2011 International Symposium on Nonlinear Theory and its Applications
NOLTA2011, Kobe, Japan, September 4-7, 2011

An Efficient Morse Theoretic Preprocessor for Computing Persistent Homology

Vidit Nanda†

†Department of Mathematics, Rutgers University
110 Frelinghuysen Rd., Piscataway, NJ 08854-8019 USA
Email: vidit@math.rutgers.edu

Abstract—We provide an efficient algorithm utilizing discrete Morse theory to dramatically reduce the size of a persistence complex while preserving its persistent homology groups. Significant gains in both time taken and memory consumed are observed when we compare to the existing methods of computing persistent homology. Our technique is not restricted to cubical, simplicial or even CW complexes.

1. Introduction

Persistent homology has become ubiquitous in the topological analysis of data since its inception in [3] and subsequent refinement in [13]. However, the standard algorithm for the computation of persistent homology groups of a persistence complex $\mathcal{F}$ with n cells over a field $\mathbf{F}$ as described in [13] relies on the Smith Normal Form over the PID $\mathbf{F}[t]$. The optimal known implementation of the Smith Normal Form over the integers, for instance, is of super-cubical complexity in n (see [10]). For many applications involving large persistent complexes, the standard algorithm is unfeasible in terms of both time and memory costs.

Some progress has been made in efficient computation of persistent homology, for instance in [12], but the efficiencies obtained via this method are applicable only to cubical datasets with little hope for generalization to other types of complexes. Recent approaches reduce the complexity to matrix multiplication time as in [8], but practical implementations of the Coppersmith-Winograd algorithm (see [2]) for matrix multiplication remain elusive.

An alternative to these methods is to reduce the persistence complex directly without changing the persistent homology groups. If we have a single complex $\mathcal{X}$ then discrete Morse theory from [4] provides an excellent theoretical tool for reducing the size of the complex. There are, however, large gaps in terms of practical implementations. Some progress has been made towards finding optimal discrete Morse functions on triangulated 2-manifolds in [7] but we know little about generalizing such results to higher dimensions and more general complexes.

Even if we are given a discrete Morse function $\mu : \mathcal{X} \to \mathbf{R}$, the process of computing homology requires summing multiplicities over all gradient paths linking critical cells of adjacent dimension (see [4, Def. 8.6]), but enumerating all these paths is a combinatorially explosive proposition. We use an extended version (see [5]) of the theory of coreduction from [9] to construct discrete Morse persistent complexes efficiently from any persistence complex. We also implement a caching strategy to avoid the exponential cost of summing over paths.

The rest of this paper is arranged as follows. In section 2 we provide the basic definitions and prior results that we have used in our work. In section 3 we describe our main algorithm. We also provide certain details of implementing this algorithm in C++ along with a comparison to existing software.

2. Preliminary Definitions and Results

Let $\mathbf{R}$ be a principal ideal domain throughout the paper.

2.1. Persistent Complexes and Persistent Homology

The following definition pertains to combinatorial complexes of cells as developed by Tucker and Lefschetz (see [11] and [6]).

A *complex* over $\mathbf{R}$ is a finite graded set $\mathcal{X} = \bigsqcup_q \mathcal{X}_q$ with each $\xi \in \mathcal{X}_q$ being called a *cell* of *dimension* q along with a function $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$ called an *incidence function* such that for any η and $\xi \in \mathcal{X}$,

1. $\kappa(\eta, \xi) \neq 0$ implies $\dim \eta = \dim \xi + 1$, and
2. $\sum_{\zeta \in \mathcal{X}} \kappa(\eta, \zeta) \cdot \kappa(\zeta, \xi) = 0$.

Consider $\mathcal{X}' \subset \mathcal{X}$ such that for any $\eta \in \mathcal{X}'$ we have $\{\xi \in \mathcal{X} \mid \kappa(\eta, \xi) \neq 0\} \subset \mathcal{X}'$. Such a subset $\mathcal{X}'$ is called a (closed) *subcomplex* of $\mathcal{X}$. We see immediately that the restriction of κ to $\mathcal{X}' \times \mathcal{X}'$ defines an incidence function on $\mathcal{X}'$.

We denote each free module $\mathbf{R}(\mathcal{X}_q)$ by $C_q(\mathcal{X})$ and call its elements the *q-chains*. The basis elements of this module are precisely the cells ξ in $\mathcal{X}_q$. We will distinguish the q chain corresponding to ξ from the cell ξ by denoting the chain as $\widehat{\xi}$. The incidence function κ induces the associated *boundary operators* $\partial_q : C_q(\mathcal{X}) \to C_{q-1}(\mathcal{X})$ via

$$\partial(\widehat{\eta}) = \sum_{\xi \in \mathcal{X}} \kappa(\eta, \xi)\widehat{\xi} \tag{1}$$

and the properties of κ ensure the familiar property $\partial_{q-1} \circ \partial_q \equiv 0$. Finally, the qth homology group of $\mathcal{X}$ is defined to be the quotient module $H_q(\mathcal{X}) = \ker \partial_q / \text{img } \partial_{q+1}$. The computation of homology groups involves expressing the matrix representation of each ∂_q with respect to the bases $\mathcal{X}_q$ and $\mathcal{X}_{q-1}$ in Smith Normal Form.

Now consider a sequence $\mathcal{F} = \left\{\mathcal{X}^k\right\}$ of complexes so that each $\mathcal{X}^k$ is a subcomplex of the subsequent $\mathcal{X}^{k+1}$. This entire sequence is called a *persistence complex*. Let κ^k denote the incidence product of each $\mathcal{X}^k$ with the understanding that $\kappa^{k+1}|_{\mathcal{X}^k \times \mathcal{X}^k} \equiv \kappa^k$. Therefore, each boundary operator ∂^{k+1} also restricts to ∂^k on $\mathcal{X}^k$. That is, each inclusion map $i^k : C(\mathcal{X}^k) \to C(\mathcal{X}^{k+1})$ is a *chain map*. The *p-persistent q-th homology group* of $\mathcal{X}^k$ as defined in [13] is the quotient module

$$H_q^p(\mathcal{X}^k) = \frac{\ker \partial_q^k}{\ker \partial_q^k \cap \text{img } \partial_{q+1}^{k+p}} \tag{2}$$

where the quotient makes sense when each module in sight is regarded as a submodule of $C_q(\mathcal{X}^{k+p})$. It has been shown in [13] that if $\mathbf{R}$ is a field then each generator of the persistent homology groups of $\mathcal{F}$ may be represented by a family of intervals of the type (k_1, k_2) with $k_1 < k_2$ where k_1 is the first value of k for which the generator is an element of $\ker \partial^k$ and k_2 is the first value of k where the generator is an element of $\text{img } \partial^k$.

In the subsequent sections, we describe how discrete Morse theory is used to reduce each $\mathcal{X}^k$ to an associated Morse complex $\mathcal{A}^k$ via chain equivalent maps $\psi_q^k : C_q(\mathcal{X}^k) \to C_q(\mathcal{A}^k)$ and $\phi_q^k : C_q(\mathcal{X}^k) \to C_q(\mathcal{A}^k)$ such that

1. Each $\mathcal{A}^k$ is a subcomplex of $\mathcal{A}^{k+1}$ with a given inclusion map $j^k : C(\mathcal{A}^k) \to C(\mathcal{A}^{k+1})$.
2. The chain maps commute with inclusions. That is, $\psi^{k+1} \circ i^k \equiv j^k \circ \psi^k$ and $\phi^{k+1} \circ j^k \equiv i^k \circ \phi^k$.

The first condition guarantees that the Morse complexes $\left\{\mathcal{A}^k\right\}$ form a persistence complex which we call $\mathcal{M}$, and the second condition guarantees that the persistent homology groups of $\mathcal{F}$ and $\mathcal{M}$ are isomorphic. Some details have been omitted here due to space considerations, but these will be supplied in an upcoming publication.

2.2. Excision of Cell Pairs

The relationship between a cell complex $\mathcal{X}$ and its Morse complex $\mathcal{A}$ has been described in [4]. Following the reinterpretation in [1], we partition the cells of $\mathcal{X}$ into three categories $\mathcal{A}$, $\mathcal{K}$ and $\mathcal{Q}$. The cells $\mathcal{A} \subset \mathcal{X}$ which eventually form the reduced Morse complex are called *critical*. The other two cell categories, denoted $\mathcal{K}$ and $\mathcal{Q}$ are required to be bijective via a correspondence $p : \mathcal{K} \to \mathcal{Q}$ such that for each K in $\mathcal{K}$ the incidence $\kappa(K, p(K))$ is a unit in $\mathbf{R}$. Such a decomposition $(\mathcal{A}, p : \mathcal{K} \to \mathcal{Q})$ of a complex $\mathcal{X}$ is known as an *acyclic matching*.

Each pair of cells $K \in \mathcal{K}$ and $Q = p(K) \in \mathcal{Q}$ may then be removed from the complex altogether. The unit incidence $\kappa(K, Q)$ is used to clear out the column of K and the row of Q in the matrix representation of the boundary operator ∂ via the elementary row and column operations

$$\kappa(\eta, \xi) \leftarrow \kappa(\eta, \xi) - \frac{\kappa(K, \xi) \cdot \kappa(\eta, Q)}{\kappa(K, Q)} \tag{3}$$

Since these operations are admissible moves towards computing the Smith Normal Form, the homology groups of the complex $\mathcal{X}$ are isomorphic to those of the reduced complex $\mathcal{X} \setminus \{K, Q\}$. Proceeding in this fashion, we may remove all the paired cells in $\mathcal{K}$ and $\mathcal{Q}$ from $\mathcal{X}$ without changing the homology groups. The boundary operator on the remaining critical cells $\mathcal{A}$ is denoted Δ and called the Morse boundary operator. It may be directly read off from the boundary matrices after all the paired cells have been removed via the operations described in equation (3).

These row operations and their inverses correspond to well-defined chain maps ϕ_q and ψ_q as described at the end of the preceding section.

2.3. Reduction of Persistence Complexes

Given a persistence complex $\mathcal{F} = \left\{\mathcal{X}^k\right\}$, we will describe the construction of acyclic matchings $(\mathcal{A}^k, p^k : \mathcal{K}^k \to \mathcal{Q}^k)$ of each $\mathcal{X}^k$ such that the obvious inclusion and restriction requirements are satisfied on the critical and paired cells. That is, we have

1. $\mathcal{A}^k \subset \mathcal{A}^{k+1}$, $\mathcal{K}^k \subset \mathcal{K}^{k+1}$ and $\mathcal{Q}^k \subset \mathcal{Q}^{k+1}$.
2. $p^{k+1}|_{\mathcal{K}^k} \equiv p^k$

Note that the second requirement forces the pairings to respect the index k. That is, cells in $\mathcal{X}^{k+1} \setminus \mathcal{X}^k$ may only be paired with each other. These conditions suffice to ensure that the following diagram commutes

$$\begin{array}{ccccc}
\ldots \xrightarrow{i^{k-1}} C(\mathcal{X}^k) & \xrightarrow{i^k} & C(\mathcal{X}^{k+1}) & \xrightarrow{i^{k+1}} \ldots \\
\psi^k \downarrow\uparrow \phi^k & & \psi^{k+1} \downarrow\uparrow \phi^{k+1} & \\
\ldots \xrightarrow{j^{k-1}} C(\mathcal{A}^k) & \xrightarrow{j^k} & C(\mathcal{A}^{k+1}) & \xrightarrow{j^{k+1}} \ldots
\end{array}$$

where ψ^k and ϕ^k are the chain equivalences between each complex $\mathcal{X}^k$ and its Morse complex $\mathcal{A}^k$. From this commutative diagram, we immediately have the two desired properties from the end of section 2.1.

3. Algorithms, Implementation and Results

Before describing any algorithms, we summarize useful notation pertaining to persistence complexes. First, we abuse notation slightly to write $\eta \in \mathcal{F}$ to indicate that η is a cell in some complex $\mathcal{X}^k$ belonging to the persistence complex $\mathcal{F}$. For any cell ξ in $\mathcal{F}$, define

$$\begin{aligned}
\mathsf{bd}(\xi) &= \{\eta \in \mathcal{F} \mid \kappa(\xi, \eta) \neq 0\} \\
\mathsf{cb}(\xi) &= \{\eta \in \mathcal{F} \mid \kappa(\eta, \xi) \neq 0\}
\end{aligned}$$

We call $\mathtt{bd}(\xi)$ the set of *boundary cells* or simply *boundaries* of ξ and $\mathtt{cb}(\xi)$ the set of *coboundary cells* or *coboundaries* of ξ. Note that for a persistence complex $\mathcal{F} = \left\{\mathcal{X}^k\right\}$, the notion of $\mathtt{bd}$ is well defined independent of the index k. That is, if $\xi \in \mathcal{X}^k$ then $\mathtt{bd}(\xi) \subset \mathcal{X}^k$ since we have the subcomplex property and κ restricts as expected to lower values of k. However, a cell ξ may have different $\mathtt{cb}$'s depending on which complex $\mathcal{X}^k$ is being considered. We remove the ambiguity by insisting that $\mathtt{cb}(\xi)$ be the union of $\mathtt{cb}(\xi)$'s over all the nested complexes $\mathcal{X}^k$.

3.1. Coreduction

The coreduction homology algorithm was first introduced in [9] to reduce the size of a single complex $\mathcal{X}$ without altering its (reduced) homology groups. An *elementary coreduction pair* is defined to be a pair of cells K and Q in a complex $\mathcal{X}$ such that $\partial\widehat{K} = u \cdot \widehat{Q}$, where u is a unit in $\mathbf{R}$. Clearly, this is a special case of the cells paired in an acyclic matching. The advantage of restricting attention to coreduction pairs is apparent when one considers the matrix operations in (3). If we excise the pair (K, Q), we only need to update $\kappa(\eta, \xi)$ when both $\kappa(K, \xi)$ and $\kappa(\eta, Q)$ are non-zero. But if Q is the only solution in $\mathcal{X}$ of $\kappa(K, *) \neq 0$, then there is no need to update κ at all and the traditional costs associated with Smith Normal Form computation are not encountered.

The coreduction algorithm relies on excising a cell of minimal dimension from a given complex $\mathcal{X}$ and then removing all possible elementary coreduction pairs until no more pairs are found. A method to keep creating and removing coreduction pairs by choosing new minimal cells once the traditional algorithm halts may be found in [5]. The only additional cost comes from computing the boundaries of the newly chosen minimal cells, but the precise formula for the new boundaries comes from discrete Morse theory.

We extend this new coreduction technique from [5] to persistent complexes so that the properties of the commuting diagram from section 2.3 are preserved.

3.2. Main Algorithm

We provide an algorithm to construct a Morse persistence complex from a persistence complex $\mathcal{F} = \left\{\mathcal{X}^k\right\}_1^K$. First, associate to each cell $\xi \in \mathcal{X}^k$ a *descending gradient path* $g(\xi)$ which is an element of $C(\mathcal{A}^k)$. The gradient path for each cell in the filtration $\mathcal{F}$ is initialized to the trivial chain and will be populated as the algorithm proceeds. On termination, the gradient path $g(A)$ for each A in the Morse persistence complex $\mathcal{M}$ will hold the chain corresponding to $\Delta\widehat{A}$, the Morse boundary. Thus, the storage of gradient paths at each step bypasses the cost of summing multiplicities over paths as in [4] completely.

The main idea is to pick as critical a cell of minimum dimension in $\mathcal{X}^k$ for the lowest k such that $\mathcal{X}^k$ is non-empty and then perform the usual coreduction algorithm from [9].

Here is a simple subroutine that allows us to populate the gradient paths of coboundaries of a given cell.

Algorithm: `UpdateGradientPath`
In: $\xi \in \mathcal{F}$; **Out:** Updates $g(\eta)$ for all $\eta \in \mathtt{cb}(\xi)$

```
for each η ∈ cb(ξ)
   if ξ ∈ A^k for some k
      g(η) ← g(η) + κ(η, ξ) · ξ̂
   else
      g(η) ← g(η) + κ(η, ξ) · g(ξ)
   end if
end for
```

Finally, here is our coreduction-based algorithm for reducing $\mathcal{F}$ to $\mathcal{M}$. For simplicity and 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 the persistence complex $\bigcup_k \mathcal{X}^k$ by setting $B^1 = \mathcal{X}^1$ and $B^k = \mathcal{X}^k \setminus \mathcal{X}^{k-1}$ for higher k.

Algorithm: `MorseReduce`
In: $\mathcal{F} = \left\{\mathcal{X}^k\right\}_1^K = \left\{B^k\right\}_1^K$; **Out:** $\mathcal{M} = \left\{\mathcal{A}^k\right\}_1^K$

```
for each k ∈ {1, ..., K}
   while B^k ≠ ∅
      Pick A ∈ B^k of min dimension
      A^k ← A^k ∪ {A};
      updateGradientPath(A)
      B^k ← B^k \ {A}
      Que := Empty Queue of Cells
      Que ← A
      while Que ≠ ∅
         Que → K
         if ∂K̂ = 0
            Que ← cb(K)
         else if ∂K̂ = u · Q̂
         and ∃k* with K, Q ∈ B^k*
            B^k* ← B^k* \ {K}
            Que ← cb(Q)
            g(Q) ← −g(K)/κ(K,Q)
            if dim Q = dim A
               updateGradientPath(Q)
            end if
            Que ← cb(Q)
            B^k* ← B^k* \ {Q}
         end if
      end while
   end while
end for
return {A^k}_1^K
```

As the algorithm picks and excises critical cells A, it updates the gradient paths of cells in $\mathtt{cb}(A)$ and enqueues those cells to check for elementary coreduction pairs that may have been created as a result of the excision. Note that while the critical cells are chosen in order of increasing k, there is no restriction on the k_* corresponding to B^{k_*} from

which the coreduction pairs are removed. Therefore, our algorithm may remove coreduction pairs from later frames when the critical elements are being picked from earlier frames.

3.3. Implementation

An implementation of our main algorithm `MorseReduce` in C++ will be made available soon. The cost of keeping the gradient path chains in memory is somewhat alleviated if we observe that these chains may be safely deleted from each removed pair (K, Q). In fact, $g(K)$ is never required after line 17, and $g(Q)$ can be removed from memory once it has been used to feed the gradient paths of its remaining coboundaries via `UpdateGradientPath` in line 19.

The algorithm can be iterated as long as there are unit-incident cells A, A' with the same k-index in the reduced persistence complex $\mathcal{M}$. That is, the output of the algorithm in its first run may be used as input for its next run and so on until the number of cell pairs removed is zero, at which point the output persistence complex can be fed into the algorithm from [13].

3.4. Results

We demonstrate our results on cubical grids (C), simplicial complexes (S) and Vietoris-Rips (V) type simplicial complexes. Most cubical complexes come from sublevel sets of finite element Cahn-Hilliard simulations and the simplicial complex arises from brain imaging data. The Vietoris-Rips complexes come from experimental fluid flow data. Our largest data set is a movie (M) involving 35 frames, each comprising about $160,000$ 3 dimensional cubes. All computations were performed on an Intel Core i5 machine with 32 GB of available RAM. The comparison is with the standard algorithm for computing persistent homology as found in [13] which we will denote `SP`. Our algorithm from section 3.2 is denoted `MR`. Further comparisons with other algorithms will be provided later.

The following table demonstrates the extent of reductions performed and the time taken. In case of the `MorseReduce` algorithm, the time taken is the sum of the time required to reduce the complex and the time required to compute the homology of the reduced complex.

Type	**Sz**	**Red. Sz**	`SP`	`MR`
C [16;2]	26.12 k	2.35 k	19 s	0.9 s
C [17;2]	1.05 M	11.89 k	13,442 s	7.3 s
S [50;5]	37.86 k	7.13 k	2,983 s	4.8 s
V [99;3]	29.5 k	8.76 k	2,866 s	1.3 s
V [100;2]	2.34 M	86.33 k	1,277 s	7.1 s
M [35;3]	54 M	323	DNF	543 s

The numbers in brackets indicate the largest index K of each persistence complex $\left\{\mathcal{X}_1^k\right\}_1^K$ followed by the top dimension of cells in the complex. DNF indicates that the algorithm failed to terminate because it ran out of memory.

4. Acknowledgements

The author thanks K. Mischaikow for persistent homological guidance. Thanks is also due to an anonymous reviewer whose suggestions have been gratefully incorporated into the final draft.

References

[1] M.K. Chari, *On Discrete Morse Functions and Combinatorial Decompositions*. Discrete Math 217, no. 1-3, 101113 2000.

[2] D. Coppersmith and S. Winograd, *Matrix Multiplication via Arithmetic Progressions*. J. Symb. Comput. 9(3): 251-280 1990.

[3] H. Edelsbrunner, D. Letscher and A. Zomorodian, *A. Topological persistence and simplification*. Discrete Comput. Geom. 28, 511533. 2002.

[4] R. Forman, *Morse Theory for Cell Complexes*. Advances in Mathematics 145, 90-145 1998.

[5] S. Harker, K. Mischaikow, M. Mrozek, V. Nanda *Computing Homology via Preprocessing by Morse Coreductions*. In preparation.

[6] S. Lefschetz, *Algebraic Topology*. American Mathematical Society Colloquium Publications, v. 27. American Mathematical Society, New York 1942.

[7] T. Lewiner, H. Lopes, and G. Tavares *Toward Optimality in Discrete Morse Theory*. Experiment. Math. Volume 12, Number 3, 271-286 2003.

[8] N. Milosavljevic, D. Morozov, P. Skraba, *Zigzag Persistent Homology in Matrix Multiplication Time*. Proceedings of the Annual Symposium on Computational Geometry, 2011.

[9] M. Mrozek and B. Batko, *The Coreduction Homology Algorithm*. Discrete & Computational Geometry 41(1), 96-118 2009.

[10] A. Storjohann, *Near Optimal Algorithms for Computing Smith Normal Forms of Integer Matrices*. Proceedings of the 1996 International Symposium on Symbolic and Algebraic Computation, ISSAC 96, pp. 267274 1996.

[11] E.W. Tucker, *Cell Spaces*. Ann. of Math. (2), pp. 92-100 1936.

[12] H. Wagner, C. Chen and E. Vuini, *Efficient Computation of Persistent Homology for Cubical Data*. Proceedings of the 4th Workshop on Topology-based Methods in Data Analysis and Visualization, 2011.

[13] A. Zomorodian and G. Carlsson, *Computing Persistent Homology*. Discrete Computational Geometry 33(2), pp. 249-274 2005.