by segmenting Rayleigh–Bénard convection data, each successive frame consisting of about $155,000$ three dimensional cubes.

Table 1 Experimental results

Type	Dim	# Frames	# Cells (M)	Red. # cells (K)	`SP`	`MR + SP`
C	2	16	0.26	2.31	2.4	0.9
C	3	25	1.24	3.35	12.7	4.2
C	3	2,400	5.25	9.50	195.8	46.6
S	5	20	12.86	27.13	73.9	8.8
S	5	5,000	0.86	99.8	1951.6	153.1
VR	2	100	2.34	86.33	1277.0	37.7
VR	3	50	9.50	18.31	286.5	47.2
VR	3	250	1.29	53.95	551.1	125.2
M	3	209	259.21	1.25	DNF	7,213
M	3	215	266.67	2.10	DNF	7,416

The implementation of `MorseReduce` benchmarked here was coded in C++ using the standard template library and compiled using the GNU C++ compiler with optimization level O3. All computations were performed on an Intel Core i5 machine with 32 GB of available RAM and virtual memory disabled. The source code for our implementation is available at [26].

The comparison is with our implementation of the standard algorithm for computing persistent homology as found in [35] which we will denote by `SP`. While this algorithm may be found in various flavors and as part of the software package **jPlex**[2] or from the **Dionysus** project,[3] the authors feel that the present comparison is fair because the same data structures are used in both cases. The `SP` results simply provide the time taken when no combinatorial Morse theoretic pre-processing is performed while holding all other implementation-specific factors constant. Thus, if more efficient implementations of `SP` exist, then we expect that preprocessing with `MorseReduce` will improve the performance of those implementations as well.

While the results of Theorem 5.1 apply to input filtrations over any PID $\mathbf{R}$, the usual computation of persistence intervals via `SP` requires $\mathbf{R}$ to be a field. In the experimental results that follow, we have performed all reductions over $\mathbb{Z}$, but we assume $\mathbf{R} = \mathbb{Z}/2\mathbb{Z}$ throughout when applying `SP` to the reduced filtration output by `MorseReduce`. Table 1 demonstrates the performance comparison of computing peristence with and without pre-processing by `MorseReduce`.

The table is arranged as follows: the first column indicates the type of complexes in the filtration (**C**ubical, **S**implicial, **V**ietoris-**R**ips or **M**ovie) while the second column contains the maximum dimension of the cells present in the filtration. The third column contains the number of frames K of each input filtration $\mathcal{F} = \left\{\mathcal{X}^k\right\}_1^K$. The next two columns provide the size (in number of cells) of the filtration before and after Morse reduction. The penultimate column provides the time taken by our implementation of `SP` to compute persistence intervals over $\mathbb{Z}/2\mathbb{Z}$ of the filtration, whereas the last

[2] http://comptop.stanford.edu/u/programs/jplex

[3] http://hg.mrzv.org/Dionysus

column provides the total time taken to first apply `MorseReduce` and then compute the persistence intervals of the reduced filtration with `SP`. DNF indicates that the given algorithm failed to terminate because it ran out of memory. All times are in seconds.

A final note to illustrate the power and scalability of the Morse theoretic approach: the movie datasets were far too large to be held in memory all at once. Our approach involved storing about 30 frames at a time and removing paired cells from all but the last frame. This freed up considerable memory which we used to input the remaining portions of the movies in pieces, each comprising 30 consecutive frames. At each stage we left the last frame unreduced so that the next piece of the movie could be attached to it, and so on. In this way, extremely large and complicated persistence computations may be brought within the scope of commodity hardware. To the best of our knowledge, there is no other publically available technique which yields persistence intervals of a large filtration from such local computations without ever holding all the cells in memory at once.

Acknowledgments The authors were partially supported by NSF Grants DMS-0915019 and CBI-0835621 and by contracts from DARPA and AFOSR.

References

1. Carlsson, G.: Topology and data. Bull. Am. Math. Soc. (N.S.) 46(2), 255–308 (2009)
2. Chari, M.K.: On discrete Morse functions and combinatorial decompositions. In: Proc. Formal Power Series and Algebraic Combinatorics (Vienna) (1997). Discrete Math. **217**(1–3), 101–113 (2000)
3. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. Discrete Comput. Geom. **37**(1), 103–120 (2007)
4. Delfinado, C.J.A., Edelsbrunner, H.: An incremental algorithm for Betti numbers of simplicial complexes on the 3-sphere: Grid generation, finite elements, and geometric design. Comput. Aided Geom. Des. **12**(7), 771–784 (1995)
5. Dumas, J.-G., Heckenbach, F., Saunders, D., Welker, V.: Computing simplicial homology based on efficient Smith normal form algorithms. In: Joswig, M., Takayama, N. (eds.) Algebra, Geometry, and Software Systems, pp. 177–206. Springer, Berlin (2003)
6. Edelsbrunner, H., Harer, J.: Persistent homology—a survey. In: Surveys on Discrete and Computational Geometry, Contemporary Mathematics, vol. 453, pp. 257–282. American Mathematical Society, Providence (2008)
7. Edelsbrunner, H., Harer, J.L.: Computational Topology. American Mathematical Society, Providence, RI. An introduction (2010)
8. Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. In: Discrete and Computational Geometry and Graph Drawing (Columbia, SC) (2001). Discrete Comput. Geom. **28**(4), 511–533 (2002)
9. Forman, R.: Morse theory for cell complexes. Adv. Math. **134**, 90–145 (1998)
10. Ghrist, R.: Barcodes: the persistent topology of data. Bull. Am. Math. Soc. (N.S.), **45**(1), 61–75 (2008)
11. Gunther, D., Reininghaus, J., Wagner, H., Hotz, I.: Memory efficient computation of persistent homology for 3d image data using discrete Morse theory. In: Proceedings of Conference on Graphics, Patterns and Images, 24 (to appear)
12. Hafner, J., McCurley, K.: Asymptotically fast triangularization of matrices over rings. SIAM J. Comput. **20**(6), 1068–1083 (1991)
13. Harker, S., Mischaikow, K., Mrozek, M., Nanda, V., Wagner, H., Juda, M., Dlotko, P.: The efficiency of a homology algorithm based on discrete Morse theory and coreductions. In: Proceedings of the 3rd International Workshop on Computational Topology in Image Context, Image A, vol. 1, pp. 41–47 (2010)
14. Harker, S., Mischaikow, K., Mrozek, M., Nanda, V.: Discrete Morse theoretic algorithms for computing homology of complexes and maps. Found. Comput. Math. (2013). doi:10.1007/s10208-013-9145-0

15. Joswig, M., Pfetsch, M.: Computing optimal Morse matchings. SIAM J. Discrete Math **20**(1), 11–25 (2006)
16. Kozlov, D.: Discrete Morse theory for free chain complexes. C. R. Math. **340**, 867–872 (2005)
17. Kozlov, D.: Combinatorial Algebraic Topology. Algorithms and Computation in Mathematics, vol. 21. Springer, Berlin (2008)
18. Lefschetz, S.: Algebraic Topology. American Mathematical Society Colloquium Publications, vol. 27. American Mathematical Society, New York (1942)
19. Lewiner, T.: Geometric discrete Morse complexes. Ph.D. Dissertation. Department of Mathematics, PUC-Rio. http://thomas.lewiner.org/pdfs/tomlew_phd_puc.pdf (2005)
20. Lewiner, T., Lopes, H., Tavares, G.: Toward optimality in discrete Morse theory. Exp. Math. **12**(3), 271–286 (2003)
21. Milosavljevic, N., Morozov, D., Skraba, P.: Zigzag persistent homology in matrix multiplication time. In: Proceedings of the 27th Annual ACM Symposium on Computational Geometry (SCG'11), pp. 216–225, Paris (2011)
22. Morozov, D.: Persistence Algorithm Takes Cubic Time in the Worst Case. BioGeometry News, Department of Computer Science, Duke University, Durham (2005)
23. Mrozek, M., Batko, B.: The coreduction homology algorithm. Discrete Comput. Geom. **41**, 96–118 (2009)
24. Mrozek, M., Wanner, T.: Coreduction homology algorithm for inclusions and persistent homology. Comput. Math. Appl. **60**(10), 2812–2833 (2010)
25. Munkres, J.R.: Elements of Algebraic Topology. The Benjamin/Cummings Publishing Company, Inc., Menlo Park (1984)
26. Perseus, the Persistent Homology Software. http://www.math.rutgers.edu/vidit/perseus
27. Robins, V., Wood, P.J., Sheppard, A.P.: Theory and algorithms for constructing discrete Morse complexes from grayscale digital images. IEEE Trans. Pattern Anal. Mach. Intell., 1–14 (2010)
28. Saunders, B.D., Wan, Z.: Smith normal form of dense integer matrices, fast algorithms into practice. In: The International Symposium on Symbolic and Algebraic Computation, pp. 274–281 (2004)
29. Spanier, E.H.: Algebraic Topology. McGraw-Hill Book Co., New York (1966)
30. Storjohann, A.: Nearly optimal algorithms for computing Smith normal forms of integer matrices, vol. 96. In: Proceedings of ISSAC, pp. 267–274 (1996)
31. Strassen, V.: Gaussian elimination is not optimal. Numer. Math. **13**, 354–356 (1969)
32. Tucker, A.W.: Cell spaces. Ann. Math. **37**, 92–100 (1936)
33. Wagner, H., Chen, C., Vucini, E.: Efficient computation of persistent homology for cubical data. In: Topological Methods in Data Analysis and Visualization II, pp. 91–106 (2012)
34. Zomorodian, A.: Topology for Computing, Cambridge Monographs on Applied and Computational Mathematics, vol. 16. Cambridge University Press, Cambridge, MA (2005)
35. Zomorodian, A., Carlsson, G.: Computing persistent homology. Discrete Comput. Geom. **33**(2), 249–274 (2005)