

Hardness Results for Homology Localization

Chao Chen *

Institute of Science and Technology Austria
Vienna University of Technology, Austria

Email: ChaoChen79@gmail.com

Daniel Freedman

Hewlett-Packard Laboratories, Israel

Email: daniel.freedman@hp.com

October 29, 2010

Abstract

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, two main results are presented. First, we prove the problem is NP-hard to approximate within any constant factor. Second, we prove that for homology of dimension two or higher, the problem is NP-hard to approximate even when the Betti number is $O(1)$. The latter result leads to the inapproximability 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, diameter and radius, we show that the localization problem is NP-hard for diameter, but is polynomial for radius.

Our work is restricted to homology over the \mathbb{Z}_2 field. Results over other fields have been studied recently by Dey et al. [13].

*Partially supported by the Austrian Science Fund under grant P20134-N13

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 [5], graphics [17], networks [12] and computational biology [11]. 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 [21, 29, 25]. 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 [20, 26]. 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 [4].

In this paper, we will address the *localization problem*, namely, 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. See Figure 1 for examples. In a disk with three holes (Figure 1(a)), cycles z_1 and z_2 are well-localized; z_3 is not. In a 2-handled torus (Figure 1(b)), the concise cycle z_1 is a better representative (than z_2) of its class, and describes the small handle better.

In Section 4.1, we use *volume*, the number of simplices of a cycle, as the criterion to minimize. For a 1-dimensional (resp. 2-dimensional) cycle, the volume is its length (resp. area).

We have two main results. First, we prove that localizing a given class with the min-

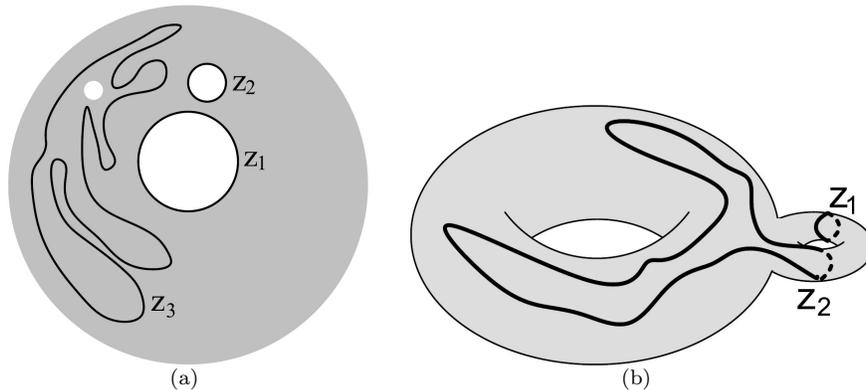


Figure 1: Motivating examples for localization.

imal volume cycle is NP-hard to approximate within any constant factor. The proof is a strict reduction from the nearest codeword problem [2]. We prove the inapproximability for homology of any dimension.

Second, we prove that for homology of dimension two or higher, computing the non-bounding cycle with the smallest volume is NP-hard to approximate within any constant factor. This is true even when the Betti number is fixed. This result leads to the inapproximability of two other problems concerning homology of dimension two or higher, namely,

- localizing a given class with the minimal volume cycle, when Betti number is fixed, and
- computing a homology cycle basis with the minimal total volume.

We conclude the paper with a discussion of other minimization criteria, including *diameter* and *radius* (Section 5). In specific, we show that localization with diameter cost function is NP-hard, where as with radius is polynomial.

Throughout this paper, the topological features we use are homology classes over \mathbb{Z}_2 field. (Thus, all the additions are mod 2 additions.)

2 Preliminaries

2.1 Homology Groups.

We briefly describe some background knowledge from algebraic topology. Refer to [24] for more details. We restrict our discussion to the combinatorial framework of simplicial homology over \mathbb{Z}_2 field.

Given a simplicial complex K , a d -chain is a formal sum of d -simplices, $c = \sum_{\sigma \in K} a_\sigma \sigma$, $a_\sigma \in \mathbb{Z}_2$. All the d -chains form the *group of d -chains*, $C_d(K)$. 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 $\partial_d : C_d(K) \rightarrow C_{d-1}(K)$ is a group homomorphism.

A d -cycle is a d -chain without boundary.¹ The set of d -cycles forms a subgroup of the chain group, which is the kernel of the boundary operator, $Z_d(K) = \ker(\partial_d)$. 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, $B_d(K) = \text{img}(\partial_{d+1})$. It is not hard to see that a d -boundary is also a d -cycle. Therefore, $B_d(K)$ is a subgroup of $Z_d(K)$. A d -cycle which is not a d -boundary, $z \in Z_d(K) \setminus B_d(K)$, 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 $H_d(K) = Z_d(K)/B_d(K)$. An element in $H_d(K)$ is a *homology class*, which is a coset of $B_d(K)$, $[z] = z + B_d(K)$ for some d -cycle $z \in Z_d(K)$. If z is a d -boundary, $[z] = B_d(K)$ is the identity element of $H_d(K)$. 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.

¹For those unfamiliar with homology, we emphasize that a 1-cycle is different from the cycle defined in graph theory. For the former definition, a 1-cycle can be a disjoint union of arbitrarily many 1-cycles. But this is not true for the latter definition.

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

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

As the dimension of the chain group is upperbounded by the cardinality of K , n , so are the dimensions of $\mathbf{B}_d(K)$, $\mathbf{Z}_d(K)$ and $\mathbf{H}_d(K)$. The Betti number can be computed with a reduction algorithm based on row and column operations of the boundary matrices [24]. Various reduction algorithms have been devised for different purposes.

A *homology basis* is a set of β_d classes generating the group $\mathbf{H}_d(K)$. 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. Given a homology cycle basis, $\{z_1, z_2, \dots, z_{\beta_d}\}$, any homology class can be written as $\sum_{i=1}^{\beta_d} a_i [z_i]$. Its cycles can be written as $[z_1, \dots, z_{\beta_d}] \alpha + \partial_{d+1} \gamma$, where $\alpha = \begin{bmatrix} a_1 \\ \vdots \\ a_{\beta_d} \end{bmatrix}$.

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 the set 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 $[b_1, \dots, b_{n_d}]$, 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. We denote the *d-skeleton* of K as the subcomplex consisting of all the d -simplices and their faces. The following notation will prove convenient. We say that a d -chain $c \in \mathbf{C}_d(K)$ is *carried by* a subcomplex K_0 when all the d -simplices of c belong to K_0 . We denote $\text{vert}(K)$ as the set of vertices of the simplicial complex K , $\text{vert}(c)$ 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 $p \in \text{vert}(K)$, $f_p : \text{vert}(K) \rightarrow \mathbb{R}$, as follows. Suppose each edge in K has a nonnegative weight, for any vertex $q \in \text{vert}(K)$, $f_p(q) = \text{dist}(p, q)$ 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 , $f_p(\sigma) = \max_{q \in \text{vert}(\sigma)} f_p(q)$. 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 , $p \in \text{vert}(K)$, $r \geq 0$, 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.

2.2 Terminology from Coding Theory.

We focus on binary linear codes and thus only use matrices over the \mathbb{Z}_2 field. For consistency, we switch the roles of the row and column indices from the standard definition. Refer to [23] 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, $\text{span}(A)$. 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 generated by A .

Given a generator matrix A , A^\perp may be computed in polynomial time. Its dimension is $(m - k) \times m$. In specific, $A^\perp A = 0$, and thus $A^T (A^\perp)^T = 0$. Any basis of the solution space of the linear equation system $A^T x = 0$ would form a valid A^\perp . This could be computed by a Gauss-Jordan elimination. For example, the generator matrix and its parity-check matrix

could be

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad A^\perp = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

2.3 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. See [3] for more details. For ease of exposition, we only discuss minimization problems. The definitions can be extended to maximization problems easily.

An *NP optimization problem* Π is a three-tuple $(\mathcal{I}, \text{Sol}, m)$ in which \mathcal{I} is the set of *instances*. For each instance $I \in \mathcal{I}$, $\text{Sol}(I)$ denotes the set of *feasible solutions* of I , and the *cost function*, $m(I, S)$, produces a value for each feasible solution $S \in \text{Sol}(I)$. Any instance can be recognized in time polynomial in its size, $\text{card}(I)$. It is also polynomial to verify whether any given S is a feasible solution, or evaluate the cost function m .

For an instance I and one of its feasible solutions, $S \in \text{Sol}(I)$, 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 optimal solution, formally,

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

where $S^*(I)$ is the optimal 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 $[1, \infty)$.

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 , $\rho_A(I) \leq r$. 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 , such that f transforms any instance I_1 in Π_1 into an instance $I_2 = f(I_1)$ in Π_2 , and g transforms any feasible solution of this I_2 , S_2 , into a feasible solution of the initial instance I_1 , $g(S_2)$.

We say the reduction is *strict* ($\Pi_1 \leq_S \Pi_2$) if in addition, for any instance $I_1 \in \mathcal{I}_{\Pi_1}$ and any feasible solution of $f(I_1)$, S_2 , the performance ratios satisfy

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

Given such a strict reduction, the optimal solution of $f(I_1)$ would lead to an optimal solution of I_1 , and furthermore, any feasible solution of $f(I_1)$ would lead to a feasible solution of I_1 with better performance ratio. It is straightforward to see that an r -approximation algorithm of Π_2 would lead to an r -approximation algorithm of Π_1 . Therefore, strict reduction preserves the membership of APX. The following lemma will be useful for our inapproximability proof.

Lemma 2.3.1. *If $\Pi_1 \leq_S \Pi_2$ and $\Pi_1 \notin \text{APX}$, then $\Pi_2 \notin \text{APX}$.*

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 cycle. Using Dijkstra's shortest path algorithm, Erickson and Whittlesey [19] computed the *shortest homology basis* of 2-manifolds, namely, the 1-dimensional homology cycle basis whose elements have the minimal total volume. Dey et al. [16] provided a polynomial time algorithm to compute such basis when the input is a general simplicial complex.

These polynomial time algorithms cannot localize an arbitrarily given class. To fill this void, Chambers et al. [7] (see [18] for a recent improvement) devised an algorithm to

localize a given class, when the input is a 2-manifold, with or without boundary. Their method precomputes the shortest representative cycles of all $2^{\beta_1} - 1$ nontrivial classes, and thus, is exponential in the 1-dimensional Betti number, β_1 .

It has been demonstrated that when $\beta_1 = \Theta(n)$, localizing a given 1-dimensional class with its shortest cycle is NP-hard, both in the cases that the topological space is a general complex [8] and a 2-manifold [7].

Due to the difficulties in localizing with the minimal volume criterion, researchers have focused on other criteria or heuristics. Some have 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 [21, 29] 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. [14] computed these handle-related cycles by computing the deformation retractions of the two components of the embedding space bounded by the given 2-manifold. A recent extension [15] improved their result based on geometric heuristics and persistent homology. Their work facilitates handle detection in real applications.

All of the aforementioned works are restricted to 1-dimensional homology. Zomorodian and Carlsson [30] 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 [10] presented a polynomial time algorithm for localizing a homology class of general dimension with the minimal radius cycle. Their algorithm can also compute a homology cycle basis with the minimal total radius. The cycle with the minimal radius, however, may be quite complicated in terms of geometry. See Section 5 for detailed discussion.

In terms of homology over other fields, Chambers et al. [6] addressed the localization problem of 1-dimensional homology over other fields by formulating a maximization problem.

They view a 1-chain as a flow of the 1-skeleton of a simplicial complex. The localization problem is formalized as finding a maximal flow homologous to a given flow under a given constraint of the edge capacities. Two 1-chains are homologous if their difference is a 1-boundary. Their algorithm is exponential in β_1 for real coefficients and $O(\beta_1^7 n \log^2 n \log^2 C)$ for integer coefficients, where C is the total sum of all the edge capacities.

Dey et al. [13] showed that the localization problem with minimal volume is polynomial time solvable for a class of spaces when the homology is defined over the integers, \mathbb{Z} . In particular, they showed that when the input space is an orientable $(d + 1)$ -dimensional manifold and the homology class in question is d -dimensional, the localization problem is polynomial over \mathbb{Z} , despite its being NP-hard over the \mathbb{Z}_2 field [7].

4 Localization with Volume

4.1 Problem Formalization and a List of Existing Results

Given an cost function defined on all the d -cycles, $\text{cost} : Z_d(K) \rightarrow \mathbb{R}$, we formalize the localization problem as a combinatorial optimization problem.

Problem 4.1.1 (Localizing Homology).

INPUT: a simplicial complex K with size n , a d -dimensional nontrivial homology class

$h = [z_0]$, $0 \leq d \leq \dim(K)$

OUTPUT: a cycle $z \in h$

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

In this section, we use volume as the cost function. Other cost functions would be discussed in Section 5.

Definition 4.1.2 (Volume). *The volume of a cycle is the number of its simplices,*

$$\text{vol}(z) = \text{card}(z)$$

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. For convenience, we denote LocHomVol as the problem of localizing a homology class with its minimal volume cycle, z_v .

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

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

Computing z_v using this general volume definition is at least as hard as using Definition 4.1.2, which is in fact a special case (when $w(\sigma) = 1, \forall \sigma \in K$). Therefore, we will only treat the unweighted volume function.

There are two other variations, which are supposed to be easier than LocHomVol, namely, computing a nonbounding cycle with the minimal volume, and computing a homology cycle basis with the minimal total volume, formally,

Problem 4.1.3 (Min-Vol Nonbounding Cycle).

INPUT: a simplicial complex K with size n

OUTPUT: a nonbounding d -cycle z , $0 \leq d \leq \dim(K)$

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

Problem 4.1.4 (Min-Vol Basis).

INPUT: a simplicial complex K with size n

OUTPUT: a homology cycle basis $\{z_1, z_2, \dots, z_{\beta_d}\}$, $0 \leq d \leq \dim(K)$

MINIMIZE: $\sum_{i=1}^{\beta_d} \text{vol}(z_i)$

We name these two problems MinVolNBCyc and MinVolBasis, respectively. For clarity, we show in Figure 2 the difference between MinVolBasis and LocHomVol. The former is easier than the latter.

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

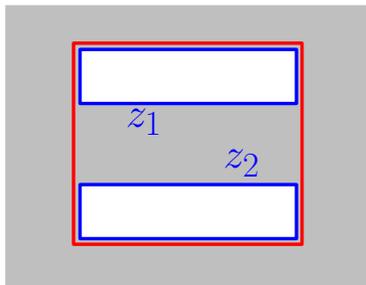


Figure 2: A rectangle with two holes. The output of MinVolBasis is the two blue 1-cycles, z_1 and z_2 . However, the minimal volume cycle in $[z_1 + z_2]$ is the red cycle, which cannot be computed using the optimal homology cycle basis.

LocHomVol , $\beta_1 = \Theta(n)$	NP-hard
LocHomVol , $\beta_1 = O(1)$, 2-manifolds	polynomial
LocHomVol , $\beta_1 = O(1)$, general complexes	unknown
MinVolNBCyc	polynomial
MinVolBasis	polynomial

Table 1: Existing results for 1-dimensional homology.

- When $\beta_1 = \Theta(n)$, LocHomVol is proved to be NP-hard by polynomial reductions from a special case of MAX-2SAT [8] and MIN-CUT with negative edge weights [7].
- Chambers et al. [7] provided a polynomial time algorithm for LocHomVol when β_1 is fixed (the algorithm is improved in [18]). The algorithm computes the shortest representative cycle for each of the $2^{\beta_1} - 1$ nontrivial classes. This work is restricted to triangulations of 2-manifolds with or without boundaries. The problem remains open when the input is a general simplicial complex.
- Erickson and Whittlesey [19] devised a polynomial time algorithm for MinVolBasis, even when $\beta_1 = \Theta(n)$. This work is restricted to triangulations of 2-manifolds. Dey et al. [16] provided a polynomial time algorithm when the input is a general simplicial complex.

We summarize these results in Table 1.

All these existing results are about 1-dimensional homology. In this paper, we will study whether LocHomVol is difficult in general dimension, and more importantly, how difficult

it is.

The existing results suggest that the localization problem might be easier if we assume fixed Betti number, or if we compute `MinVolNBCyc` or `MinVolBasis` instead. Therefore, we would also like to find out how difficult these problems could be. We prove the inapproximability of a special case of `MinVolNBCyc`, namely, when $\beta_d = 1$, 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 the sake of clarity, we list all the new results as follows.

- When the homology in question is 1-dimensional or higher and the Betti number is $\Theta(n)$, it is NP-hard to approximate `LocHomVol` within any constant factor (Theorem 4.2.4).
- When the homology in question is 2-dimensional or higher, we prove that `MinVolNBCyc` is NP-hard to approximate within any constant factor (Theorem 4.3.3). So do `LocHomVol` with $\beta_d = O(1)$ and `MinVolBasis` (Corollary 4.3.4).
- A polynomial time algorithm to compute the minimal volume nonbounding cycle for a special case: when the pertinent space is embedded in \mathbb{R}^N and the pertinent homology is $(N - 1)$ -dimensional.

4.2 LocHomVol is NP-hard to approximate within any constant factor

We prove by a strict reduction from the nearest codeword problem (`NearestCodeword`), which cannot be approximated within any constant factor [2]. Problems used in previous reductions to `LocHomVol` [8, 7] have constant approximation ratios, and thus cannot be used for our proof.

Problem 4.2.1 (Nearest Codeword Problem).

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

OUTPUT: a vector $y \in y_0 + \text{span}(A)$

MINIMIZE: the Hamming weight of y

Lemma 4.2.2. *For 1-dimensional homology, LocHomVol cannot be approximated within any constant factor.*

Proof. We prove by a strict reduction from NearestCodeword, namely,

$$\text{NearestCodeword} \leq_S \text{LocHomVol}.$$

Given an instance of NearestCodeword, 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. Note that we are abusing notation when we call T a cell complex, as these cells may not be homeomorphic to closed balls. See Figure 3 for an example with a 4×2 generator matrix

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

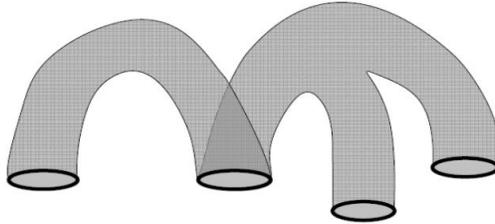


Figure 3: The constructed cell complex, T . Two 2-cells (pipes) share four 1-cells (thickened circles), corresponding to two columns and four rows of A .

The column space $\text{span}(A)$ one-to-one corresponds to the set of boundaries of T . The m -dimensional binary vector space, \mathbb{Z}_2^m , one-to-one corresponds to the set of 1-cycles of T . Therefore, NearestCodeword is identical to the problem of computing the minimal volume representative cycle of a given 1-dimensional class of T , $[y_0]$. However, this problem, denoted as LocHomVol-T, is different from LocHomVol, whose input is a simplicial complex. We

will prove in Lemma 4.2.3 that we can strictly reduce LocHomVol-T to LocHomVol by subdividing T into a simplicial complex K . This concludes the proof. \square

For convenience, we introduce some notations before proving Lemma 4.2.3. We call a 1-chain c a *simple path* if $\text{card}(c) = \text{card}(\text{vert}(c)) + 1$, and there is a non-repeating sequence of $\text{vert}(c)$, (v_1, v_2, \dots, v_k) , such that any two consecutive vertices in the sequence is connected by an edge of c ². The first and last vertices are the *end vertices*. If we identify the two end vertices, that is, $v_1 = v_k$, the chain c is called a *simple cycle*. In this case, $\text{card}(c) = \text{card}(\text{vert}(c))$. We extend the definition of homologous to chains. Two chains are *homologous* to each other if their difference is a boundary.

Lemma 4.2.3.

$$\text{LocHomVol-T} \leq_S \text{LocHomVol}$$

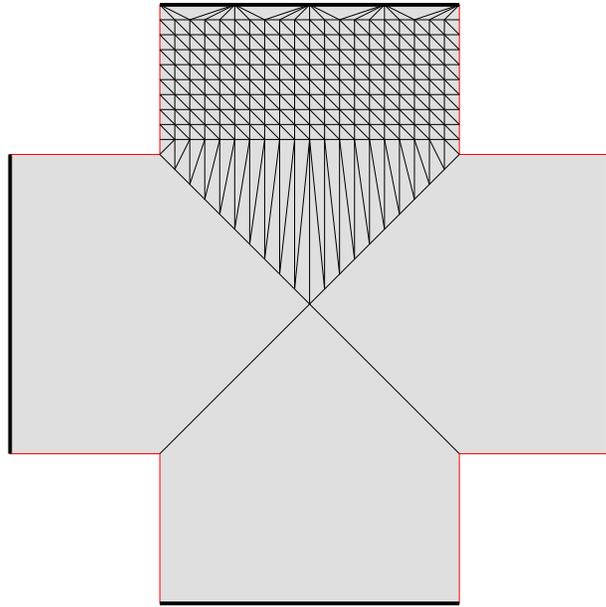
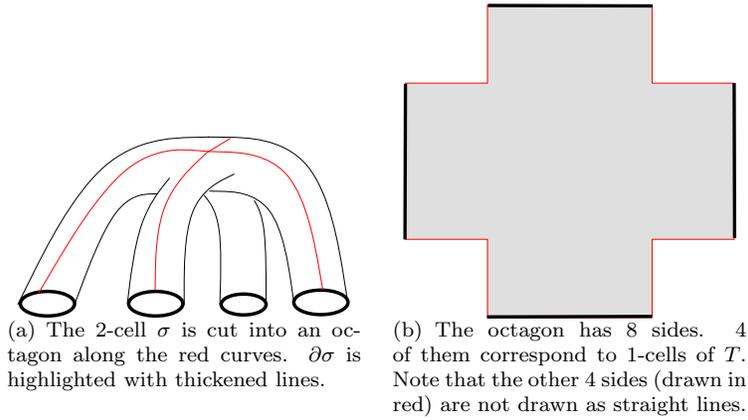
Proof. We prove by subdividing T into a simplicial complex K , whose size is polynomial to the size of T . Recall m is the number of 1-cells of T . For each 2-cell of T , σ , we triangulate it as follows.

1. Cut σ into a polygon with $2a$ sides, where a is the cardinality of $\partial\sigma$. For example, in Figure 4, σ is cut into an octagon.
2. Triangulate each of the a sides corresponding to 1-cells of T into t_1 edges.
3. Triangulate the interior of the polygon fine enough so that for any two vertices from two non-adjacent sides of the polygon, any path connecting them has at least $t_1 m$ edges. See Figure 4 for the case when $\text{card}(\partial\sigma) = 4$, $t_1 = 4$ and $m = 5$.

Note that the size of K is polynomial to the size of T , and thus can be produced in polynomial time.

For convenience, we denote the triangulation of all 1-cells of T as K_1 , which is a sub-complex of K . The number of edges in K_1 is mt_1 . 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 , y , and its corresponding 1-cycle of K_1 , $\phi(y)$, the ratio of their volumes is $1 : t_1$.

²This definition is consistent with the definition in graph theory.



(c) A fine triangulation of the polygon. For simplicity, we only draw 1/4 of the triangulation. The size of this triangulation is polynomial to m and t_1 .

Figure 4: The triangulation of a 2-cell of T , σ , with $\text{card}(\partial\sigma) = 4$, $t_1 = 4$ and $m = 5$.

Our construction provides a polynomial transformation of every instance of LocHomVol-T , (T, y_0) , into an instance of LocHomVol , $(K, z_0 = \phi(y_0))$. For any such instance, and any feasible solution $z \in [z_0]$, we will show a polynomial time algorithm transforming z into a cycle $z' \in [z_0]$ carried by K_1 , such that $\text{vol}(z') \leq \text{vol}(z)$. This leads to a solution of LocHomVol-T , $\phi^{-1}(z')$. For convenience, we denote this solution $g(z)$.

This reduction is strict. First, the optimal solution of LocHomVol, z_v , is a cycle of K_1 , whose corresponding solution of LocHomVol-T, $g(z_v) = \phi^{-1}(z_v)$ is the optimal solution of LocHomVol-T. The ratio of their volumes is $\text{vol}(z_v) : \text{vol}(g(z_v)) = t_1 : 1$. Second, for any feasible solution z , the volume of its corresponding solution in LocHomVol-T is

$$\text{vol}(g(z)) = \text{vol}(\phi^{-1}(z')) = \frac{1}{t_1} \text{vol}(z') \leq \frac{1}{t_1} \text{vol}(z),$$

and therefore, $\frac{\text{vol}(g(z))}{\text{vol}(g(z_v))} \leq \frac{\text{vol}(z)}{\text{vol}(z_v)}$.

This guarantees Inequality (1), and thus the strictness of the reduction.

To conclude this proof, we show that for any given cycle of K , $z \in [z_0]$ with z_0 carried by K_1 , there is a polynomial time algorithm which computes a cycle $z' \in [z_0]$ carried by K_1 , such that $\text{vol}(z') \leq \text{vol}(z)$. We partition z into simple cycles and simple paths by finding all repeating vertices and vertices of K_1 . Each simple cycle has no vertex from K_1 . Each simple path has no vertices from K_1 except for the two end vertices. Next, we deal with these simple cycles and simple paths one by one. For each simple cycle or simple path, which is carried by the triangulation of one 2-cell of T , σ , we find a 1-chain homologous to it and is carried by K_1 . The new chain has a smaller or equal volume.

There are three cases. Recall that ϕ maps a 1-chain of T to its subdivision.

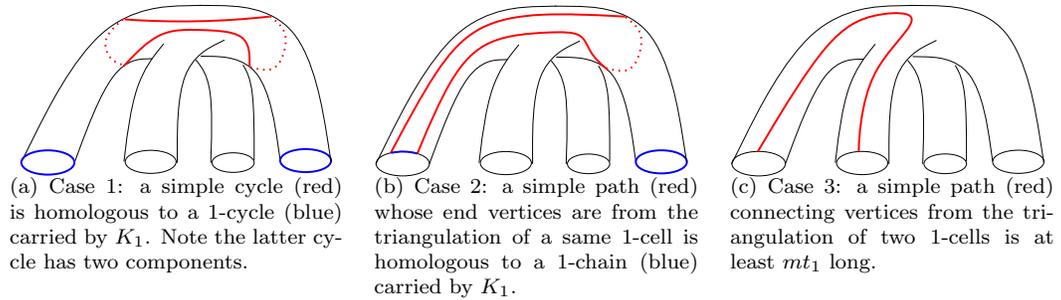


Figure 5: Different cases for generating z' .

1. If a simple cycle is trivial, replace it with an empty chain. Otherwise, this cycle is homologous to a cycle carried by the triangulation of $\partial\sigma$, $\phi(\partial\sigma) \subseteq K_1$. The latter cycle has a smaller or equal volume, due to the way we subdivide σ . See Figure 5(a)

for an example.

2. For a simple path whose both end vertices are from the triangulation of a same 1-cell of T , $\tau \in \partial\sigma$, it is homologous to a path connecting the two end vertices within $\phi(\tau)$ plus cycles which are triangulations of other cells of $\partial\sigma$. The latter chain has a smaller or equal volume, due to the way we subdivide σ . See Figure 5(b).
3. Suppose it is a simple path connecting vertices from the triangulations of two different 1-cells (Figure 5(c)). Such path has a volume of at least mt_1 . In such case, we let z' be the input z_0 , whose volume is no greater than mt_1 , and thus no greater than $\text{vol}(z)$.

z' is computed after we transform all simple paths and simple cycles into homologous chains and cycles carried by K_1 , or we let $z' = z_0$ if Case 3 happens. \square

Lemma 4.2.2 is about 1-dimensional homology. We extend the result to homology of any higher dimension.

Theorem 4.2.4. *For any $d \geq 1$, LocHomVol for d -dimensional homology cannot be approximated within any constant factor.*

Proof. We show that when $d \geq 2$, LocHomVol for $(d - 1)$ -dimensional homology can be strictly reduced to LocHomVol for d -dimensional homology, namely, $\text{LocHomVol}_{d-1} \leq_S \text{LocHomVol}_d$. Together with Lemma 4.2.2, the theorem is proved.

Next, we explain the reduction. Given a simplicial complex of LocHomVol_{d-1} , we build a *suspension* of it, namely, two cones of the complex glued together at their base [28]. There is a one-to-one correspondence between the $(d - 1)$ -dimensional cycle group of the original complex and the d -dimensional cycle group of the new complex. This correspondence also works for the boundary groups. Since the volume of each $(d - 1)$ -cycle is $1/2$ of the volume of its corresponding d -cycle, this is a strict reduction. \square

Restriction to a manifold. A natural question is whether the localization problem could be made easier if we restrict the input to be the triangulation of a manifold. We could

modify Lemma 4.2.2 and its proof to accommodate this manifold assumption. Specifically, we embed the cell complex T in \mathbb{R}^N . By thickening the underlying space of T and taking its boundary as a new topological space, we get an $(N - 1)$ -manifold (one less dimension than the ambient space). This manifold can be triangulated in a similar way as we triangulate T , and thus leads to the inapproximability of LocHomVol for 1-dimensional homology when the input is the triangulation of an $(N - 1)$ -manifold. We omit a formal description of such reduction. See Figure 4.2 for an example when T is 1-dimensional and $N = 2$.

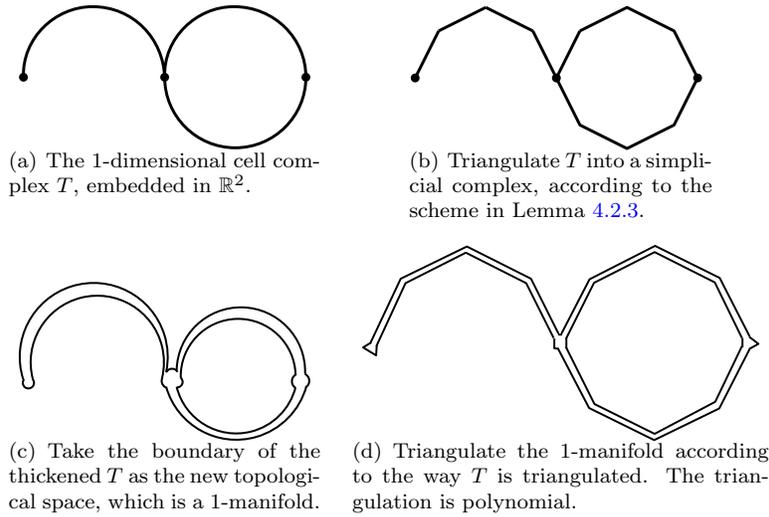


Figure 6: An example of modifying proof of Lemma 4.2.3 to accommodate the manifold assumption.

A classical result suggests that we can embed the 2-dimensional cell complex T in \mathbb{R}^5 . Using an analog of book embedding an arbitrary graph in \mathbb{R}^3 [27], we can embed T in \mathbb{R}^4 . Therefore, we prove the problem is NP-hard to approximate for 1-dimensional homology when the input is the triangulation of a 3-manifold. This raises the open question that whether localizing a 1-dimensional class of a 2-manifold is NP-hard to approximate (it has already been proved to be NP-hard to compute).

A similar argument can be applied to other problems we will discuss in the next section, except that in Lemma 4.3.1, the relevant homology is 2-dimensional, the cell complex T is 3-dimensional and the manifold is 4-dimensional.

4.3 MinVolNBCyc is NP-hard to approximate within any constant factor

In the previous section, the simplicial complex we constructed for LocHomVol has $\Theta(n)$ Betti number. It has been revealed for 1-dimensional homology that

- MinVolNBCyc and MinVolBasis can be solved in polynomial time, and
- LocHomVol with $\beta_1 = O(1)$ can be solved in polynomial time when the input is the triangulation of a 2-manifold, with or without boundary.

This raises the question of whether these three problems are hard for homology of dimension two or higher. Our main result in this section is the inapproximability proof of a special case of MinVolNBCyc (Theorem 4.3.3). This trivially leads to the inapproximability of all the aforementioned problems (Corollary 4.3.4).

Lemma 4.3.1. *For 2-dimensional homology, even when $\beta_2 = 1$, MinVolNBCyc is NP-hard to approximate within any constant factor.*

Proof. We prove by a strict reduction from NearestCodeword, namely,

$$\text{NearestCodeword} \leq_S \text{MinVolNBCyc}.$$

Given an instance of NearestCodeword, we consider the generator matrix $C = [A, y_0]$ and its parity-check matrix C^\perp (the dimension is $(m - k - 1) \times m$). Following a scheme similar to Lemma 4.2.2 (illustrated in Figure 3), we construct a cell complex T_2 using C^\perp as the 2-dimensional boundary matrix. T_2 has $m - k - 1$ 1-cells and m 2-cells. There is a one-to-one correspondence between the 2-dimensional cycle group of T_2 and $\text{nullspace}(C^\perp) = \text{span}(C)$. This cycle group has dimension $k + 1$ and is spanned by the column vectors of A and y_0 .

Next, for each column vector of A , we seal the corresponding 2-cycle in T_2 with a 3-cell. T_2 is the 2-skeleton of the augmented complex, which is denoted as T . The one and only nontrivial 2-dimensional homology class of T is identical to the coset $y_0 + \text{span}(A)$. Finding the smallest volume nonbounding 2-cycle of T , denoted as MinVolNBCyc-T, is

equal to finding the minimal Hamming weight vector in this coset and thus equal to solving NearestCodeword. It suffices to show that MinVolNBCyc-T can be strictly reduced to MinVolNBCyc, by subdividing T .

In order to triangulate T into a simplicial complex K , we first subdivide the 2-skeleton, T_2 , into a simplicial complex K_2 , in which all 2-cells are triangulated into the same number of triangles (say, t_2). There is a one-to-one correspondence between the 2-dimensional cycle groups $Z_2(K_2)$ and $Z_2(T_2) = Z_2(T)$. The volume of each 2-cycle of K_2 is t_2 times that of its corresponding cycle.

Next, while keeping K_2 intact, we triangulate interior of the 3-cells fine enough so that for any nonbounding 2-cycle of K , $z \in [z_0]$ with z_0 carried by K_2 , we can always find in polynomial time a nonbounding 2-cycle of K_2 , $z' \in [z_0]$, with a smaller or equal volume. (This is similar to the triangulation strategy in Lemma 4.2.3.) Due to the one-to-one correspondence between $Z_2(K_2)$ and $Z_2(T)$ and the $t_2:1$ ratio of their volumes, we have a strict reduction from MinVolNBCyc-T to MinVolNBCyc. \square

Remark 4.3.2. *Whereas β_2 and β_3 of the constructed K are 1 and 0 respectively, the 1-dimensional Betti number, β_1 , could be linear in the size of K . However, we can remedy this by computing an arbitrary 1-dimensional homology cycle basis and seal all its elements with additional triangles. This will not influence the reduction. This way, we prove the inapproximability for complexes with bounded Betti numbers of all dimensions.*

Similar to Theorem 4.2.4, we can extend the result to any higher dimension by a suspension-building-based strict reduction of any MinVolNBCyc problem for $(d-1)$ -dimensional homology to that for the d -dimensional homology.

Theorem 4.3.3. *Even when the relevant Betti number is 1, MinVolNBCyc is NP-hard to approximate within any constant factor for homology of dimension two or higher.*

So far the inapproximability proof is for MinVolNBCyc with $\beta_d = 1$. This trivially leads to the inapproximability of the general MinVolNBCyc. Furthermore, we extend the inapproximability to the other two problems.

Corollary 4.3.4. *For homology of dimension two or higher, the following problems are NP-hard to approximate within any constant factor:*

1. MinVolBasis;
2. LocHomVol with fixed Betti number.

Proof. We show that the special case MinVolNBCyc can be computed in polynomial time from the output of the other two problems. This leads to the inapproximability.

Given the output of MinVolBasis, the homology cycle basis with the minimal total volume, the minimal volume nonbounding cycle is in this basis.

For LocHomVol with fixed Betti number, we enumerate all nontrivial classes and find their minimal volume representatives. The minimal volume nonbounding cycle is one of those representatives. □

4.4 A Polynomial Special Case

There is, however, a special case in which MinVolNBCyc can be computed in polynomial time, even with linear Betti number: when K is an N -dimensional complex embedded in \mathbb{R}^N and the pertinent nonbounding cycle is $(N - 1)$ -dimensional. In this section, we provide a polynomial time algorithm, inspired by [22, 7]. It is not hard to generalize this algorithm to MinVolBasis and LocHomVol.

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 $\mathbb{R}^N \setminus |K|$. 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.

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' . Denote vertex sets of G and G' as V and V' , respectively. The set of new vertices $V' \setminus V$ is dual to the set of new N -cells. See Figure 7 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 $V' \setminus V$. 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, $(v_1, v_2) \in (V' \setminus V) \times (V' \setminus V)$. Compute the minimal $(v_1 - v_2)$ -cut for each pair. The one with the smallest cardinality is the desired one.

Since the cardinality of $V' \setminus V$ is $\beta_{N-1} + 1$, the complexity of this algorithm is $O(\beta_{N-1}^2 f(n))$ where n is the size of the simplicial complex and $f(n)$ is the complexity of the MIN-CUT algorithm. Using MIN-CUT algorithms whose complexity is $O(n^2 \log n)$, the whole algorithm has complexity $O(\beta_{N-1}^2 n^2 \log n)$.

Remark 4.4.1. *The idea can be carried over to the case of a weighted volume function, but only if the weight function is non-negative.*

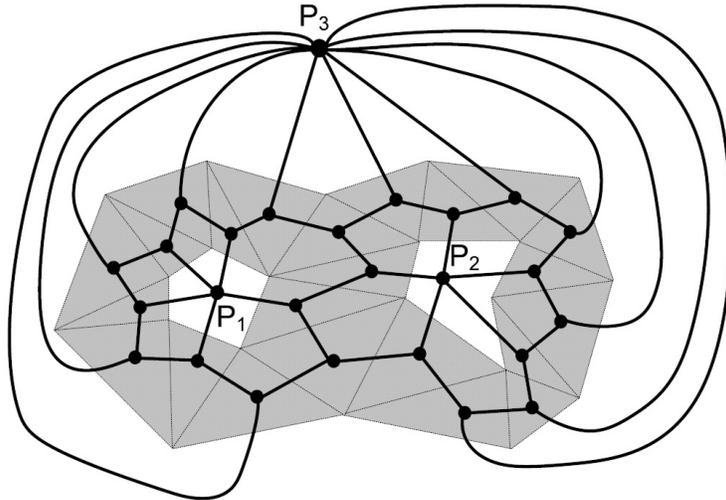


Figure 7: A 2-dimensional simplicial complex embedded in \mathbb{R}^2 . The dual graph G and G' are drawn in solid lines and vertices. Their difference, $G' \setminus G$, includes vertices p_1, p_2, p_3 and their incident edges.

5 Localization with Other Geometric Criteria

5.1 Diameter

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

Definition 5.1.1 (Diameter). *The diameter of a cycle is the diameter of its vertex set, $\text{diam}(z) = \text{diam}(\text{vert}(z))$, in which the diameter of a set of vertices is the maximal discrete geodesic distance between them, formally,*

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

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 5.2 by comparison against other criteria. We prove in Theorem 5.1.4 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 [1]. The theorem has been stated in our previous paper [9]. But the proof has not been published.

Remark 5.1.2. *We do not address the approximability of LocHomDiam, as we realize that z_d suffers from a “wiggling problem” and consequently may be geometrically complex (see Section 5.2). However, it is not hard to see that the reduction in Theorem 5.1.4 is strict, which implies that LocHomDiam 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 5.1.3 (Multiple-Choice Cover Problem).

INPUT: a set of vertices, $V = \{v_1, v_2, \dots, v_n\}$; a distance function $\text{dist} : V \times V \rightarrow \mathbb{R}^+$ satisfying triangular inequality; Disjoint subsets of V , S_1, S_2, \dots, 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: $\text{diam}(C)$

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 5.1.4. *LocHomDiam is NP-hard to compute.*

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

We first embed the vertex set V in any metric space preserving the pairwise distance $\text{dist}(\cdot, \cdot)$. Without loss of generality, we assume V is embedded in 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 through each vertex of S_i once without self-intersection, $\xi_i = (v_1, v_2, \dots, v_{\text{card}(S_i)})$, which contains $\text{card}(S_i) - 1$ edges. The edge lengths are the same as the distances between corresponding vertices. See Figure 8(a). We construct a slender threadlike tube T_i , which carries the path ξ_i . T_i has $(3 \text{card}(S_i))$ vertices, $S_i \cup S'_i \cup S''_i$, where

$$S'_i = \{v'_1, v'_2, \dots, v'_{\text{card}(S_i)}\}, \quad \text{and} \quad S''_i = \{v''_1, v''_2, \dots, v''_{\text{card}(S_i)}\}.$$

For any j , v'_j and v''_j lie very close to v_j . Corresponding to the $\text{card}(S_i) - 1$ edges in ξ_i , T_i consists of $\text{card}(S_i) - 1$ 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 8(b) for one such triangular cylinder.

We construct the simplicial complex, K , as follows. For any i , $T_i \subseteq K$. For any two vertices $v_1, v_2 \in V$, if they are not neighbors, add an edge connecting them, whose length

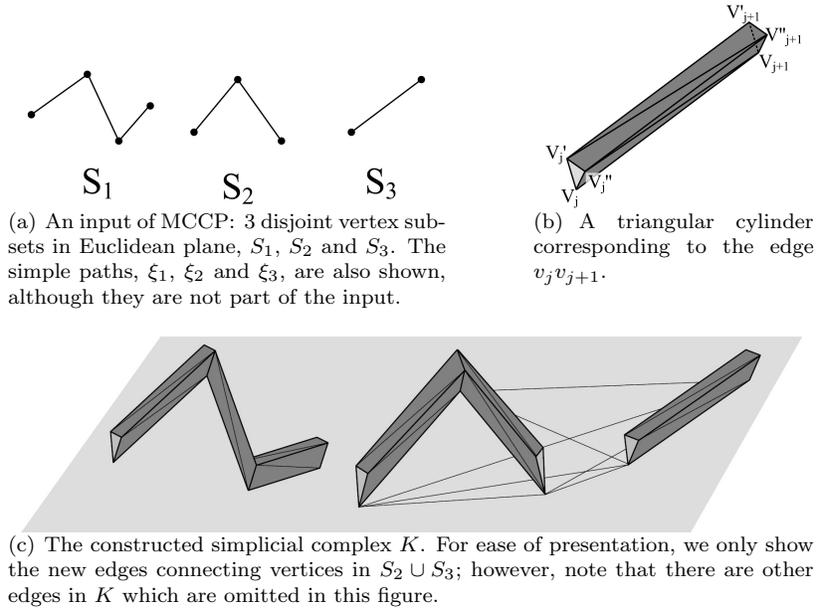


Figure 8: Explanation of theorem 5.1.4 proof.

is their Euclidean distance in the Euclidean plane \mathbb{R}^2 . See Figure 8(c) for the complex constructed from the input in Figure 8(a). 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 `LocHomDiam` 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 `LocHomDiam`. 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 `LocHomDiam`, 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 , $\text{vert}(z)$. 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,

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

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. \square

5.2 Radius

Another option for the cost function is radius.

Definition 5.2.1 (Radius). *The radius of a cycle is the radius of the smallest geodesic ball carrying it, formally,*

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

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 [9, 10]. Intuitively, z_r is the cycle whose vertices are as close to a vertex of K as possible. Theorem 5.2.3 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 9(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 9(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 9(a), its diameter is twice of its radius. For the cycle z_d in Figure 9(b), its diameter is equal to its radius.

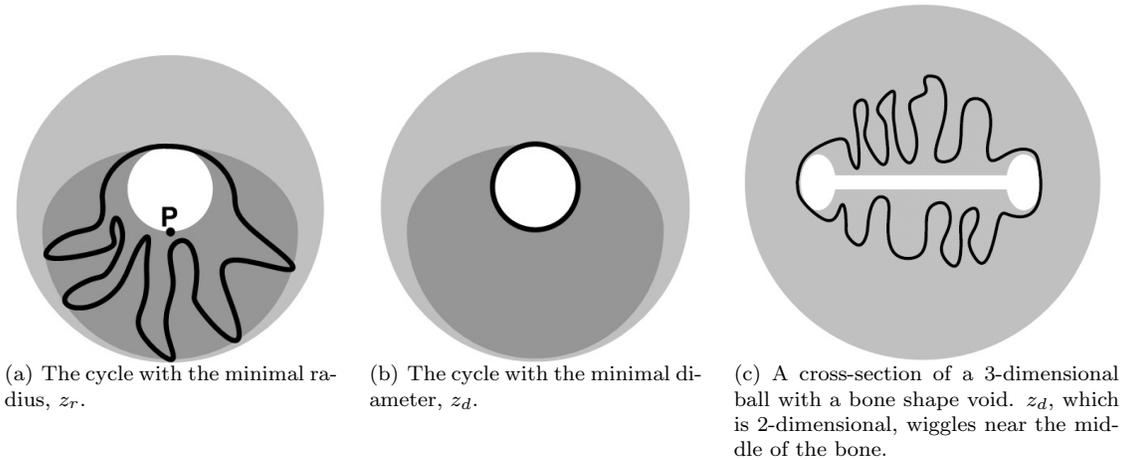


Figure 9: Cycles with the minimal radius and diameter.

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

Theorem 5.2.2. $\text{diam}(z_r) \leq 2 \text{diam}(z_d)$.

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

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

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 $\text{diam}(z_d)$. This ball carries z_d . Finally,

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

□

As shown in Figure 9(a) and 9(b), this bound is a tight bound.

However, in general, the minimal diameter cycle also suffers from the wiggling problem. In Figure 9(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 time algorithm of Chen and Freedman [10].

Theorem 5.2.3. *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, $\{z_1, z_2, \dots, z_{\beta_d}\}$, 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, \dots, z_i , namely,

$$z_0 = [\partial_{d+1}, z_1, z_2, \dots, z_i]\gamma, \quad (2)$$

where d is the dimension of the class in question. This can be computed by applying the persistence reduction to the matrix $[\partial_{d+1}, z_1, z_2, \dots, z_{\beta_d}]$. 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 $B_p^{r(p)}$, $[0, z_1, z_2, \dots, z_i]\gamma$.

The algorithm applies persistent homology algorithm $O(\text{card}(\text{vert}(K)))$ times. Therefore, the complexity is $O(n^4)$, where n is the size of the simplicial complex. \square

6 Conclusion

In this paper, we have addressed the localization problem with regard to three different measures. For volume, we have proved inapproximability results. We have also proved 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 time algorithm exists has also been discussed.

For diameter, we have proved 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.

We summarize all the constructed strict reductions in this paper as follows. The arguments of a problem – specifically, the dimension d of the relevant homology group, and Betti numbers of the input simplicial complex – are contained in square brackets.

- $\text{NearestCodeword} \leq_S \text{LocHomVol-T}[d = 1] \leq_S \text{LocHomVol}[d = 1]$
 $\leq_S \text{LocHomVol}[d \geq 1]$
- $\text{LocHomVol} \leq_S \text{LocHomVol}$, with manifold assumption
- $\text{NearestCodeword} \leq_S \text{MinVolNBCyc}[d = 2, \beta_2 = 1]$
 $\leq_S \text{MinVolNBCyc}[d \geq 2, \beta_d = 1]$
- $\text{MinVolNBCyc}[d \geq 2, \beta_d = 1] \leq_S \text{MinVolBasis}[d \geq 2]$
- $\text{MinVolNBCyc}[d \geq 2, \beta_d = 1] \leq_S \text{LocHomVol}[d \geq 2, \beta_d = O(1)]$
- $\text{MCCP} \leq_S \text{LocHomDiam}$

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 the volume measure has been shown to be extremely hard.

Acknowledgment

The authors thank David Cohen-Steiner and Omid Amini for constructive discussions. We thank anonymous reviewers for helpful comments.

References

- [1] E. M. Arkin and R. Hassin. Minimum-diameter covering problems. *Networks*, 36(3):147–155, 2000.
- [2] 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.
- [3] G. Ausiello and V. T. Paschos. Reductions, completeness and the hardness of approximability. *European Journal of Operational Research*, 172(3):719–739, 2006.
- [4] 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.
- [5] G. Carlsson. Topology and data. *Bull. Amer. Math. Soc.*, 46(2):255–308, 2009.
- [6] E. W. Chambers, J. Erickson, and A. Nayyeri. Homology flows, cohomology cuts. In *STOC*, 2009.
- [7] E. W. Chambers, J. Erickson, and A. Nayyeri. Minimum cuts and shortest homologous cycles. In *Symposium on Computational Geometry*, 2009.
- [8] C. Chen and D. Freedman. Quantifying homology classes ii: Localization and stability. *CoRR*, abs/0709.2512, 2007.

- [9] 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.
- [10] C. Chen and D. Freedman. Measuring and computing natural generators for homology groups. *Computational Geometry: Theory and Applications*, 43(2):169–181, February 2010.
- [11] 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.
- [12] V. de Silva and R. Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7:339–358, 2007.
- [13] T. K. Dey, A. N. Hirani, and B. Krishnamoorthy. Optimal homologous cycles, total unimodularity, and linear programming. In *STOC*, pages 221–230, 2010.
- [14] T. K. Dey, K. Li, and J. Sun. On computing handle and tunnel loops. In *IEEE Proc. NASAGEM*, 2007.
- [15] 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.
- [16] T. K. Dey, J. Sun, and Y. Wang. Approximating loops in a shortest homology basis from point data. In *Symposium on Computational Geometry*, pages 166–175, 2010.
- [17] J. Erickson and S. Har-Peled. Optimally cutting a surface into a disk. *Discrete & Computational Geometry*, 31(1):37–59, 2004.
- [18] J. Erickson and A. Nayyeri. Minimum cuts and shortest non-separating cycles via homology covers. In *SODA*, 2011.
- [19] 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.

- [20] 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.
- [21] I. Guskov and Z. J. Wood. Topological noise removal. In *Proceedings of the Graphics Interface 2004 Conference*, pages 19–26, 2001.
- [22] D. Kirsanov and S. J. Gortler. A discrete global minimization algorithm for continuous variational problems. Technical Report TR-14-04, Harvard University, 2004.
- [23] D. J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, 2002.
- [24] J. R. Munkres. *Elements of Algebraic Topology*. Addison-Wesley, Redwood City, California, 1984.
- [25] 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.
- [26] 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.
- [27] Wikipedia. Book embedding. http://en.wikipedia.org/wiki/Book_embedding.
- [28] Wikipedia. Suspension. [http://en.wikipedia.org/wiki/Suspension_\(topology\)](http://en.wikipedia.org/wiki/Suspension_(topology)).
- [29] Z. J. Wood, H. Hoppe, M. Desbrun, and P. Schröder. Removing excess topology from isosurfaces. *ACM Trans. Graph.*, 23(2):190–208, 2004.
- [30] A. Zomorodian and G. Carlsson. Localized homology. In *Proceedings of the 2007 International Conference on Shape Modeling and Applications*, pages 189–198, 2007.