Hardness Results for Homology Localization

Chao Chen, Daniel Freedman

HP Laboratories
HPL-2009-374

Abstract:
We address the problem of localizing homology classes, namely, finding the cycle representing a given class with the most concise geometric measure. We focus on the volume measure, 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 $O(1)$. A side effect is 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. We also discuss other geometric measures (diameter and radius) and show their disadvantages in homology localization. Our work is restricted to homology over the Z2 field.

Hardness Results for Homology Localization[*]

Chao Chen[†]   Daniel Freedman[‡]

Abstract

We address the problem of localizing homology classes, namely, finding the cycle representing a given class with the most concise geometric measure. We focus on the volume measure, 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 $O(1)$. A side effect is 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. We also discuss other geometric measures (diameter and radius) and show their disadvantages in homology localization. Our work is restricted to homology over the $\mathbb{Z}_2$ field.

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 [4], graphics [14], networks [11] and computational biology [10]. 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 [17, 25, 21]. 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 [16, 22]. 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 [3].

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.
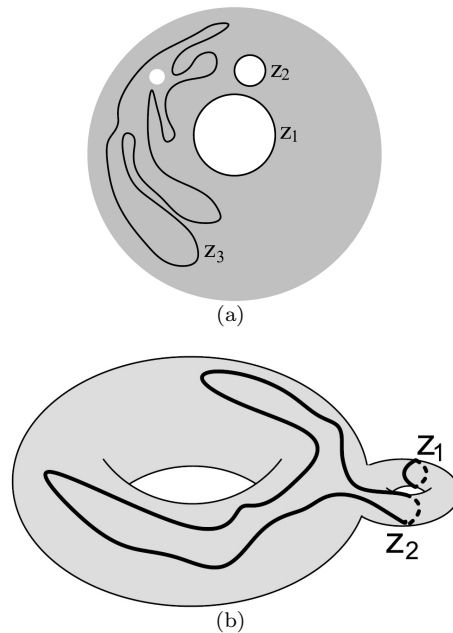

Figure 1: Motivating examples for localization.

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

[†]Institute of Science and Technology Austria and Vienna University of Technology, Austria.

[‡]Hewlett-Packard Laboratories, Israel.

localizing a given class with the minimal volume cycle is NP-hard to approximate within any constant factor. The proof is a strict reduction from the nearest codeword problem. We prove the inapproximability for homology of any dimension.

Second, we prove that for homology of dimension two or higher, computing the nonbounding 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 two dimensions 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 short discussion of other minimization criteria, including *diameter* and *radius*.

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

2 Preliminaries

2.1 Homology Groups. We briefly describe some background knowledge from algebraic topology. Please refer to [20] for more details. For simplicity, 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*, $\mathsf{C}_d(K)$. 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. [1] 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, $\mathsf{B}_d(K) = \mathrm{img}(\partial_{d+1})$. 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 [20]. Various reduction algorithms have been devised for different purposes.

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 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 $[b_1, ..., b_{n_d}]$, 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 $c \in \mathsf{C}_d(K)$ is *carried by* a subcomplex K_0 when all the d-simplices of c belong to K_0. We denote $\mathrm{vert}(K)$ as the set of vertices of the simplicial complex K, $\mathrm{vert}(c)$ 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 re-

---

[1] For 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.

alization. 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.

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. Please refer to [19] 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 by a Gauss-Jordan elimination of the transpose of A. Its dimension is $(m - k) \times m$.

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. Please see [2] 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 $I \in \mathcal{I}$, Sol(I) denotes the set of *feasible solutions* of I, and the *objective 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, card(I). It is also polynomial to verify whether any given S is a feasible solution, or evaluate the objective 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, $g(I_1, S_2)$.

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 $f(I_1)$, $S_2 \in \text{Sol}_{\Pi_2}(f(I_1))$, the performance ratios satisfy

$$(2.1) \qquad \rho_{\Pi_2}(f(I_1), S_2) \geq \rho_{\Pi_1}(I_1, g(I_1, S_2)).$$

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 2.1. *If* $\Pi_1 \leq_S \Pi_2$ *and* $\Pi_1 \notin$ APX*, then* $\Pi_2 \notin$ 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 [15] 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. [6] 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 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 manifold with boundary [7] and without boundary [6].

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 [17, 25] 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. [12] 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 [13] 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 [26] 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 [9] 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. Please see Section 8 for detailed discussion.

In terms of homology over other fields, the problem of finding the minimal volume representative does not have a direct analogy. A cycle with real coefficients can have arbitrarily small but nonzero volume. A minimal volume cycle with integer coefficients is not all that different in conception from the corresponding cycle over the $\mathbb{Z}_2$ field, but may be more complicated due to the torsion.

Chambers et al. [5] 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.

4 Problem Formalization and a List of Results

Given an objective function defined on all the d-cycles, cost : $\mathsf{Z}_d(K) \to \mathbb{R}$, we formalize the localization problem as a combinatorial optimization problem.

PROBLEM 4.1. (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: cost(z)

In this paper, we use volume as the objective function.

DEFINITION 4.1. (VOLUME) *The volume of a cycle is the number of its simplices,* vol$(z) =$ 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, $w : K \to \mathbb{R}$, 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, 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 4.2. (MIN-VOL NONBOUNDING CYCLE)
INPUT: a simplicial complex K with size n
OUTPUT: a nonbounding d-cycle z
MINIMIZE: vol(z)

PROBLEM 4.3. (MIN-VOL BASIS)
INPUT: a simplicial complex K with size n
OUTPUT: a homology cycle basis $\{z_1, z_2, \cdots, z_{\beta_d}\}$
MINIMIZE: $\sum_{i=1}^{\beta_d}$ vol(z_i)

For short, we name these two problems MinVolNBCyc and MinVolBasis, respectively.

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

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 has been proven to be NP-hard by polynomial reductions from a special case of MAX-2SAT [7] and MIN-CUT with negative edge weights [6].

- Chambers et al. [6] provided a polynomial algorithm for LocHomVol when β_1 is fixed. 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 [15] devised a polynomial algorithm for MinVolBasis, even when $\beta_1 = \Theta(n)$. This work is restricted to triangulations of 2-manifolds. A natural extension of the algorithm (together with [9]) can compute MinVolNBCyc and MinVolBasis in polynomial time 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 5.1).

- When the homology in question is 2-dimensional or higher, we prove that MinVolNBCyc is NP-hard to approximate within any constant factor (Theorem 6.1). So do LocHomVol with $\beta_d = O(1)$ and MinVolBasis (Corollary 6.1).

- 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.

5   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 [1]. Problems used in previous reductions to LocHomVol [7, 6] have constant approximation ratios, and thus cannot be used for our proof.

PROBLEM 5.1. (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 5.1. *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. Please 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 2 for an example with a 4×2 generator matrix

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

As each 1-chain of T is a 1-cycle, it is not hard to see that 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 which
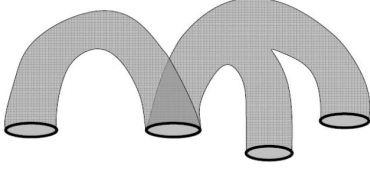
Figure 2: The constructed cell complex, T. Two 2-cells (pipes) share four 1-cells (thickened circles), corresponding to two columns and four rows of A.

is supposed to be a triangulation of a topological space. Next, we subdivide T into a simplicial complex K. With this construction, we will strictly reduce LocHomVol-T to LocHomVol.

We first triangulate each 1-cell of T into t_1 edges, with t_1 fixed and small. For convenience, we denote the triangulation of all 1-cells of T as 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, y, and its corresponding 1-cycle of K_1, $\phi(y)$, the ratio of their volumes is $1 : t_1$.

Next, we triangulate the 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, $z \in [z_0]$, we can compute in polynomial time a cycle z' carried by K_1, which is homologous to z and has a smaller or equal volume. More details of the triangulation and generating z' from z can be found in Appendix A.

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 transform z into z' and then into a solution of LocHomVol-T, $\phi^{-1}(z')$. For convenience, we denote this solution $g(z)$. Lastly, we prove 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. The ratio of their volumes is $\mathrm{vol}(z_v) : \mathrm{vol}(g(z_v)) = t_1 : 1$. Second, for any feasible solution z, the volume of its corresponding solution in LocHomVol-T is

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

and therefore, $\dfrac{\mathrm{vol}(z)}{\mathrm{vol}(z_v)} \geq \dfrac{\mathrm{vol}(g(z))}{\mathrm{vol}(g(z_v))}.$

This guarantees Inequality (2.1), and thus the strictness of the reduction. $\square$

Lemma 5.1 is about 1-dimensional homology. We extend the result to homology of any higher dimension.

THEOREM 5.1. *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, LocHomVol$_{d-1}$ $\leq_S$ LocHomVol$_d$. Together with Lemma 5.1, 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 [24]. 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 then modify Lemma 5.1 and its proof to accommodate this manifold assumption. Specifically, we can 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. This leads to the inapproximability of LocHomVol for 1-dimensional homology when the input is the triangulation of an $(N-1)$-manifold.

A classical result suggests that we can embed the 2-dimensional cell complex T in $\mathbb{R}^5$. By using an analog of book embedding an arbitrary graph in $\mathbb{R}^3$ [23], 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 proven 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 6.1, the relevant homology is 2-dimensional, the cell complex T is 3-dimensional and the manifold is 5-dimensional.

6 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

1. MinVolNBCyc and MinVolBasis can be solved in polynomial time, and

2. 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 6.1). This trivially leads to the inapproximability of all the aforementioned problems (Corollary 6.1).

LEMMA 6.1. *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 5.1 (illustrated in Figure 2), we construct a cell complex T_2 using $C^\perp$ as the 2-dimensional boundary matrix. There is a one-to-one correspondence between the 2-dimensional cycle group of T_2 and nullspace($C^\perp$) = span(C). This cycle group has rank(A) + 1 = $k + 1$ 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, 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 $\mathsf{Z}_2(K_2)$ and $\mathsf{Z}_2(T_2) = \mathsf{Z}_2(T)$. 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 as finely as possible so that for any nonbounding 2-cycle of K, z, we can always find in polynomial time a nonbounding 2-cycle of K_2, z',

which is homologous to z. (This is similar to the triangulation strategy in Lemma 5.1, which is explained in Appendix A.) Due to the one-to-one correspondence between $\mathsf{Z}_2(K_2)$ and $\mathsf{Z}_2(T)$ and the t_2:1 ratio of their volumes, we have a strict reduction from MinVolNBCyc-T to MinVolNBCyc. $\qquad\square$

REMARK 6.1. *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. It is not hard to see that this will not influence the reduction. This way, we prove the inapproximability for complexes with bounded Betti numbers of all dimensions.*

Similar to Theorem 5.1, 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 of the d-dimensional homology.

THEOREM 6.1. *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 6.1. *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. $\qquad\square$

7 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 algorithm, inspired by [18, 6]. 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 $\mathbb{R}^N \backslash |K|$. 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.

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 $V' \backslash V$ 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.

Since the cardinality of $V' \backslash V$ 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 7.1. *The idea can be carried over to the case of a weighted volume function, but only if the weight function is non-negative.*
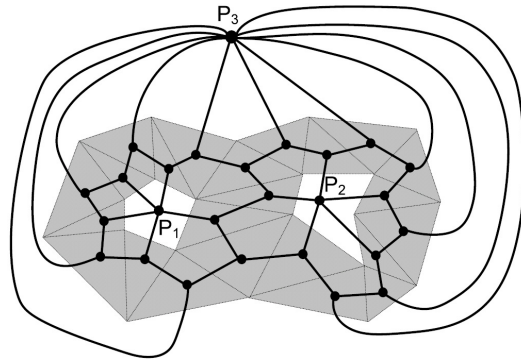

Figure 3: A 2-dimensional simplicial complex embedded in $\mathbb{R}^2$. The dual graph G and G' are drawn in solid lines and vertices. Their difference, $G' \backslash G$, includes vertices p_1, p_2, p_3 and their incident edges.

8 Localizing with Other Geometric Criteria

Since localizing a class with the minimal volume is extremely difficult, we could resort to other geometric criteria as the objective function for optimization. In this section, we discuss two such criteria, diameter and radius of a cycle. We briefly explain the definitions, and then show that these criteria suffer from a "wiggling problem". We end by quoting relevant results which have been proven in our previous work [8, 7].

Given a simplicial complex K and nonnegative lengths defined on each of its edges, the *discrete geodesic distance* between any two vertices, $d : \text{vert}(K) \times \text{vert}(K) \to \mathbb{R}$, is defined as the length of the shortest path in the 1-skeleton of K. Given $p \in \text{vert}(K)$, $r \geq 0$, the *discrete geodesic ball*, B_p^r, centered at p with radius r, is the maximal subcomplex whose vertices' discrete geodesic distances from p are no greater than r. The *diameter* of a cycle, z, is the maximal pairwise discrete geodesic distance of the vertices in $\text{vert}(z)$, $\max_{p,q \in \text{vert}(z)} d(p,q)$. The *radius* of z is the smallest radius of discrete geodesic balls carrying z.

We denote z_d (resp. z_r) as the representative cycle of a given class with the minimal diameter (resp. radius). These cycles seem to be good substitutes for the minimal volume representative cycle, z_v. However, both z_d and z_r suffer from a "wiggling problem" and are not geometrically concise.

For example, in an annulus (Figure 4(a)), z_r wiggles freely inside the geodesic ball (centered at p, dark grey area) carrying it. In Figure 4(b), we show 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 between 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 $\text{diam}(z_d)$.

For completeness, we quote previous results concerning the computation of z_d and z_r.

- z_d is NP-hard to compute;

- z_r can be computed in polynomial time;

- $\text{diam}(z_r) \leq 2 \text{diam}(z_d)$. This is a tight bound.

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

(b) A cross-section of a 3-ball with a bone shape void, and the 2-dimensional z_d.
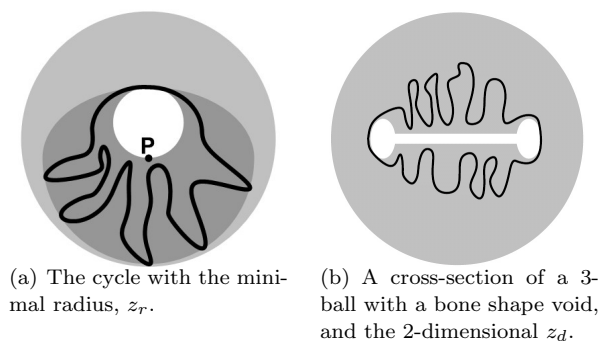
Figure 4: Wiggling cases.

9 Discussion

In this paper, we have proved inapproximability of localization with minimal volume. An open question is whether we can use other discrete geodesic distance related measures for localization, 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.

For the volume measure, there are still unsolved questions when the relevant homology is 1-dimensional. For example, is there a polynomial-time algorithm for LocHomVol with a fixed Betti number? Is LocHomVol with $\beta_1 = \Theta(n)$ NP-hard to approximate when the input is a 2-manifold?

References

[1] 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.

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

[3] 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.

[4] G. Carlsson. Topology and data. *Bull. Amer. Math. Soc.*, 46(2):255–308, 2009.

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

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

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

[8] 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.

[9] C. Chen and D. Freedman. Measuring and computing natural generators for homology groups. *Computational Geometry: Theory and Applications*, 43(2):169–181, February 2010.

[10] 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.

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

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

[13] 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.

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

[15] 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.

[16] 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.

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

[18] D. Kirsanov and S. J. Gortler. A discrete global minimization algorithm for continuous variational prob-

lems. Technical Report TR-14-04, Harvard University, 2004.

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

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

[21] 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.

[22] 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.

[23] Wikipedia. Book embedding. `http://en.wikipedia.org/wiki/Book\_embedding`.

[24] Wikipedia. Suspension. `http://en.wikipedia.org/wiki/Suspension\_(topology)`.

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

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

Appendix

### A Details of Subdividing T in Lemma 5.1

We explain details of triangulating T finely, so that for any cycle of K, $z \in [z_0]$, we can compute in polynomial time a cycle $z' \in [z_0]$, which is carried by the subcomplex K_1, with the volume $\mathrm{vol}(z') \leq \mathrm{vol}(z)$.

For convenience, we introduce some notations. We call a 1-chain c a *simple path* if $\mathrm{card}(c) = \mathrm{card}(\mathrm{vert}(c)) + 1$, and there is a non-repeating sequence of $\mathrm{vert}(c)$, $(v_1, v_2, \ldots, v_k)$, 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, $\mathrm{card}(c) = \mathrm{card}(\mathrm{vert}(c))$. We extend the definition of homologous to chains. Two chains are *homologous* to each other if their difference is a boundary.

Recall that we triangulate each 1-cell of T into t_1 edges. The triangulation of the 1-skeleton of T is a subcomplex K_1. Recall m is the number of 1-cells of T. For each 2-cell of T, we triangulate it as fine as possible while keeping K_1 intact. See Figure 5 for the triangulation of a 2-cell σ whose boundary has 4 1-cells.

Given this triangulation, the polynomial time trans-

---

[2]This definition is consistent with the definition in graph theory.
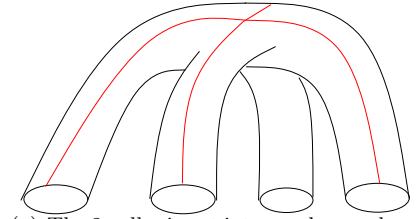
(a) The 2-cell σ is cut into a polygon along the red curves.
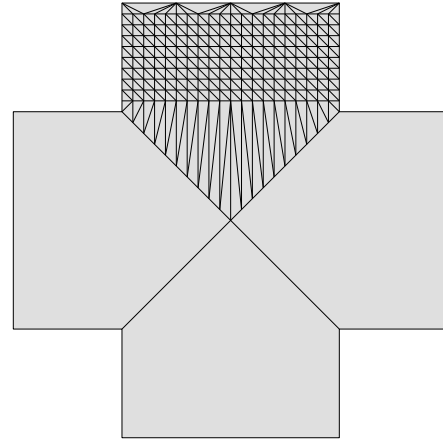
(b) A fine triangulation of the polygon. For simplicity, we only draw 1/4 of the triangulation.

Figure 5: The triangulation of a 2-cell of T whose boundary has 4 1-cells, when $t_1 = 4$ and $m = 5$.

formation of z into z' can be achieved as follows. 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. There are three cases. Recall that ϕ maps a chain of T to its subdivision.

1. Any simple cycle or simple path is carried by the triangulation of one 2-cell of T, σ. A simple 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. See Figure 6(a) for an example.

2. For a simple path whose both end vertices are from the triangulation of a same 1-cell $\tau \in \partial\sigma$, 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. See Figure 6(b).

3. Suppose it is a simple path connecting vertices from the triangulations of two different 1-cells (Figure 6(c)). We triangulate the 2-cell σ as fine as possible

so that any such path has a volume of at least mt_1. In such case, we just let z' be the input z_0, whose volume is no greater than mt_1, and thus no greater than $\text{vol}(z)$. The fine triangulation in Figure 5 achieves this objective when $t_1 = 4$ and $m = 5$.

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.
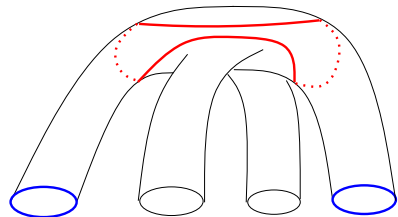
(a) Case 1: a simple cycle (red) is homologous to a 1-cycle (blue) carried by K_1. Note the latter cycle has two components.
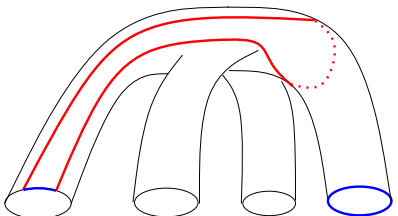
(b) Case 2: a simple path (red) whose end vertices are from the triangulation of a same 1-cell is homologous to a 1-chain (blue) carried by K_1.
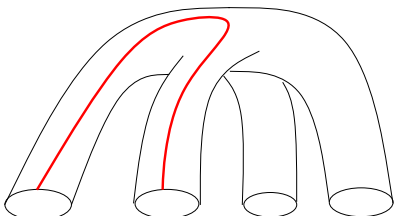
(c) Case 3: a simple path (red) connecting vertices from the triangulation of two 1-cells is at least mt_1 long.

Figure 6: Different cases for generating z'.