Hardness Results for Homology Localization

Chao Chen
Rensselaer Polytechnic Institute
110 8th Street, Troy, NY 12180, United States
Email: chenc3@cs.rpi.edu
Homepage: www.cs.rpi.edu/∼chenc3

Daniel Freedman
Rensselaer Polytechnic Institute
110 8th Street, Troy, NY 12180, United States
Email: freedman@cs.rpi.edu
Homepage: www.cs.rpi.edu/∼freedd

June 17, 2009

Abstract

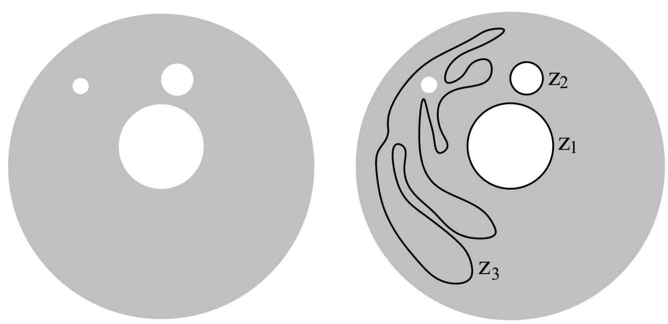
In this paper, we address the problem of localizing homology classes, namely, finding the cycle representing a given class with the most concise geometric measure. We study the problem with different measures: volume, diameter and radius. For volume, that is, the 1-norm of a cycle, we prove the problem is NP-hard to approximate within any constant factor. We also prove that for homology of 2-dimension or higher, the problem is NP-hard to approximate even when the Betti number is $O(1)$. A side effect is the inapproximability proof of the problem of computing the nonbounding cycle with the smallest volume, and computing cycles representing a homology basis with the minimal total volume. As for the other two measures defined by pairwise geodesic distance, we show that the localization problem is NP-hard for diameter, but is polynomial for radius.

1 Introduction

The problem of computing the topological features of a space has recently drawn much attention from researchers in various fields, such as high-dimensional data analysis [6, 24], graphics [21, 7], networks [16] and computational biology [1, 14]. Topological features are often preferable to purely geometric features, as they are more qualitative and global, and tend to be more robust. If the goal is to characterize a space, therefore, features which incorporate topology seem to be good candidates.

While topological features are global, the need to "localize" them has been raised in a variety of applications. In graphics and manifold learning, one wants to detect and remove topological noise such as the small holes and handles that are introduced in data acquisition; this is often done in the context of traditional signal-noise analysis, and finite sampling of continuous spaces [25, 36, 31]. In the area of sensor networks, holes of the coverage region, caused by physical constraints, should be accurately identified and described so as to produce as robust a network as possible [23, 32]. In the study of shape, 3D shapes may be enriched with properties such as curvatures associated with tangent vectors at each tangent plane. The new augmented shape lives in high dimension, whose topological features can be localized and reveal geometric features of the original shape [5].

In this paper, we will address the localization problem. The topological features we use are homology classes over $\mathbb{Z}_2$ field, due to their ease of computation. (Thus, throughout this paper, all the additions are mod 2 additions.)



(a) The space in question is a disk with three holes.

(b) Cycles z_1 and z_2 are well-localized; z_3 is not.

Figure 1: A motivating example.

1.1 Problem Definition

The *localization problem* is finding the smallest representative cycle of a homology class with regard to a given natural criterion of the size of a cycle. The criterion should be deliberately chosen so that the corresponding smallest cycle is concise in not only mathematics but also intuition. Such a cycle is a "well-localized" representative cycle of its class. For example, in Figure 1(b), the cycles z_1 and z_2 are well-localized representatives of their respective homology classes; whereas z_3 is not.

Formally, given an objective function defined on all the d-cycles,

$$\text{cost} : \mathsf{Z}_d(K) \to \mathbb{R},$$

we formalize the localization problem as a combinatorial optimization problem.

Problem 1.1.1 (Localizing Homology.)**.**

INPUT: a simplicial complex K with size n, a d-dimensional nontrivial homology class $h = [z_0]$

OUTPUT: a cycle $z \in h$

MINIMIZE: $\text{cost}(z)$

In this paper, we explore three options of the objective function, $\text{cost}(z)$, i.e. the *volume*, *diameter* and *radius*.

1. **Volume:** the number of simplices in a cycle. The definition can be extended to incorporate weight function defined on simplices.

2. **Diameter:** the maximum of pairwise discrete geodesic distances between vertices of the cycle.

3. **Radius:** the radius of the smallest discrete geodesic ball carrying the cycle.

For convenience, we call the localization problems using these three objective functions LHV, LHD and LHR, respectively.

1.2 Contributions

1. For the volume function,

   - We prove LHV is NP-hard to approximate within any constant factor when the Betti number of the pertinent homology is $\Theta(n)$;

- For homology of 2-dimension or higher, we prove a simpler problem, finding the nonbounding cycle with the minimal volume (MVNC), is NP-hard to approximate within any constant factor. This leads to the inapproximability of LHV with fixed Betti number and computing the homology cycle basis with minimal total volume;
- We provide a polynomial algorithm for a special case, using min-cut algorithm.

2. Prove that LHD is NP-hard to compute.
3. Prove that LHR is polynomial time computable by reciting our previous polynomial algorithm.

2 Preliminaries

Homology Groups. We briefly describe some background knowledge from algebraic topology. Please refer to [30] for more details. For simplicity, we restrict our discussion to the combinatorial framework of simplicial homology over $\mathbb{Z}_2$ field.

Within a given simplicial complex K, a *d-chain* is a formal sum of d-simplices in K,

$$c = \sum_{\sigma \in K} a_\sigma \sigma, a_\sigma \in \mathbb{Z}_2.$$

All the d-chains form the *group of d-chains*, $\mathsf{C}_d(K)$. The *boundary* of a d-chain is the sum of the $(d-1)$-faces of all the d-simplices in the chain. The boundary operator $\partial_d : \mathsf{C}_d(K) \to \mathsf{C}_{d-1}(K)$ is a group homomorphism.

A *d-cycle* is a d-chain without boundary. The set of d-cycles forms a subgroup of the chain group, which is the kernel of the boundary operator, $\mathsf{Z}_d(K) = \ker(\partial_d)$. A *d-boundary* is the boundary of a $(d+1)$-chain. The set of d-boundaries forms a group, which is the image of the boundary operator, $\mathsf{B}_d(K) = \operatorname{img}(\partial_{d+1})$. It is not hard to see that a d-boundary is also a d-cycle. Therefore, $\mathsf{B}_d(K)$ is a subgroup of $\mathsf{Z}_d(K)$. A d-cycle which is not a d-boundary, $z \in \mathsf{Z}_d(K) \backslash \mathsf{B}_d(K)$, is a *nonbounding cycle*. In our case, the coefficients belong to a field, namely $\mathbb{Z}_2$; when this is the case, the groups of chains, boundaries and cycles are all vector spaces. Note that this is not true when the homology is over a ring which is not a field, such as $\mathbb{Z}$.

The *d-dimensional homology group* is defined as the quotient group $\mathsf{H}_d(K) = \mathsf{Z}_d(K)/\mathsf{B}_d(K)$. An element in $\mathsf{H}_d(K)$ is a *homology class*, which is a coset of $\mathsf{B}_d(K)$, $[z] = z + \mathsf{B}_d(K)$ for some d-cycle $z \in \mathsf{Z}_d(K)$. If z is a d-boundary, $[z] = \mathsf{B}_d(K)$ is the identity element of $\mathsf{H}_d(K)$. Otherwise, when z is a nonbounding cycle, $[z]$ is a *nontrivial homology class* and z is called a *representative cycle* of $[z]$. Cycles in the same homology class are *homologous* to each other, which means their difference is a boundary.

The dimension of the homology group, which is referred to as the *Betti number*,

$$\beta_d = \dim(\mathsf{H}_d(K)) = \dim(\mathsf{Z}_d(K)) - \dim(\mathsf{B}_d(K)).$$

As the dimension of the chain group is upper bounded by the cardinality of K, n, so are the dimensions of $\mathsf{B}_d(K)$, $\mathsf{Z}_d(K)$ and $\mathsf{H}_d(K)$. The Betti number can be computed with a reduction algorithm based on row and column operations of the boundary matrices [30]. Various reduction algorithms have been devised for different purposes [26, 20, 37].

A *homology basis* is a set of β_d classes generating the group $\mathsf{H}_d(K)$. We call a set of β_d nonbounding cycles representing a homology basis a *homology cycle basis*. Any d-cycle can be written as the linear combination of a homology cycle basis and boundaries.

Note that since the field is $\mathbb{Z}_2$, the set of d-chains is in one-to-one correspondence with the set of subsets of d-simplices. A d-chain corresponds to a n_d-dimensional vector, whose nonzero entries correspond to the included d-simplices. Here n_d is the number of d-simplices in K. Computing the boundary of a d-chain corresponds to multiplying the chain vector with a boundary matrix $[b_1, ..., b_{n_d}]$, whose column vectors are boundaries of d-simplices in K. By slightly abusing notation, we call the boundary matrix ∂_d.

We call a subset of simplices of a given simplicial complex a *subcomplex*, if this subset itself is a simplicial complex. The following notation will prove convenient. We say that a d-chain $c \in \mathsf{C}_d(K)$ is *carried by* a subcomplex K_0 when all the d-simplices of c belong to K_0, formally, $c \subseteq K_0$. We denote $\mathrm{vert}(K)$ as the set of vertices of the simplicial complex K, vert(c) as that of the chain c. Denote $|K|$ as the underlying space of K, $|c|$ as that of the chain c.

Replacing simplices by their continuous images in a given topological space gives singular homology. The simplicial homology of a simplicial complex is naturally isomorphic to the singular homology of its geometric realization. This implies, in particular, that the simplicial homology of a space does not depend on the particular simplicial complex chosen for the space. In figures of this paper, we often ignore the simplicial complex and only show the continuous images of chains.

The Discrete Geodesic Distance. To introduce the diameter and radius functions, we need a notion of distance. As we will deal with a simplicial complex K, it is most natural to introduce a discrete metric, and corresponding distance functions. We define the *discrete geodesic distance* from a vertex $p \in \mathrm{vert}(K)$, $f_p : \mathrm{vert}(K) \to \mathbb{R}$, as follows. Suppose each edge in K has a nonnegative weight, for any vertex $q \in \mathrm{vert}(K)$, $f_p(q) = \mathrm{dist}(p, q)$ is the length of the shortest path connecting p and q, in the 1-skeleton of K. We may then extend this distance function from vertices to higher dimensional simplices naturally. For any simplex $\sigma \in K$, $f_p(\sigma)$ is the maximal function value of the vertices of σ, $f_p(\sigma) = \max_{q \in \mathrm{vert}(\sigma)} f_p(q)$. This extension has the same effect as linearly interpolating the function on the interiors of the simplices (the sublevel sets of the two extensions are homotopy equivalent). Finally, we define a *geodesic ball* B_p^r, $p \in \mathrm{vert}(K)$, $r \geq 0$, as the subset of K, $B_p^r = \{\sigma \in K \mid f_p(\sigma) \leq r\}$. It is straightforward to show that these subsets are in fact subcomplexes.

Terminology from Coding Theory. We focus on binary linear codes and thus only use matrices over $\mathbb{Z}_2$ field. For consistency, we switch the roles of the row and column indices from the standard definition. Please refer to [29] for details.

Given an $m \times k$ ($m > k$) full rank matrix A, we define a *linear code* as the k-dimensional column space of A, namely, span(A). Each element of the linear code is called a *codeword*. This matrix is called the *generator matrix* as it is a basis of the linear code. By slightly abusing notation, we call a full rank matrix $A^\perp$ the *parity-check matrix* if its nullspace is the linear code. Given a generator matrix A, $A^\perp$ may be computed in polynomial time.

The Hardness of Approximability and Strict Reductions. We will prove several optimization problems are NP-hard to approximate within any constant factor. Relevant definitions will be presented in this section. Please see [4] for more details.

An *NP optimization problem* Π is a four-tuple $(\mathcal{I}, \mathrm{Sol}, m, \mathrm{opt})$ such that

1. $\mathcal{I}$ is the set of *instances* of Π; any $I \in \mathcal{I}$ can be recognized in time polynomial with its size, $\mathrm{card}(I)$;

2. given $I \in \mathcal{I}$, $\mathrm{Sol}(I)$ denotes the set of *feasible solutions* of I; for every $S \in \mathrm{Sol}(I)$, $\mathrm{card}(S)$ is polynomial in $\mathrm{card}(I)$; given any S polynomial in $\mathrm{card}(I)$, one can decide in polynomial time whether $S \in \mathrm{Sol}(I)$;

3. given $I \in \mathcal{I}$ and $S \in \mathrm{Sol}(I)$, $m(I, S)$ denotes the value of S; m is polynomially computable and is commonly called *objective function*;

4. $\mathrm{opt} \in \{\min, \max\}$ indicates the *type* of optimization problem.

Given an NP optimization problem, Π, for an instance I and one of its feasible solutions, $S \in \mathrm{Sol}(I)$, we define the *performance ratio*, $\rho_\Pi(I, S)$, as the ratio of the value $m(I, S)$ (assume $m(\cdot, \cdot) \geq 0$) over the value of the optimum solution, formally,

$$\rho_\Pi(I, S) = \frac{m(I, S)}{m(I, S^*(I))}$$

where $S^*(I)$ is the optimum solution of I. The quality of a polynomial approximation algorithm, A, is measured by the *approximation ratio* $\rho_A(I) = \rho_\Pi(I, A(I))$. For minimization problems, therefore, the approximation ratio is in $[1, \infty)$, while for maximization problems it is in $[0, 1]$.

An NP optimization problem Π belongs to the class APX if there exists a polynomial approximation algorithm A and a value $r \in \mathbb{Q}$ such that given any instance I of Π, $\rho_A(I) \leq r$ (resp. , $\rho_A(I) \geq r$) if Π is a minimization problem (resp. , a maximization problem). In such case, A is called an r-approximation algorithm of Π.

Given two problems Π_1 and Π_2, we *reduce* Π_1 to Π_2 by providing two polynomial time computable functions f and g that satisfy

- $f : \mathcal{I}_{\Pi_1} \to \mathcal{I}_{\Pi_2}$ such that $\forall I_1 \in \mathcal{I}_{\Pi_1}$, $f(I_1) \in \mathcal{I}_{\Pi_2}$; in other words, given an instance I_1 in Π_1, f allows to build an instance $I_2 = f(I_1)$ in Π_2;
- $g : \mathcal{I}_{\Pi_1} \times \mathrm{Sol}_{\Pi_2} \to \mathrm{Sol}_{\Pi_1}$ such that, for all $(I_1, S_2) \in \mathcal{I}_{\Pi_1} \times \mathrm{Sol}_{\Pi_2}(f(I_1))$, $g(I_1, S_2) \in \mathrm{Sol}_{\Pi_1}(I_1)$; in other words, starting from a solution S_2 of the instance $I_2 = f(I_1)$, g determines a solution $S_1 = g(I_1, S_2)$ of the initial instance I_1.

When both problems are minimization problems. We say the reduction is *strict* if in addition, for any instance $I_1 \in \mathcal{I}_{\Pi_1}$ and any feasible solution of $f(I_1)$, $S_2 \in \mathrm{Sol}_{\Pi_2}(f(I_1))$, the performance ratios satisfy

$$\rho_{\Pi_2}(f(I_1), S_2) \geq \rho_{\Pi_1}(I_1, g(I_1, S_2)). \tag{1}$$

Formally, we say $\Pi_1 \leq_S \Pi_2$. It is not hard to see that the strict reduction preserves the membership of APX. The following lemma will be useful for our inapproximability proof.

Lemma 2.0.1. *If* $\Pi_1 \leq_S \Pi_2$ *and* $\Pi_1 \notin \mathrm{APX}$*, then* $\Pi_2 \notin \mathrm{APX}$.

In other words, if Π_1 is strictly reducible to Π_2 and cannot be approximated within any constant factor, neither can Π_2.

3 Related Work

Researchers have been interested in localizing 1-dimensional homology classes with the minimal volume cycle, namely, the shortest representative 1-cycle. Using Dijkstra's shortest path algorithm, Erickson and Whittlesey [22] computed the *shortest homology basis*, namely, the 1-dimensional homology cycle basis whose elements have the minimal total volume. The authors also showed how the idea carries over to finding the optimal generators of the first fundamental group, though the proof is considerably harder in this case.

This polynomial algorithm cannot localize an arbitrarily given class. To fill this void, Chambers et al. [10] devised an algorithm to localize a given class. Their method precomputes the shortest representative cycles of all $2^{\beta_1} - 1$ nontrivial classes, and thus, is exponential to the 1-dimensional Betti number.

It has been demonstrated that when $\beta_1 = \Theta(n)$, localizing a given 1-dimensional class with its shortest cycle is NP-hard, no matter the topological space is a manifold with [11] or without boundary [10]. However, Chambers et al. [9] showed that this is not the case for 1-dimensional homology over real or integer coefficients.

For completeness, we refer to some related works which compute a single or a set of non-trivial cycles satisfying certain topological and geometrical restrictions on 2-manifolds [21, 28, 8, 15].

Due to the difficulties in localizing with the minimal volume criterion, researchers focused on other criteria or heuristics. Some researchers computed 1-dimensional cycles closely related to handles which are much more meaningful in low dimensional applications such as graphics and CAD. Guskov and Wood [25, 36] detected small handles of a 2-manifold using the Reeb graph of the manifold. Given a 2-manifold embedded in S^3, Dey et al. [17] computed these handle-related cycles by computing the deformation retractions of the two components of the the embedding space bounded by the given 2-manifold. A recent extension [18] improved their result based on geometric heuristics and the persistent homology. Their work facilitates handle detection in real applications. However, the computed 1-cycles are not guaranteed to be geometrically concise.

All the aforementioned works are restricted to 1-dimensional homology. Zomorodian and Carlsson [38] took a different approach to solving the localization problem for general dimension. Their method starts with a topological space and a cover, which is a set of spaces whose union contains the original space. They computed a homology basis and localized classes of it, using tools from algebraic topology and persistent homology. However, both the quality of the localization and the complexity of the algorithm depend strongly on the choice of cover; there is, as yet, no suggestion of a canonical cover.

Chen and Freedman [12, 13] presented a polynomial algorithm for localizing a homology class of general dimension with the radius function (LHR). Their algorithm can also compute a homology cycle basis with the minimal total radius. The cycle with the minimal radius, however, could be quite complicated in terms of geometry. Please see Section 6 for detailed discussion.

4 Volume

The first choice of the objective function is volume.

Definition 4.0.2 (Volume)**.** *The volume of a cycle is the number of its simplices,* $\mathrm{vol}(z) = \mathrm{card}(z)$.

For example, the volume of a 1-cycle, a 2-cycle and a 3-cycle are the numbers of their edges, triangles and tetrahedra, respectively. The cycle with the smallest volume, denoted as z_v, agrees intuitively with the notion of a "well-localized" cycle. The problem of localizing a homology class with its minimal volume cycle, z_v, has been proven to be NP-hard [11, 10]. In this paper, we prove that this problem is even NP-hard to approximate.

More generally, we can extend the the volume definition to be the sum of the weights assigned to simplices of the cycle, given an arbitrary weight function, $w : K \to \mathbb{R}$, defined on all the simplices of K, formally,

$$\mathrm{vol'}(z) = \sum_{\sigma \in z} w(\sigma).$$

Computing z_v using this general volume definition is at least as hard as using Definition 4.0.2, which is in fact a special case (when $w(\sigma) = 1$, $\forall \sigma \in K$). Therefore, in this paper, we will only treat the unweighted volume function.

Results. There are some existing hardness results, when the homology classes in question are 1-dimensional.

1. When the Betti number, $\beta_1 = \Theta(n)$, LHV has been proven to be NP-hard by polynomial reductions from a special case of MAX-2SAT [11] and MIN-CUT with negative edge weights [10].

2. Chambers et al [10] provided a polynomial algorithm when β_1 is fixed. The algorithm computes the shortest representative cycle for each of the $2^{\beta_1} - 1$ nontrivial classes.

3. Erickson and Whittlesey [22] computed in polynomial time the homology cycle basis with the minimal total volume, even when $\beta_1 = \Theta(n)$.

All these existing results are about 1-dimensional homology. We will study whether LHV is difficult in general dimension, and more importantly, how difficult it is.

The existing results suggest that the problem might be easier if we assume fixed Betti number, or if we compute the homology cycle basis with the minimal total volume instead. Therefore, we would also like to find out how difficult these problems could be. We prove hardness result for a simpler problem, computing the nonbounding cycle with the minimal volume, denoted as MVNC, which in turn shows that all the problems we are interested in are NP-hard to approximate when the homology is 2-dimensional or higher.

For clearance, we list all the new results as follows.

1. When the homology in question is 1-dimensional or higher and the Betti number is $\Theta(n)$, it is NP-hard to approximate LHV within any constant factor (Theorem 4.1.4).
2. Concerning the problem MVNC, we have the following results:
   - When the homology in question is 2-dimensional or higher, we prove that MVNC is NP-hard to approximate within any constant factor even with fixed Betti number (Theorem 4.2.6).
   - A polynomial algorithm for MVNC for a special case: when the pertinent space is embedded in $\mathbb{R}^N$ and the pertinent homology is $(N-1)$-dimensional.

4.1 LHV is NP-hard to approximate within any constant factor

We prove by a strict reduction from the nearest codeword problem (NCP), which cannot be approximated within any constant factor [3]. Problems used in previous reductions [11, 10] have constant approximation ratios, and thus cannot be used for our proof.

Problem 4.1.1 (Nearest Codeword Problem)**.**

INPUT: an $m \times k$ *generator matrix* A *over* $\mathbb{Z}_2$ *and a vector* $y_0 \in \mathbb{Z}_2^m \setminus \mathrm{span}(A)$.

OUTPUT: a vector $x \in \mathbb{Z}_2^k$

MINIMIZE: the Hamming distance between Ax *and* y_0

This is equivalent to finding the vector $y = Ax + y_0$ with the minimal Hamming weight.

Lemma 4.1.2. LHV *cannot be approximated within any constant factor.*

Proof. We prove by a strict reduction from NCP, namely, NCP $\leq_S$ LHV.

Given an instance of NCP, namely, a generator matrix A and a vector y_0, we first construct a cell complex, T, whose 2-dimensional boundary matrix is A. T has m 1-cells and k 2-cells corresponding to the m rows and k columns of A. Each 1-cell is a 1-dimensional cycle. Each 2-cell is a pipe with multiple openings. See Figure 2 for an example. Please note that we are abusing notation when we call T a cell complex, as these cells may mot be homeomorphic to closed balls.

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

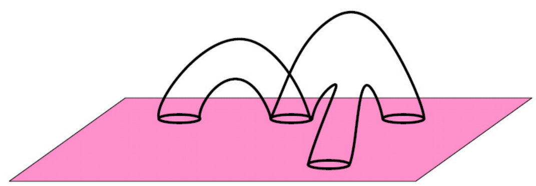


Figure 2: A 4×2 generator matrix A and the constructed cell complex, T.

As each 1-chain of T is a 1-cycle, it is not hard to see that NCP is identical to the problem of computing the minimal volume representative cycle of a given 1-dimensional class of T, denoted as LHV_T. However, LHV_T is different from LHV, whose input is a simplicial complex which is supposed to be a triangulation of a topological space. Next, we embed T in Euclidean space $\mathbb{R}^4$, and then triangulate $|T|$ into a simplicial complex K. With this construction, we will strictly reduce LHV_T to LHV.

The embedding is an analog of book embedding an arbitrary graph in $\mathbb{R}^3$ [34]. Take a 2-dimensional plane P in $\mathbb{R}^4$. There are infinitely many 3-dimensional subspaces passing P. We embed all the 1-cells of T in the plane P. We embed each 2-cell (pipe) in one of those 3-dimensional subspaces. Since these 3-dimensional subspaces only intersect in P, these 2-cells do not intersect each other except for their boundaries.

We first triangulate each 1-cell of T into t_1 edges, with t_1 fixed and small. For convenience, we call the triangulation of all 1-cells K_1, which is a subcomplex of K. There is a one-to-one correspondence between 1-cycles of T and 1-cycles of K_1, denoted as ϕ. For any

1-cycle of T and it corresponding 1-cycle of K_1, the ratio of their volumes is $1/t_1$.

Next, we triangulate interior of 2-cells of T (pipes) while keeping K_1 intact. The triangulation is fine enough so that for any 1-cycle of K, z, homologous to some 1-cycle of K_1, we can compute in polynomial time a cycle z' of K_1, which is homologous to z and has smaller or equal volume.

Our construction provides a polynomial transformation of every instance of LHV_T, (T, y_0), into an instance of LHV, $(K, z_0 = \phi(y_0))$. For any such instance, and any feasible solution $z \in [z_0]$, we transform z into z' and then into a solution of LHV_T, $\phi^{-1}(z')$. This reduction is a strict reduction, as the ratio of the volumes of the optimal solution of the two problems is always $1/t_1$, and

$$\mathrm{vol}(z) \geq \mathrm{vol}(z') = t_1 \, \mathrm{vol}(\phi^{-1}(z')).$$

□

Remark 4.1.3 (Extension to a manifold)**.** *We extend the proof to the case when K is the triangulation of a manifold without boundary. The idea is as follows. Thicken the underlying space of T, then take its boundary as a new topological space. We get a 3-manifold (one less dimension than the ambient space), which can be triangulated similarly. This extension applies to all proofs in the rest of the paper.*

Lemma 4.1.2 is about 1-dimensional homology. We extend the result to homology of any higher dimension.

Theorem 4.1.4. *For any $N \geq 1$,* LHV *for N-dimensional homology cannot be approximated within any constant factor.*

Proof. We show that when $N \geq 2$, LHV for $(N-1)$-dimensional homology can be strictly reduced to LHV for N-dimensional homology, namely, $\mathrm{LHV}_{N-1} \leq_S \mathrm{LHV}_N$. Together with Lemma 4.1.2, the theorem is proved.

Next, we explain briefly the reduction. Take any given simplicial complex of LHV_{N-1} and build a *suspension* of it, namely, two cones of the complex glued together at their base [35]. There is a one-to-one correspondence between the $(N-1)$-dimensional cycle group of the original complex and the N-dimensional cycle group of the new complex. This correspondence also works for the boundary groups. Since the volume of each $(N-1)$-cycle is 1/2 of the volume of its corresponding N-cycle, this is a strict reduction. □

4.2 Computing the minimal volume nonbounding cycle

In the previous section, the simplicial complex we constructed for LHV has $\Theta(n)$ Betti number. It has been revealed for 1-dimensional homology that LHV is polynomial for fixed Betti number, and it is polynomial to compute a homology cycle basis with the minimal total volume. In this section, we show that these two problems are both NP-hard to approximate when the pertinent homology is of 2-dimension or higher.

We prove the inapproximability of a simpler problem, denoted as MVNC.

Problem 4.2.1 (Minimal Volume Nonbounding Cycle Problem.)**.**
INPUT: a simplicial complex K with size n
OUTPUT: a nonbounding d-cycle z
MINIMIZE: $\mathrm{vol}(z)$

We prove that this problem is NP-hard to approximate within any constant factor, even when the pertinent Betti number is 1. This trivially leads to the inapproximability of the problem of LHV with fixed Betti number and the problem of computing the optimal homology cycle basis, as MVNC for $\beta_d = 1$ is a special case of these two problems.

We prove the inapproximability of MVNC by a strict reduction from the minimum distance problem (MDP) for binary linear code, which cannot be approximated within any constant factor [19].

Problem 4.2.2 (Minimum Distance Problem)**.**
INPUT: an $m \times k$ generator matrix A over $\mathbb{Z}_2$.
OUTPUT: a vector $x \in \mathbb{Z}_2^k$.
MINIMIZE: the Hamming weight of Ax.

We compute the corresponding parity-check matrix $A^\perp$. The MDP problem is then equal to finding the nonzero vector in the nullspace of $A^\perp$ with the minimal Hamming weight.

Lemma 4.2.3. *MVNC for 2-dimensional homology cannot be approximated within any constant factor.*

Proof. We prove by a strict reduction from MDP, namely, MDP $\leq_S$ MVNC.

Similar to Lemma 4.1.2, we construct a cell complex, T, with $A^\perp$ as the 2-dimensional boundary matrix. Again, we embed T in $\mathbb{R}^4$ and triangulate it into a simplicial complex. In the triangulation, instead of triangulating 2-cells as fine as possible, we triangulate all of them into a same number of triangles, say t_2. This guarantees a one-to-one correspondence between the group of 2-cycles of K and the nullspace of $A^\perp$. The volume of each 2-cycle is t_2 times of the Hamming weight of its corresponding vector. Since any nonzero 2-cycle is nonbounding, we have a strict reduction. □

In Lemma 4.2.3, the constructed simplicial complex has $\Theta(n)$ Betti number. Next, we prove that the problem is NP-hard to approximate within any constant factor, even when β_2 is 1. We prove this by a reduction from the NCP problem. To facilitate the proof, we rewrite the problem in the following form.

Problem 4.2.4 (Nearest Codeword Problem (revised version))**.**
INPUT: an $m \times k$ generator matrix A over $\mathbb{Z}_2$ and a vector $y_0 \in \mathbb{Z}_2^m$
OUTPUT: a vector $z \in y_0 + \text{span}(A)$
MINIMIZE: the Hamming weight of z

Lemma 4.2.5. *For 2-dimensional homology,* MVNC *is NP-hard to approximate within any constant factor, even when $\beta_2 = 1$.*

Proof. We prove by a strict reduction from NCP.

Given a generator matrix $C = [A, y]$, we compute its parity-check matrix $C^\perp$. Following the method in Lemma 4.2.3, we construct a cell complex T using $C^\perp$ as the 2-dimensional boundary matrix. There is a one-to-one correspondence between the 2-dimensional cycle group of T and $\text{nullspace}(C^\perp) = \text{span}(C)$. Next, for each column vector of A, we find the corresponding 2-cycle in T and seal it with a 3-cell. In the augmented complex, T', the one and only nontrivial 2-dimensional homology class is identical to the coset $y_0 + \text{span}(A)$. Finding the smallest volume nonbounding 2-cycle of T', denoted as $\text{MVNC}_{T'}$, is equal to finding the minimal weight vector in this coset and thus equal to solving the nearest codeword problem. We next strictly reduce $\text{MVNC}_{T'}$ to MVNC.

We first embed T in $\mathbb{R}^4$. Although the 3-cells in $T' \backslash T$ might intersect each other in $\mathbb{R}^4$. By extending the ambient space to $\mathbb{R}^6$ and using an analog of book embedding, we could safely embed them.

We triangulate T' into a simplicial complex K' as follows (in a similar scheme as in Lemma 4.1.2). We first triangulate the 2-skeleton of T', T, into a simplicial complex K, in which all 2-cells are triangulated into the same number of triangles (t_2). There is a one-to-one correspondence between $\mathsf{Z}_2(K)$ and $\mathsf{Z}_2(T) = \mathsf{Z}_2(T')$.

Next, while keeping K intact, we triangulate interior of 3-cells as fine as possible so that for any nonbounding 2-cycle of K', we could always find in polynomial time a nonbounding 2-cycle of K which is homologous to it. As the volume of any 2-cycle of K is t_2 times of that of its corresponding 2-cycle of T', we have a strict reduction from $\text{MVNC}_{T'}$ to MVNC. $\square$

Similar to Theorem 4.1.4, we extend the result to any higher dimension. We conclude this section with the following theorem.

Theorem 4.2.6. MVNC *is NP-hard to approximate within any constant factor for homology of* 2*-dimension or higher, even when the pertinent Betti number is* 1*.*

4.3 A special case

There is, however, a special case in which MVNC can be computed in polynomial time: when K is a N-dimensional complex embedded in $\mathbb{R}^N$ and the pertinent nonbounding cycle is $(N-1)$-dimensional. In this section, we provide a polynomial algorithm, inspired by [27].

We add new N-cells to K to get a new complex K', whose underlying space is $\mathbb{R}^N$. Each new cell covers one component of $\mathbb{R}^N \backslash |K|$. There are $\beta_{N-1} + 1$ new cells, one of which covers the infinity component. The boundary of each new cell is one component of the

$(N-1)$-dimensional boundary of K. Here we are abusing notation again as the new cells may not be homeomorphic to closed balls.

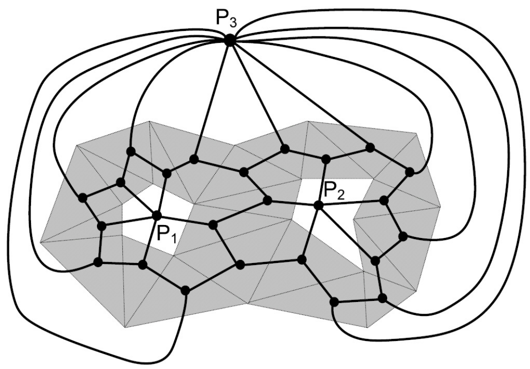


Figure 3: A 2-dimensional simplicial complex embedded in $\mathbb{R}^2$. The dual graph G and G' are drawn in solid lines and vertices. Their difference, $G' \backslash G$, includes vertices p_1, p_2, p_3 and their incident edges.

We use the MIN-CUT algorithm on the dual graphs to solve the problem. The dual graph of K, G, is a subgraph of the dual of K', G'. The set of new vertices $V' \backslash V$ is dual to the set of new N-cells. See Figure 3 for an example when $N = 2$.

We call a cycle *minimal* if none of its non-empty subsets is a cycle. We denote $C(G', G)$ as the set of *minimal edge cuts* (cuts whose subsets are not cuts) of G' which cut G' into two partitions each of which contains at least one vertex of $V' \backslash V$. There is a one-to-one correspondence between the set of minimal nonbounding $(N-1)$-cycles of K and the set of cuts $C(G', G)$. The volume of each cycle is equal to the cardinality of its corresponding cut. As the nonbounding $(N-1)$-cycle with the smallest volume has to be one of the minimal cycles, it can be computed by computing the cut in $C(G', G)$ with the smallest cardinality.

To compute the minimal cardinality cut in $C(G', G)$, we enumerate all pairs of vertices, $(v_1, v_2) \in (V' \backslash V) \times (V' \backslash V)$. Compute the minimal ($v_1$-$v_2$)-cut for each pair. The one with the smallest cardinality is the desired one.

Remark 4.3.1. *The idea can be carried to the weighted volume function, but only if the weight function is non-negative.*

5 Diameter

When LHV is proved to be NP-hard to approximate, we resort to discrete geodesic distance related objective functions, diameter and radius.

Definition 5.0.2 (Diameter)**.** *The diameter of a cycle is the diameter of its vertex set,* $\mathrm{diam}(z) = \mathrm{diam}(\mathrm{vert}(z))$*, in which the diameter of a set of vertices is the maximal discrete geodesic distance between them, formally,*

$$\mathrm{diam}(S) = \max_{p,q \in S} \mathrm{dist}(p, q).$$

Intuitively, a representative cycle of h with the minimal diameter, denoted z_d, is the cycle whose vertices are as close to each other as possible. The intuition will be further illustrated in Section 6 by comparison against other criteria. We prove in Theorem 5.0.5 that computing z_d of h is NP-hard, by reduction from a special case of the NP-hard *Multiple-Choice Cover Problem* (MCCP) of Arkin and Hassin [2]. The theorem has been stated in our previous paper [12]. But the proof has not been published.

Remark 5.0.3. *We do not address the approximability of* LHD*, as we realize that* z_d *suffers from a "wiggling problem" and consequently may be geometrically complex (see Section 6). However, it is not hard to see that the reduction in Theorem 5.0.5 is strict, which implies that* LHD *cannot be approximated any better than this special case of MCCP, which cannot be approximated within* $2 - \epsilon$ *for any* $\epsilon > 0$*, though we do not establish this formally.*

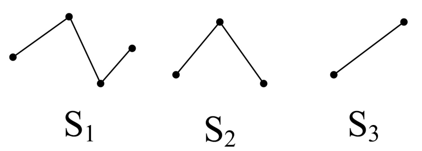
Problem 5.0.4 (Multiple-Choice Cover Problem)**.**
INPUT: a set of vertices, $V = \{v_1, v_2, ..., v_n\}$*; a distance function* $\mathrm{dist} : V \times V \to \mathbb{R}^+$ *satisfying triangular inequality; Disjoint subsets of* V*,* $S_1, S_2, ..., S_m$*, such that* $\bigcup_{i=1}^{m} S_i = V$
OUTPUT: a cover $C \subseteq V$ *containing one and exactly one vertex from each subset* S_i
MINIMIZE: $\mathrm{diam}(C)$

Please note that the original MCCP problem of Arkin and Hassin only requires the cover to have nonempty intersection with each subset S_i. We revise the problem to facilitate our proof, without influencing the NP-hardness. The reason is the optimal result of the revised problem is clearly an optimal result of the original problem.

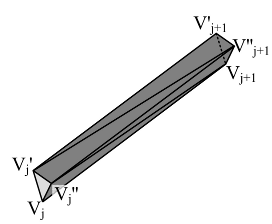
Theorem 5.0.5. LHD *is NP-hard to compute.*

Proof. We present a polynomial-time algorithm transforming an input of MCCP into an input of LHD. Later we will show that the solution of LHD gives us the solution of MCCP. As part of the input of LHD, the constructed simplicial complex K consisting of m tubes, T_1,...,T_m, as well as extra edges connecting vertices.

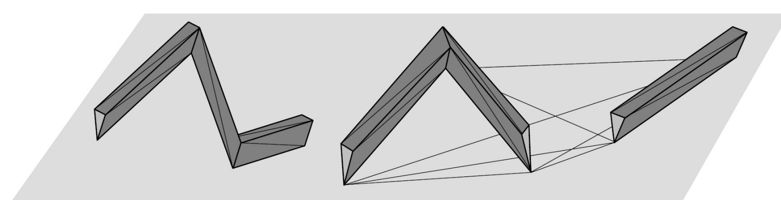
We first embed the vertex set V in any metric space preserving the pairwise distance dist. Without loss of generality, we assume V is embedded the Euclidean plane $\mathbb{R}^2$, for ease of explanation. For each vertex subset S_i, we find a simple path in $\mathbb{R}^2$, going though each vertex of S_i once without self-intersection, $\xi_i = (v_1, v_2, ..., v_{\text{card}(S_i)})$, which contains $\text{card}(S_i) - 1$ edges. The edge lengths are the same as the distances between corresponding vertices. See Figure 4(a). We construct a a slender threadlike tube T_i, which carries the



(a) An input of MCCP: 3 disjoint vertex subsets in Euclidean plane, S_1, S_2 and S_3. The simple paths, ξ_1, ξ_2 and ξ_3, are also shown, although they are not part of the input.



(b) A triangular cylinder corresponding to the edge $v_j v_{j+1}$.



(c) The constructed simplicial complex K. For ease of presentation, we only show the new edges connecting vertices in $S_2 \cup S_3$; however, note that there are other edges in K which are omitted in this figure.

Figure 4: Explanation of theorem 5.0.5 proof.

path ξ_i. T_i has $(3\,\mathrm{card}(S_i))$ vertices, $S_i \cup S_i' \cup S_i''$, where

$$S_i' = \{v_1', v_2', ..., v_{\mathrm{card}(S_i)}'\}, \quad \text{and} \quad S_i'' = \{v_1'', v_2'', ..., v_{\mathrm{card}(S_i)}''\}.$$

For any j, v_j' and v_j'' live very close to v_j. Corresponding to the $\mathrm{card}(S_i) - 1$ edges in ξ_i, T_i consists of $\mathrm{card}(S_i) - 1$ triangular cylinder concatenated together. By a *triangular cylinder* we mean the surface of a 3-prism with the two end triangles missing. To facilitate the concatenation, corresponding edges of the end triangles may not be parallel to each other, as in a standard 3-prism. Each edge $v_j v_{j+1}$ corresponds to a triangular cylinder with vertices v_j, v_j', v_j'', v_{j+1}, v_{j+1}' and v_{j+1}''. In the triangular cylinder, the short edges are very short, say, no longer than ϵ. The long edges have the length similar to the length of edge $v_j v_{j+1}$. See Figure 4(b) for one such triangular cylinder.

We construct the simplicial complex, K, as follows. For any i, $T_i \subseteq K$, and for any two vertices $v_1, v_2 \in V$, if they are not neighbors, add an edge connecting them, whose length is their Euclidean distance in the Euclidean plane $\mathbb{R}^2$. See Figure 4(c) for the complex constructed from the input in Figure 4(a). Please note that although in the figure, the embedding of K in $\mathbb{R}^3$ has self-intersection, the simplicial complex K can be embedded in Euclidean space of higher dimension, as we did in previous proofs.

For the constructed complex K, we use LHD to localize the 1-dimensional class $\sum_{i=1}^m h_i$, where h_i is the only 1-dimensional class carried by the tube T_i. We need a cycle to represent it as the input for LHD. We use $z_0 = \sum_{i=1}^m z_{i_0}$, where z_{i_0} is the 1-cycle whose vertices are v_{i_0}, v_{i_0}' and v_{i_0}'', in which v_{i_0} is an arbitrary vertex in S_i.

Next, we construct a cover C from the solution of LHD, z, and show that C is the solution of MCCP. We construct an intermediate vertex set $C_0 \subseteq V$ as follows. A vertex v belongs to C_0 if and only if any of v_i, v_i' and v_i'' belongs to the vertex set of z, $\mathrm{vert}(z)$. The solution z is in the form $\sum_{i=1}^m z_i$, where z_i represents class h_i. Therefore, C_0 has nonempty intersection with each vertex set S_i. We compute the cover C by picking one vertex from each $S_i \cap C_0$.

Within the simplicial complex,

$$\mathrm{diam}(C) = \mathrm{diam}(C_0) \quad \text{and} \quad |\,\mathrm{diam}(C_0) - \mathrm{diam}(z)| \leq 2\epsilon.$$

Furthermore, C has the same diameter in the simplicial complex, K, and in the Euclidean plane, $\mathbb{R}^2$. Since ϵ is arbitrarily small, we can see that C is the cover with the minimal diameter in the Euclidean plane, and thus, is the solution of MCCP. □

6 Radius

The third option for the objective function is radius.

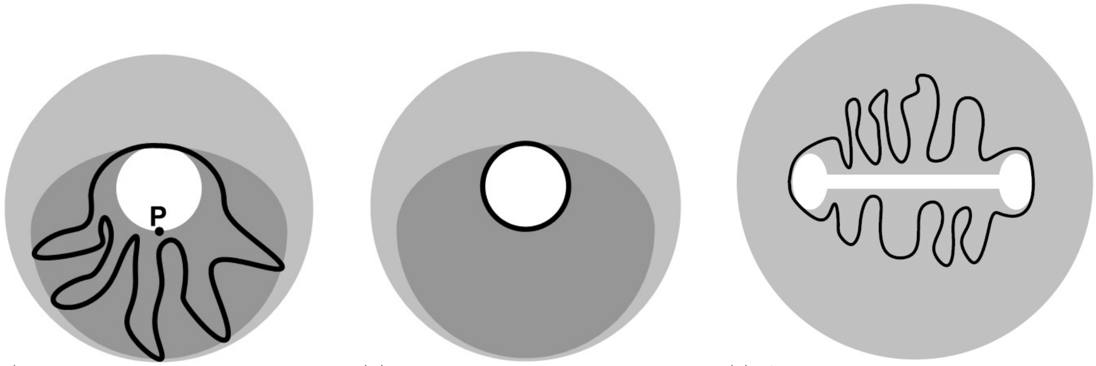
Definition 6.0.6 (Radius)**.** *The radius of a cycle is the radius of the smallest geodesic ball carrying it, formally,*

$$\mathrm{rad}(z) = \min_{p \in \mathrm{vert}(K), z \subseteq B_p^r} r$$

Given a homology class, the representative cycle with the minimal radius, denoted as z_r, is the same as the localized cycle defined in our previous work [12, 13]. Intuitively, z_r is the cycle whose vertices are as close to a vertex of K as possible. Theorem 6.0.8 shows that z_r can be computed in polynomial time.

However, in spite of its ease of computation, z_r may not necessarily be concise in an intuitive sense. It wiggles freely inside the smallest geodesic ball carrying it. See Figure 5(a) for example, in which we localize the only nontrivial homology class of an annulus (the light gray area). The dark gray area is the smallest geodesic ball carrying the class, whose center is p. Note that the geodesic ball of the annulus may not seem like a disc in the embedded Euclidean plane.

By contrast, the cycle with the minimal diameter, z_d, avoids this wiggling problem in this case and is concise in an intuitive sense (Figure 5(b)). This figure also illustrates that the radius and the diameter of a cycle are not strictly related. For the cycle z_r in Figure 5(a), its diameter is twice of its radius. For the cycle z_d in Figure 5(b), its diameter is equal to its radius.



(a) The cycle with the minimal radius, z_r.

(b) The cycle with the minimal diameter, z_d.

(c) A cross-section of a 3-dimensional ball with a bone shape void. z_d, which is 2-dimensional, wiggles near the middle of the bone.

Figure 5: Cycles with the minimal radius and diameter.

We prove that z_r is a 2-approximation of z_d.

Theorem 6.0.7. $\mathrm{diam}(z_r) \leq 2\,\mathrm{diam}(z_d)$.

Proof. First, the triangle inequality of the geodesic distance suggests that for any two vertices of z_r, p_1 and p_2, their geodesic distance is

$$\mathrm{dist}(p_1, p_2) \leq \mathrm{dist}(p_1, p_0) + \mathrm{dist}(p_0, p_2) \leq 2\,\mathrm{rad}(z_r),$$

where p_0 is the center of the smallest geodesic ball carrying the cycle z_r and the class. This implies that the diameter of z_r is no greater than twice of its radius.

Second, the diameter of z_d is no less than its radius. To see this, pick a geodesic ball centered at any vertex of z_d with radius $\mathrm{diam}(z_d)$. This ball carries z_d. Finally,

$$\mathrm{diam}(z_r) \leq 2\,\mathrm{rad}(z_r) \leq 2\,\mathrm{rad}(z_d) \leq 2\,\mathrm{diam}(z_d).$$

□

As shown in Figure 5(a) and 5(b), this bound is a tight bound.

However, in general, the minimal diameter cycle also suffers from the wiggling problem. In Figure 5(c), we show an example in which the topological space is a closed 3-dimensional ball with a bone shape void in the middle. The minimal diameter 2-cycle, z_d, representing the only nontrivial 2-dimensional class, can freely wiggle near the middle of the bone, as the diameter is determined by the distance of the two ends of the bone. The reason for this phenomenon is in finding the minimal diameter cycle, we minimize the maximum of all pairwise geodesic distances. It is not hard to see that z_d does not wiggle only if for any $v \in \text{vert}(z_d)$, its longest distance from other vertices in z_d is close to $\text{diam}(z_d)$.

We conclude this section by showing that z_r can be computed in polynomial time. The proof is a short sketch of a polynomial algorithm of Chen and Freedman [13].

Theorem 6.0.8. *We can compute z_r in polynomial time.*

Proof. For each vertex, p, we find the smallest geodesic ball centered at p carrying any cycle in $[z_0]$, namely, $B_p^{r(p)}$, as well as the carried cycle. Iterating through all vertices $p \in \text{vert}(K)$, the one with the smallest $r(p)$ gives us z_r.

To compute $B_p^{r(p)}$, we apply persistent homology on the complex using f_p as the filter function. Persistent homology algorithm computes a homology cycle basis, $\{z_1, z_2, \cdots, z_{\beta_d}\}$, sorted according to the time they enter the sublevel set. We find the smallest index i so that z_0 is a linear combination of boundaries and z_1, z_2, $\cdots$, z_i, namely,

$$z_0 = [\partial_{d+1}, z_1, z_2, \cdots, z_i]\gamma, \tag{2}$$

where d is the dimension of the class in question. The time z_i enters the sublevel set is the radius $r(p)$. Replacing ∂_{d+1} with 0, we get a representative cycle of $[z_0]$ carried by $B_p^{r(p)}$, $[0, z_1, z_2, \cdots, z_i]\gamma$. We can use a sparse matrix rank computation over finite field to improve the time complexity, when the dimension of the homology and the Betti number are small enough ($O(\log n)$) [33]. $\square$

7 Conclusion

In this paper, we have addressed the localization problem with regard to three different measures. For volume, we have proven inapproximability results for the cases when Betti number is $\Theta(n)$ and is fixed. We have also proven the inapproximability of computing the nonbounding cycle with the minimal volume and computing the homology cycle basis with the minimal total volume. A special case in which polynomial algorithm exists has also been discussed.

For diameter, we have proven that the localization problem is NP-hard; for radius, by contrast, we have stated a polynomial time algorithm. Both of these two measures, however, suffer from the "wiggling problem", namely, that the output of the localization may be geometrically quite complex.

An open question is whether we can use other discrete geodesic distance related measures, besides diameter and radius, which do not suffer from the wiggling problem. For example, can we use the normalized sum of the pairwise geodesic distances? Furthermore, what if we restrict the geodesic distance to be within the cycle (rather than the entire complex)? It is conceivable that these distance related measures might be easier to compute, as localization with volume measure has been shown to be extremely hard.

Acknowledgment

The authors thank David Cohen-Steiner and Omid Amini for constructive discussion.

References

[1] P. K. Agarwal, H. Edelsbrunner, J. Harer, and Y. Wang. Extreme elevation on a 2-manifold. *Discrete & Computational Geometry*, 36:553–572, 2006.

[2] E. M. Arkin and R. Hassin. Minimum-diameter covering problems. *Networks*, 36(3):147–155, 2000.

[3] S. Arora, L. Babai, J. Stern, and Z. Sweedyk. The hardness of approximate optima in lattices, codes, and systems of linear equations. *J. Comput. Syst. Sci.*, 54(2):317–331, 1997.

[4] G. Ausiello and V. T. Paschos. Reductions, completeness and the hardness of approximability. *European Journal of Operational Research*, 172(3):719–739, 2006.

[5] E. Carlsson, G. Carlsson, and V. de Silva. An algebraic topological method for feature identification. *Int. J. Comput. Geometry Appl.*, 16(4):291–314, 2006.

[6] G. Carlsson. Persistent homology and the analysis of high dimensional data. Fields-Ottawa Workshop on the Geometry of Very Large Data Sets, February 2005.

[7] C. Carner, M. Jin, X. Gu, and H. Qin. Topology-driven surface mappings with robust feature alignment. In *Proceedings of the 16th IEEE Visualization Conference*, pages 543–550, 2005.

[8] E. W. Chambers, É. Colin de Verdière, J. Erickson, F. Lazarus, and K. Whittlesey. Splitting (complicated) surfaces is hard. In *Proceedings of the 22nd ACM Symposium on Computational Geometry*, pages 421–429, 2006.

[9] E. W. Chambers, J. Erickson, and A. Nayyeri. Homology flows, cohomology cuts. In *STOC*, 2009.

[10] E. W. Chambers, J. Erickson, and A. Nayyeri. Minimum cuts and shortest homologous cycles. In *Symposium on Computational Geometry*, 2009.

[11] C. Chen and D. Freedman. Quantifying homology classes ii: Localization and stability. *CoRR*, abs/0709.2512, 2007.

[12] C. Chen and D. Freedman. Quantifying homology classes. In *Proceedings of the 25th Annual Symposium on Theoretical Aspects of Computer Science*, pages 169–180, 2008.

[13] C. Chen and D. Freedman. Measuring and computing natural generators for homology groups. *Computational Geometry: Theory and Applications*, to appear.

[14] D. Cohen-Steiner, H. Edelsbrunner, and D. Morozov. Vines and vineyards by updating persistence in linear time. In *Proceedings of the 22nd ACM Symposium on Computational Geometry*, pages 119–126, 2006.

[15] É. Colin de Verdière and J. Erickson. Tightening non-simple paths and cycles on surfaces. In *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 192–201, 2006.

[16] V. de Silva and R. Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7:339–358, 2007.

[17] T. K. Dey, K. Li, and J. Sun. On computing handle and tunnel loops. In *IEEE Proc. NASAGEM*, 2007.

[18] T. K. Dey, K. Li, J. Sun, and D. Cohen-Steiner. Computing geometry-aware handle and tunnel loops in 3d models. *ACM Trans. Graph.*, 27(3), 2008.

[19] I. Dumer, D. Micciancio, and M. Sudan. Hardness of approximating the minimum distance of a linear code. *IEEE Transactions on Information Theory*, 49(1):22–37, 2003.

[20] H. Edelsbrunner and J. Harer. Persistent homology: A survey. *Twenty Years After*, to appear.

[21] J. Erickson and S. Har-Peled. Optimally cutting a surface into a disk. *Discrete & Computational Geometry*, 31(1):37–59, 2004.

[22] J. Erickson and K. Whittlesey. Greedy optimal homotopy and homology generators. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1038–1046, 2005.

[23] Q. Fang, J. Gao, and L. Guibas. Locating and bypassing routing holes in sensor networks. In *Mobile Networks and Applications*, volume 11, pages 187–200, 2006.

[24] R. Ghrist. Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc.*, 45(1):61–75, 2008.

[25] I. Guskov and Z. J. Wood. Topological noise removal. In *Proceedings of the Graphics Interface 2004 Conference*, pages 19–26, 2001.

[26] T. Kaczynski, M. Mrozek, and M. Slusarek. Homology computation by reduction of chain complexes. *Computers and Math. Appl.*, 35:59–70, 1998.

[27] D. Kirsanov and S. J. Gortler. A discrete global minimization algorithm for continuous variational problems. Technical Report TR-14-04, Harvard University, 2004.

[28] M. Kutz. Computing shortest non-trivial cycles on orientable surfaces of bounded genus in almost linear time. In *Proceedings of the 22nd ACM Symposium on Computational Geometry*, pages 430–438, 2006.

[29] D. J. C. MacKay. *Information Theory, Inference & Learning Algorithms.* Cambridge University Press, 2002.

[30] J. R. Munkres. *Elements of Algebraic Topology.* Addison-Wesley, Redwook City, California, 1984.

[31] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete and Computational Geometry*, 39(1):419–441, 2008.

[32] R. Sarkar, X. Yin, J. Gao, F. Luo, and X. D. Gu. Greedy routing with guaranteed delivery using ricci flows. In *Proc. of the 8th International Symposium on Information Processing in Sensor Networks (IPSN'09)*, pages 121–132, April 2009.

[33] D. H. Wiedemann. Solving sparse linear equations over finite fields. *IEEE Transactions on Information Theory*, 32(1):54–62, 1986.

[34] Wikipedia. Book embedding. http://en.wikipedia.org/wiki/Book_embedding.

[35] Wikipedia. Suspension. http://en.wikipedia.org/wiki/Suspension_(topology).

[36] Z. J. Wood, H. Hoppe, M. Desbrun, and P. Schröder. Removing excess topology from isosurfaces. *ACM Trans. Graph.*, 23(2):190–208, 2004.

[37] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.

[38] A. Zomorodian and G. Carlsson. Localized homology. In *Proceedings of the 2007 International Conference on Shape Modeling and Applications*, pages 189–198, 2007.