World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Research Article)

Check for updates

An approximation to Čech complex using median of triangles for computing Betti numbers of some point cloud data

Md. Morshed Bin Shiraj [1, *], Md. Mizanur Rahman [1], Md. Manik Hossain [1], Md. Dalim Haque [1, 2], Md. Al-Imran [1], Mst Zinia Afroz Liza [1], Md. Masum Murshed [1] and Nasima Akhter [1]

[1] *Department of Mathematics, University of Rajshahi, Rajshahi-6205, Bangladesh.*

[2] *Department of Mathematics, Rajshahi University of Engineering and Technology, Rajshahi-6204, Bangladesh.*

World Journal of Advanced Engineering Technology and Sciences, 2023, 09(02), 037–048

Publication history: Received on 31 May 2023; revised on 07 July 2023; accepted on 10 July 2023

Article DOI: https://doi.org/10.30574/wjaets.2023.9.2.0201

Abstract

An approach has been developed to create an approximated simplicial complex in between the Vietoris-Rips complex and the Čech complex using median of triangles for computing Betti numbers of some point cloud data. The Vietoris-Rips complex has been built first for this. Then the sample points have been classified into three classes based on three conditions of the median (l) of any triangle's maximum edge $(2r)$ for any three points in $\mathbb{R}^n$. Then the values of the filtration (ε) have been chosen in such a way that $\varepsilon = r$ for $l < r$, $\varepsilon = r$ for $l = r$, and $\varepsilon = r + \frac{(l-r)}{3}$ for $l > r$. The approach has been extended for higher dimensional triangles calculating l by the distance of the centroid from the opposite vertex of the maximum face and considering r as the filtration value of the maximum edge. Then an algorithm has been introduced to calculate the simplicial complex after building simplices for each filtration value. Finally, to validate the study results of the approximated simplicial complex have been compared with the Vietoris-Rips complex and the Čech complex. The proposed approximated simplicial complex has been found computationally effective than the Čech complex and its filtration values are lying between filtration values of the Vietoris-Rips complex and the Čech complex without any loss of persistent data.

Keywords: Approximated simplicial complex; Čech complex; Vietoris–Rips complex; Persistent homology; Topological data analysis; Point cloud data; Betti numbers.

1. Introduction

The development of modern science and technology is highly depending on data. Every day data are generated persistently in different fields. A large amount of data are high dimensional with a lot of noise. Analyzing these types of data is very challenging. Topological data analysis (TDA) is one of the most popular and vital tools to analyze these data that can analyze higher dimensional data with noise. TDA mainly analyzes data to calculate the shape of data using persistent homology. Persistent homology is a methodology of calculating the homology of a chain complex for some filtrations to track persistence of the homology. Though the calculation of homology is basically introduced via singular homology [1], the algebra of infinitely generated modules makes its computation very difficult to use in practice. In contrast, simplicial homology is quite easy to calculate because of its finite nature.

Simplicial homology is the homology calculated from a simplicial complex which is a collection of simplices. *Simplices* are points, edges, triangles, and higher dimensional triangles. There are many ways to construct a simplicial complex from some point cloud data which begins with the nerve of a topological space.

Let U be an underlying topological space of a set of vertices A. Let $C = \{C_\alpha\}_{\alpha\epsilon A}$ be any covering of X.

* Corresponding author: Md. Morshed Bin Shiraj (mbshiraj@gmail.com)

Definition [2]: The *nerve* of C, $\mathcal{N}(C)$ is an abstract simplicial complex of A where a k-simplex forms if the intersection of all coverings of $k+1$ vertices of A is non-empty, i.e., if for any $\{a_1, a_2, \dots, a_k\} \subset A, C_{a_0} \cap C_{a_1} \cap \dots \cap C_{a_k} \neq \emptyset$, $\{a_1, a_2, \dots, a_k\}$ forms a k-simplex.

As mentioned in [3], *Nerve Theory* confirms that the nerve $\mathcal{N}(C)$ is homotopic equivalent to U. Now the mathematical question is how to build the covering. To answer this question there are many approaches have been introduced. In this study, only two of them have been reviewed to be on the track we followed.

Firstly, open balls can be considered as coverings of the space U, $B_k(U) = \{B_k(u)\}_{u \epsilon U}$. For any $U_v \subset U$ such that $U_v \in \cup_{v \epsilon U_v} B_k(v)$, the Nerve can be defined as a simplicial complex of $\{B_k(\text{v})\}_{v \epsilon U_v}$ which is known as *Čech complex* $\check{\text{C}}(U_v, k)$. Building Čech complex is computationally difficult because of determining the intersection of open balls. It requires building higher dimensional balls to go through the process which is computationally expensive.

Secondly, one can make a simplicial complex called *Vietoris-Rips complex* without building open balls by using the information of vertices and edges only. The Vietoris-Rips complex is a simplicial complex for a certain parameter ε, such that k-simplex forms for any $\{a_1, a_2, \dots, a_k\} \subset A$ if and only if $d(a_i, a_j) \leq \varepsilon, \forall\, 0 \leq i, j \leq k$, i.e., any two points are pairwise close enough than ε.

The Vietoris-Rips complex is computationally fast but it loses some information about small cycles between certain parameters, say ε_1, for which vertices of the cycle are pairwise close enough and ε_2 for which intersection of open balls of all vertices of the cycle is non-empty. It means there is a possibility to build a new simplicial complex between the Vietoris-Rips complex and the Čech complex.

There are few studies like [4]-[16] where different constructions of the Čech complex, the Vietoris-Rips complex, and other approximated complexes to these two complexes have been introduced. But none of these study counts geometrical approximation concerning the median of triangles.

In this study, the Vietoris-Rips complex has been constructed using a set-theoretic approach and then an approximation technique has been introduced to construct a simplicial complex between Vietoris-Rips and Čech complex to calculate Betti numbers of some point cloud data.

2. Background

2.1. Statement of the problem

To understand the problem of loss of persistent data due to the Vietoris-Rips complex, Vietoris-Rips complex and Čech complex of a sampled data of 4 points [(0,0), (1,0), (0.75,0.75) , (0,0.95)] have been computed and shown in Figure 1.

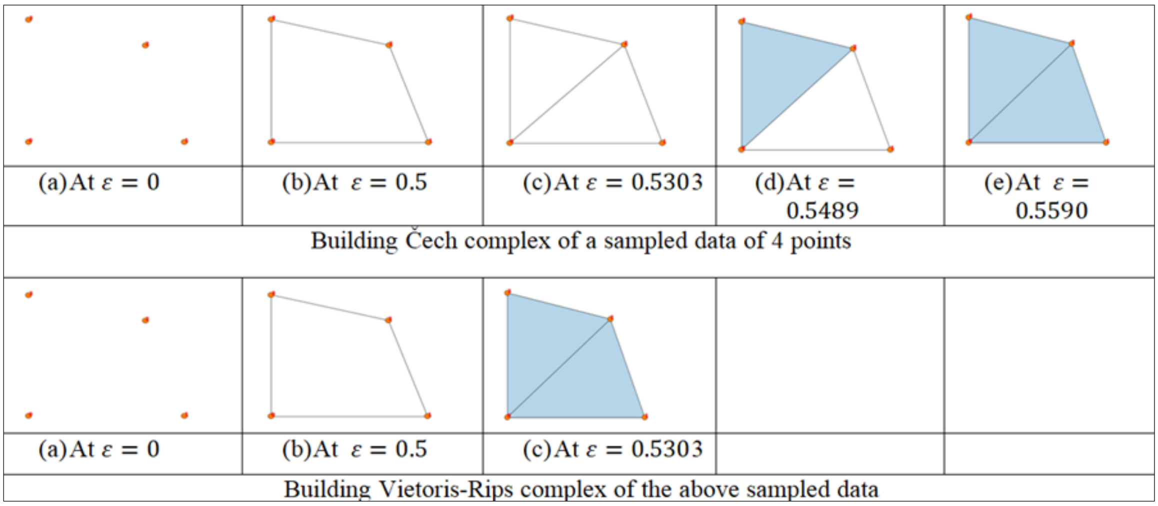


Figure 1 loss of information of cycles due to the Vietoris-Rips complex compared to the Čech complex.

Let ε be a parameter indicating the radius of each mini ball centered at each of the sampled data points. For $\varepsilon = 0.5$ both the Čech complex and Vietoris-Rips complex have formed the cycle [0,1,2,3]. But for $\varepsilon = 0.53$ Čech complex is detecting two loops where the Vietoris-Rips complex has no cycle. For $\varepsilon = 0.597$ Čech complex has one loop and for $\varepsilon = 0.617$ Čech complex has no cycle left. Thus Čech complex detected three cycles. Each cycle has birth and death. The value of the parameter ε for which a cycle forms is called *birth* and for which value of ε any two vertices of the cycle joined with an edge is called *death* of the cycle. Thus we can form a point $(birth,\ death)$ as persistent data of the cycle. Thus the cycles of Čech complex and Vietoris-Rips complex generate the following Table 1 between $0 \leq \varepsilon \leq 0.617$.

Table 1 Cycles of the sampled data showed in Figure 1.

Čech complex		**Vietoris-Rips Complex**	
Cycles	**(birth, death)**	**Cycles**	**(birth, death)**
[0123]	(0.5, 0.5303)	[0123]	(0.5, 0.5303)
[023]	(0.5303, 0.5489)		
[012]	(0.5303, 0.5590)		

From Table 1 and Figure 1, it is clear that the Vietoris-Rips complex is an approximation to Čech complex but it loses persistent data of small cycles.

On the other hand, the Čech complex is computationally expansive. Therefore the problem is to construct such a simplicial complex that is computationally effective and can detect all cycles as Čech complex.

2.2. Construction of Vietoris-Rips complex

To construct a simplicial complex between the Čech complex and Vietoris-Rips complex from a set of point cloud data, Vietoris-Rips complex has been constructed and then it has been developed into a desired simplicial complex. Let X be the set of some sample data points. Then the following algorithm has been followed to construct the Vietoris-Rips complex of the sample points.

2.2.1. ALGORITHM 1: Constructing Vietoris-Rips complex

- **Step 1**: Let $m =$ number of sample points. Define each point of X as a set of Natural numbers $\mathbb{N}$, then the indexed set $I = \{0, 1, 2, \dots, m-1\}$.
- **Step 2**: Construct a matrix $[D]_{m \times m}$ of pairwise distances assuming indexed points as rows and columns. Then $[D] = \{a_{ij}$ is the pairwise distance between i^{th} and j^{th} indexed point: $i, j \epsilon I\}$.
- **Step 3**: Make a list L of all entries of $[D]$ without repeating any numbers.
- **Step 4**: Consider $\varepsilon = \frac{L}{2}$ for each element of L.
- **Step 5**: For each ε, calculate the list of all points as 0-simplices denoted as $C0$, where $C0 = [[i]: i \epsilon I]$.
- **Step 6**: For each ε, calculate the list of 1-simplices, denoted as $C1$, where $C1 = [[i, j]$: i^{th} and j^{th} points pairwise distance $d_{ij} \leq 2\varepsilon$].
- **Step 7**: For each ε, calculate the list of 2-simplices, denoted as $C2$, where $C2 = [[i, j, k]$: for any two elements of $C1$ that have the common intersection of a singleton point, if the symmetric difference of the two elements is in $C1]$. In other words, $C2 = [[i, j, k]$: if the length of $[i, j] \cap [j, k]$ or, $[i, k] \cap [j, k]$ or, $[i, j] \cap [i, k] = 1$ and symmetric difference of any pairs among $\{[i, j], [j, k], [i, k]\} \epsilon C1]$.
- **Step 8**: Calculate the list of 3-simplices, denoted as $C3$, where $C3 = [[i, j, k, l]$: for any three elements of $C2$ that have the common intersection of a singleton point, if the union of the three elements is in $C2]$.
- **Step 9**: Accumulate the simplicial complex as a list, say SC, where $SC = [C0, C1, C2, C3]$.

2.3. Approximation of Vietoris-Rips complex to Čech complex:

To compute such an approximated simplicial complex between Vietoris-Rips and Čech complex, the main problem is to calculate ε values for which 2-simplices and 3-simplices will be formed. Before going to the algorithm, let's discuss the observation that gives us a strong hypothesis of triangulation.

Let us assume three random sample points x, y, z. They can be chosen from anywhere on a circle. Assume that these three points are in such an arrangement that $d(x,y) > d(y,z) \geq d(x,z)$, where d refers to the distance between two points. Let m be the middle point of x and y and let l be the line connecting z and m, called the *median* of the xy line.

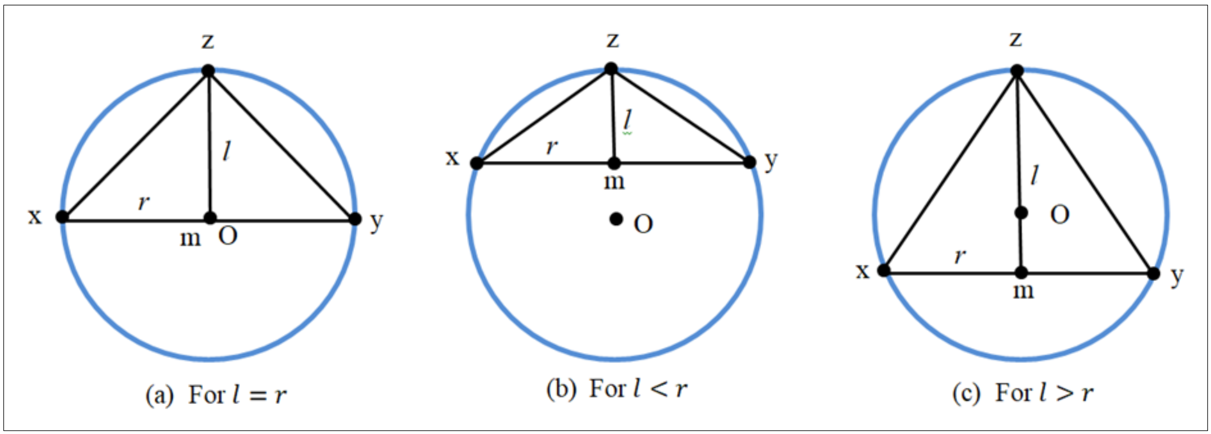


Figure 2 Three classes of the choice of 3 random points from a circle to form a 2-simplex.

Let $xm = my = r$. Then the three classes shown in Figure 2 of these three points can be made for $(a)\ l = r$, (b) for $l < r$, and (c) for $l > r$.

For case $(a) l = r$, x, y, and z are equidistant from m and so the triangle $\triangle\, xyz$ forms at $\varepsilon = r = l$.

For case $(b)\ l < r$, $\triangle\, xyz$ will form immediately joining xy because z is closer than x to m. That is $\triangle\, xyz$ forms at $\varepsilon = r$.

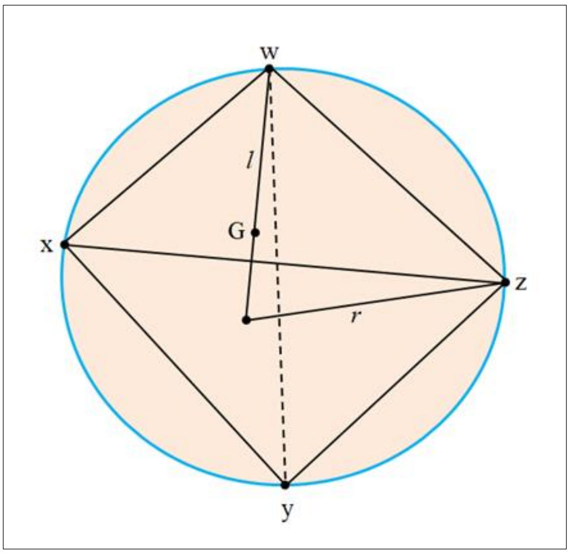


Figure 3 Four random sample points on a sphere to form a 3-simplex.

For case $(c)\ l > r$, point z is far from x to m, and of course $\varepsilon > r$. Let point z is approaching towards m such that $l \to r$ then $\varepsilon \to r$. That is, $(l - r) \to 0$. Let us assume $\varepsilon = r + \frac{(l-r)}{3}$ considering the above-approaching nature.

To calculate 3-simplices of any four sample points, let x, y, z, w be such random four points on a sphere shown in Figure 3 that values of ε to form four triangles $\triangle\, xyz$, $\triangle\, xzw$, $\triangle\, xyw$, $\triangle\, yzw$ are $\varepsilon_{xyz} > \varepsilon_{xzw} > \varepsilon_{xyw} > \varepsilon_{yzw}$.

To follow the same approach discussed earlier to form 2-simplices, let $r_h = max\,(\varepsilon_{xyz}, \varepsilon_{xzw}, \varepsilon_{xyw}, \varepsilon_{yzw})$. Since $\triangle\, xyz$ forms at last, let us assume l_h has been drawn from the w to xyz plane. Let G be the centroid of the tetrahedron $xyzw$. Then $d(w, G)$ can be calculated easily.

Since centroid divides each median by 2: 1 ratio, assume that, $l_h = d(w, G) + \frac{d(w,G)}{3}$. Thus the three classes of a tetrahedron become (a) for $l_h = r_h$, (b) for $l_h < r_h$ and (c) for $l_h > r_h$. For cases (a) and (b) $l_h \leq r_h$, the value of ε is $\varepsilon = r_h$. For case (c) $l_h < r_h$, the value of ε is $\varepsilon = r_h + \frac{(l_h - r_h)}{3}$.

To calculate the simplicial complex induced from the median of triangles calculated earlier, the following algorithm has been followed.

2.3.1. ALGORITHM 2: Constructing an approximated simplicial complex to Čech complex

- **Step 1**: Follow steps 1 to 4 of ALGORITHM 1.
- **Step 2:** Let $Triags$ be all possible triangles taking a combination of three points out of all sample points.
- **Step 3:** For each triangle $\in Triags$, find the maximum distant points and let m be the midpoint of them.
- **Step 4:** For each triangle $\in Triags$, let $r = x_m$ as discussed above.
- **Step 5:** For each triangle $\in Triags$, let $l = d(z, m)$ as discussed above.
- **Step 6:** For each triangle $\in Triags$, calculate ε for three classes as discussed above (a) $\varepsilon = r$ if $r \geq l$ and (b) $\varepsilon = r + \frac{(l-r)}{3}$ if $r < l$.
- **Step 7:** Let $Thedrons$ be all possible tetrahedrons taking a combination of four points out of all sample points.
- **Step 8:** For each tetrahedron $\in Thedrons$, find the triangle formed at last, i.e., for which ε is maximum.
- **Step 9:** For each tetrahedron $\in Thedrons$, find r_h =maximum ε calculated in step 8.
- **Step 10:** Calculate the centroid G of each tetrahedron.
- **Step 11:** Calculate $d(w, G)$ as discussed above.
- **Step 12:** Calculate $l_h = d(w, G) +) + d\frac{(w,G)}{3}$.
- **Step 13:** For each tetrahedron $\in Thedrons$, calculate ε, where (a) $\varepsilon = r_h$ if $r_h \geq l_h$ and (b) $\varepsilon = r_h + \frac{(l_h - r_h)}{3}$ if $r_h < l_h$.
- **Step 14:** Merge all ε into one list.
- **Step 15:** Follow steps 5-6 of ALGORITHM 1 for each ε.
- **Step 16:** For each ε, calculate the list of 2-simplices, denoted by $C2$, where $C2$ = {all of those triangles that have $\varepsilon \leq$ the trial ε}.
- **Step 17:** For each ε, calculate the list of 3-simplices, denoted as $C3$, where $C3$ = {all of those tetrahedrons that have $\varepsilon \leq$ the trial ε}.
- **Step 18:** Accumulate the desired simplicial complex, say SC, where $SC = [C0, C1, C2, C3]$.

3. Methodology

The following research methodology has been followed to finish this study:

- Literatures have been reviewed to find the motivation and background of this research.
- A Python routine has been written to build the Vietoris-Rips complex of some sampled data points following ALGORITHM 1 in section 2.2.1.
- Several sets of data points have been made to check the Python routine.
- Another Python routine has been coded to calculate Betti numbers ($\beta_0, \beta_1, \beta_2$).
- Then following ALGORITHM 2 in section 2.3.1 a separate Python routine has been executed and Betti numbers ($\beta_0, \beta_1, \beta_2$) have been calculated using the mogutda package developed by Kwan-Yuet Ho, Isla Staden, and Filip Cornell.
- To visualize simplicial complex, Python routine has been prepared following [17].
- A particular example of calculating Čech complex from 11 point data has been found in [18] that has been compared with our result.
- Vietoris-Rips complex has been computed using the Gudhi package and results have been compared to validate our model.
- Finally, the results have been analyzed in section 4.

4. Results and discussion

An array of 11 points $[(1,0), (0,1), (2,1), (3,2), (0,3), (3+\sqrt{3}, 3), (1,4), (3,4)\ (2, 4+\sqrt{3}), (0,4),$ and $(-0.5, 2)]$ has been used to compare our result with [16]. According to the example of computing Čech complex from the array of 11 points in [16], the ε-values are 0, 0.5, 0.559017, 0.707107, and 1 which are the same as calculated ε-values (see Table 2) within [0, 1] interval. There are three 2-simplex have been found [9,6,4] at 0.707107; [2,1,0], and [10,4,1] at $\varepsilon = 1$.

Table 2 Comparison between filtration values of Vietoris-Rips complex and the proposed approximated simplicial complex

ε −values of Vietoris-Rips complex	ε −values of the proposed approximated simplicial complex
0.0	0.0
0.5	0.5
0.5590169943749475	0.5590169943749475
0.7071067811865476	0.7071067811865476
0.9999999999999991	**0.804737854124365**
1.0	0.9999999999999991
1.0307764064044151	1.0
1.25	1.0307764064044151
1.3228756555322954	**1.1937129433613967**
1.346291201783626	**1.2142305476310067**
1.4142135623730951	1.25
1.5	1.3228756555322954
1.5811388300841898	1.346291201783626
1.6929339632083815	**1.3603796100280632**
1.75	**1.408963380382927**
1.8027756377319946	1.4142135623730951
1.931851652578136	**1.4698553182767933**
1.9318516525781364	1.5
2.0	**1.52704627669473**
2.0155644370746373	1.5811388300841898
2.0615528128088303	1.6929339632083815
2.1213203435596424	1.75
2.23606797749979	1.8027756377319946
2.2460077487775676	1.931851652578136
2.3660254037844384	1.9318516525781364
2.3941701709713277	2.0
2.41827959743147	2.0155644370746373
2.5686720715874407	2.0615528128088303
2.6633792282071913	2.1213203435596424
2.9093129111764093	2.23606797749979
	2.2460077487775676
	2.3660254037844384

	2.3941701709713277
	2.41827959743147
	2.5686720715874407
	2.6633792282071913
	2.9093129111764093

Then radius of mini balls ε's have been calculated from the data for both Vietoris-Rips and the approximated simplicial complex which have been compared in Table 2. From Table 2, one can find other ε values dissimilar to the Vietoris-Rips complex in the approximated simplicial complex very easily that have been colored in blue. All ε-values of the Vietoris-Rips complex are contained in the set of ε-values of the proposed approximated simplicial complex.

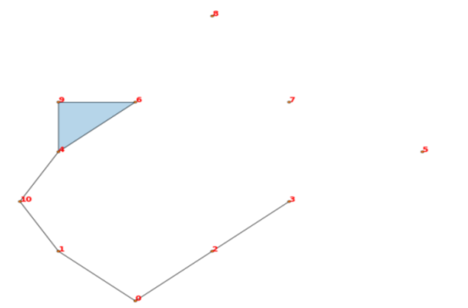
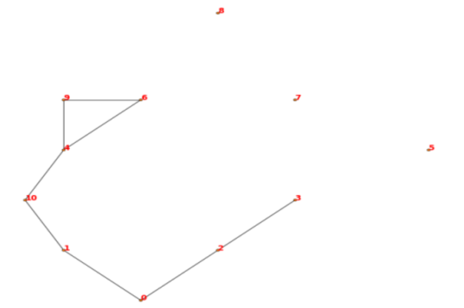
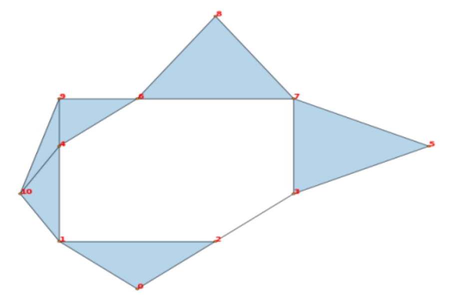
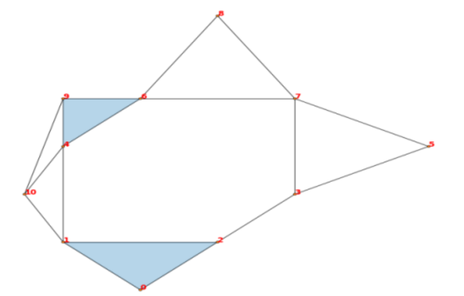
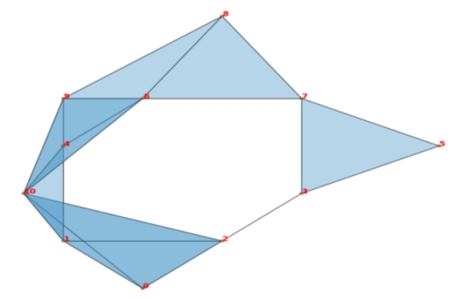
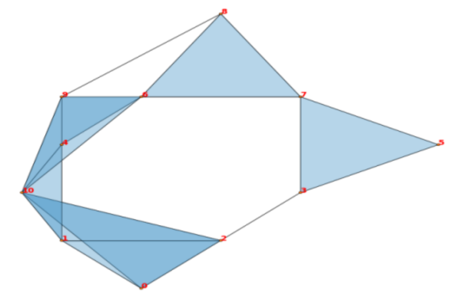
Table 3 Comparison of the approximated simplicial complex with Vietoris-Rips and Čech complexes.

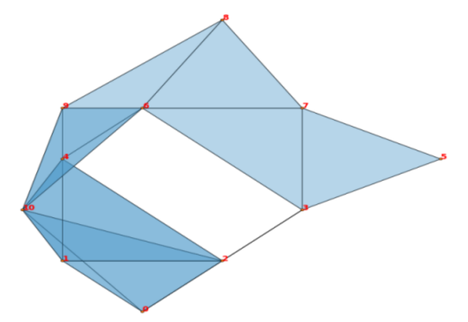
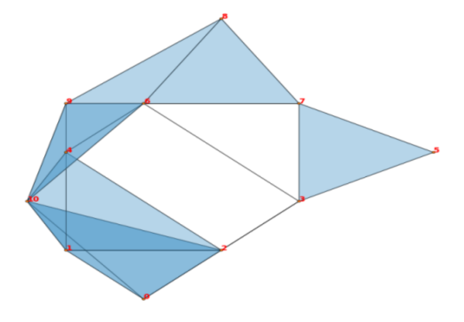
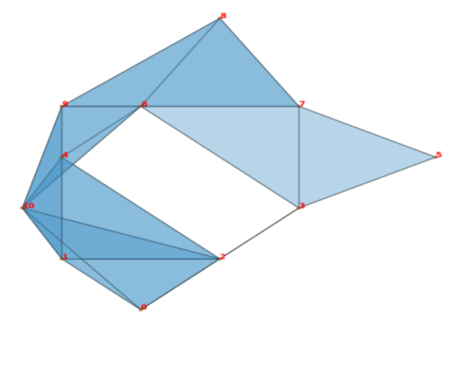
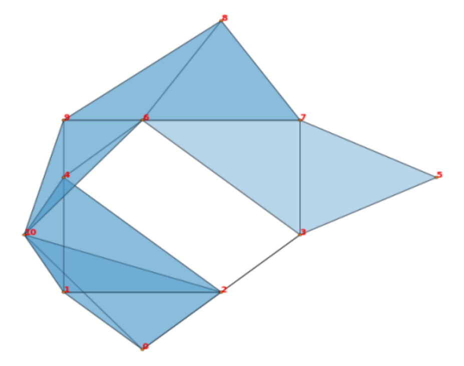
Filtration values	**Name of the Simplicial complex**	**Simplicial complex**
For $\varepsilon = 0.0$	Vietoris-Rips Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10]]
	Approximated Simplicial Complex	**[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10]]**
	Čech Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10]]
For $\varepsilon = 0.5$	Vietoris-Rips Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [4, 9], [6, 9]]
	Approximated Simplicial Complex	**[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [4, 9], [6, 9]]**
	Čech Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [4, 9], [6, 9]]
For $\varepsilon = 0.5590$	Vietoris-Rips Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [1, 10], [4, 9], [4, 10], [6, 9]]
	Approximated Simplicial Complex	**[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [1, 10], [4, 9], [4, 10], [6, 9]]**
	Čech Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [1, 10], [4, 9], [4, 10], [6, 9]]
For $\varepsilon = 0.70711$	Vietoris-Rips Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [0, 1], [0, 2], [1, 10], [2, 3], [4, 6], [4, 9], [4, 10], [6, 9], [4, 6, 9]]
	Approximated Simplicial Complex	**[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [0, 1], [0, 2], [1, 10], [2, 3], [4, 6], [4, 9], [4, 10], [6, 9]]**
	Čech Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [0, 1], [0, 2], [1, 10], [2, 3], [4, 6], [4, 9], [4, 10], [6, 9]]
For $\varepsilon = 0.80474$	Vietoris-Rips Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [0, 1], [0, 2], [1, 10], [2, 3], [4, 6], [4, 9], [4, 10], [6, 9], [4, 6, 9]]
	Approximated Simplicial Complex	**[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [0, 1], [0, 2], [1, 10], [2, 3], [4, 6], [4, 9], [4, 10], [6, 9], [9, 4, 6]]**
	Čech Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [0, 1], [0, 2], [1, 10], [2, 3], [4, 6], [4, 9], [4, 10], [6, 9], [9, 4, 6]]
For $\varepsilon = 0.999999$	Vietoris-Rips Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [0, 1], [0, 2], [1, 10], [2, 3], [3, 5], [4, 6], [4, 9], [4, 10], [5, 7], [6, 9], [4, 6, 9]]
	Approximated Simplicial Complex	**[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [0, 1], [0, 2], [1, 10], [2, 3], [3, 5], [4, 6], [4, 9], [4, 10], [5, 7], [6, 9], [9, 4, 6]]**
	Čech Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [0, 1], [0, 2], [1, 10], [2, 3], [3, 5], [4, 6], [4, 9], [4, 10], [5, 7], [6, 9], [9, 4, 6]]

For ε =1.0	Vietoris-Rips Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [0, 1], [0, 2], [1, 2], [1, 4], [1, 10], [2, 3], [3, 5], [3, 7], [4, 6], [4, 9], [4, 10], [5, 7], [6, 7], [6, 8], [6, 9], [7, 8], [0, 1, 2], [1, 4, 10], [3, 5, 7], [4, 6, 9], [6, 7, 8]]
	Approximated Simplicial Complex	**[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [0, 1], [0, 2], [1, 2], [1, 4], [1, 10], [2, 3], [3, 5], [3, 7], [4, 6], [4, 9], [4, 10], [5, 7], [6, 7], [6, 8], [6, 9], [7, 8], [0, 1, 2], [9, 4, 6]]**
	Čech Complex	[[0], [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [0, 1], [0, 2], [1, 2], [1, 4], [1, 10], [2, 3], [3, 5], [3, 7], [4, 6], [4, 9], [4, 10], [5, 7], [6, 7], [6, 8], [6, 9], [7, 8], [0, 1, 2], [9, 4, 6]]

In Table 3, a comparison of the approximated simplicial complex with Vietoris-Rips and Čech Complexes has been performed. From the comparison, 2-simplices of Vietoris-Rips complexes have been highlighted with blue color which are absent in the approximated simplicial complex. They have been formed for greater filtration values of the other two simplicial complexes. On the contrary, the approximated simplicial complex and Čech complexes are similar for the same filtration values. Thus, the approximated simplicial complex is containing the Vietoris-Rips complex which is similar to Čech Complex.

Table 4 Visualization of Vietoris-Rips complex and approximated simplicial complex showing differences in building 2-simplices for the corresponding filtration values.

Filtration values	Vietoris-Rips Complex	Approximated Simplicial Complex
For ε =0.70711		
For ε =1.03078		
For ε =1.34629		

For ε =1.41421		
For ε =1.5		

Differences have been found in the visualization of simplicial complexes for some selected values of ε in Table 4. In the figure, approximated simplicial complexes have formed cycles before building a 2-simplex for corresponding filtration values which ensures a solution to the data loss problem of the Vietoris-Rips complex.

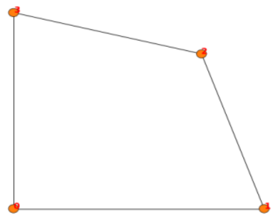
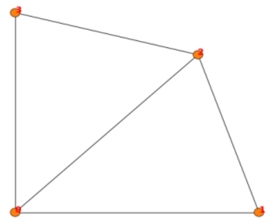
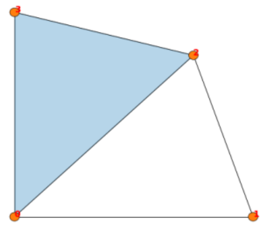
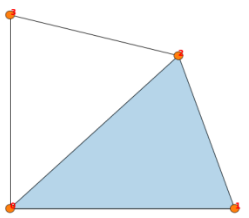
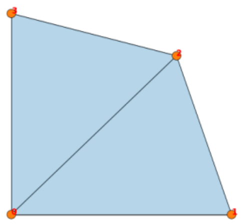
Since the visualization in Table 4 is a 2-dimensional projection, 3-simplexes can't be understood properly. In that case, Table 5 has been calculated to recognize the exact number of holes in the simplicial complexes of the figure. Table 5 shows clear evidence of a greater number of loops ($\boldsymbol{\beta_1}$) in the approximated simplicial complex than in the Vietoris-Rips complex. Since faces are generating faster in the Vietoris-Rips complex, a number of voids ($\boldsymbol{\beta_2}$) will appear earlier than the approximated simplicial complex.

Table 5 Calculated Betti numbers of Vietoris-Rips and approximated simplicial complexes.

Filtration values	**Betti numbers of Vietoris-Rips complex**			**Betti numbers of approximated simplicial complex**		
	$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$
For ε =0.70711	4	0	0	4	1	0
For ε =1.03078	1	1	0	1	5	0
For ε =1.34629	1	1	2	1	2	0
For ε =1.41421	1	1	3	1	3	2
For ε =1.5	1	1	5	1	2	3

Some values may differ from the equivalent filtration values to the Čech complex shown in Table 6 because the filtration values of the proposed approach to generate approximated simplicial complex have been calculated using an approximation to the Čech complex. Another notable distinction is between the simplicial complexes that were formed

Table 6 Filtration values of simplicial complexes building for Čech complex, approximated simplicial complex, and Vietoris-Rips complex.

Simplicial complex	**ε-values for Čech complex**	**ε-values for approximated simplicial complex**	**ε-values for Vietoris-Rips complex**
	0	0	0
	0.5	0.5	0.5
	0.5303	0.5303	-
	0.5489	-	-
	-	0.59651	-
	0.5590	0.61708	0.5303

at 0.5489 for the Čech complex but not for the approximated simplicial complex and those that were formed at 0.5965 for the approximated simplicial complex but not for the Čech complex shown in Table 6.

The suggested approach to find an approximated simplicial complex can be a useful tool to calculate much more quickly and without any data loss in light of all the findings that have been described so far.

5. Conclusion

Vietoris–Rips complex is an approximation to the Čech complex where some persistent data loss has been found. On the other hand, computing Čech complex is computationally expensive. In this study, a simplicial complex between the Vietoris-Rips complex and the Čech complex has been introduced in section 2.3. For this, an algorithm has been written to construct the Vietoris-Rips complex in a new way in section 2.2.1. Then another algorithm for constructing an approximated simplicial complex to Čech complex has been proposed in section 2.3.1. Figure 2 shows the construction of the approximated simplicial complex of the given sample points that have been classified into three classes according to $l < r$, $l = r$, and $l > r$. Then the filtration values of the triangles have been defined by $\varepsilon = r$ for $l \leq r$ and $\varepsilon = r + \frac{(l-r)}{3}$ for $l > r$ in section 2.3. For higher dimensional data similar approach has been followed and the result showed in section 2.3. From Table 3, it can be found that the proposed approximated simplicial complex is containing the Vietoris-Rips complex. Table 4 and Table 6 show that the same amount of filtration values (ε) to the Čech complex that lay between the filtration values of the Vietoris-Rips complex and the Čech complex. Table 5 provides convincing evidence that the approximated simplicial complex has more loops ($\boldsymbol{\beta_1}$) than the Vietoris-Rips complex and number of voids ($\boldsymbol{\beta_2}$) has been appeared earlier than the approximated simplicial complex. As a result, the computationally efficient approximated simplicial complex algorithm, which approximates the Čech complex and solves the data loss problem, has been developed.

The results of this study can be applied in real data especially in network data to get better result than Vietoris-Rips complex with less computational cost. Similar investigations can be performed to construct a more effective approximation to the Čech complex. Also, this study can be conducted over a greater number of points. Persistent homology of the approximated simplicial complex can be calculated to understand persistence in more detail compared with other relevant studies.

Compliance with ethical standards

Acknowledgments

The **Ministry of Science and Technology** (MoST) of Bangladesh provided partial funding for this study through the award of fellowships to Md. Morshed Bin Shiraj and Md. Manik Hossain for the 2022–2023 academic year. Nasima Akhter also received research funding from the Dean of the Faculty of Science at the University of Rajshahi, with grant number **A–117/5/52/RU./Science-01/2022–2023**. For their support in sponsoring this study, the authors thank the MoST and the Dean of the Faculty of Science of the University of Rajshahi.

Disclosure of conflict of interest

None of the authors are involved in any conflicts of interest.

References

[1] Eilenberg S (1944). Singular homology theory. Annals of Mathematics, 407-447.

[2] Carlsson G (2009). Topology and data. Bulletin of the American Mathematical Society, 46(2), 255-308.

[3] Björner A (1995). Topological methods. Handbook of combinatorics, 2, 1819-1872.

[4] Ahmad MZ and Peters J (2017). Proximal Čech Complexes in Approximating Digital Image Object Shapes. Theory and Application. Theory and Applications of Mathematics & Computer Science, 7(2), 81-123.

[5] Choudhary A (2017). Approximation algorithms for Vietoris-Rips and Čech filtrations, DOI: http://dx.doi.org/10.22028/D291-26959.

[6] Choudhary A, Kerber M and Raghvendra S (2017). Improved approximate rips filtrations with shifted integer lattices and cubical complexes. Journal of Applied and Computational Topology, 5, 425 - 458.

[7] Choudhary A, Kerber M and Raghvendra S (2016). Polynomial-Sized Topological Approximations Using the Permutahedron. Discrete & Computational Geometry, 61, 42-80.

[8] Chu J, Vejdemo-Johansson M, and Ji P (2022). An improved algorithm for Generalized Čech complex construction. arXiv preprint arXiv:2209.15574.

[9] Dantchev S, Ivrissimtzis I (2012). Efficient construction of the Čech complex, Computers & Graphics, 36(6), 708-713, ISSN 0097-8493, DOI: https://doi.org/10.1016/j.cag.2012.02.016.

[10] Dey TK, Shi D and Wang Y (2019), SimBa: An Efficient Tool for Approximating Rips-filtration Persistence via Simplicial Batch Collapse. ACM J. Exp. Algorithmics 24, Article 1.5, 16 pages. DOI: https://doi.org/10.1145/3284360.

[11] Kerber M and Sharathkumar R (2013). Approximate Čech complex in low and high dimensions. In Algorithms and Computation: 24th International Symposium, ISAAC 2013, Hong Kong, China, December 16-18, Proceedings 24 (pp. 666-676). Springer Berlin Heidelberg.

[12] Le N-K, Martins P, Decreusefond L and Vergne A (2015). "Construction of the Generalized Czech Complex," 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), Glasgow, UK, pp. 1-5, DOI: https://doi.org/10.1109/VTCSpring.2015.7145759.

[13] Le N-K, Vergne A, Martins P and Decreusefond L (2017), "Optimize wireless networks for energy saving by distributed computation of Čech complex," 2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Rome, Italy, 2017, pp. 1-8, DOI: https://doi.org/10.1109/WiMOB.2017.8115760.

[14] Sheehy D (2013). Linear-Size Approximations to the Vietoris–Rips Filtration. Discrete & Computational Geometry, 49, 778 - 796.

[15] Yuchen W, Jialiang L, and Martins P (2018). "Distributed Coverage Hole Detection Algorithm Based on Čech Complex." In Communications and Networking: 12th International Conference, ChinaCom 2017, Xi'an, China, October 10–12, 2017, Proceedings, Part II 12, pp. 165-175. Springer International Publishing, 2018.

[16] Zomorodian A (2010), Fast construction of the Vietoris-Rips complex, Computers & Graphics, Volume 34, Issue 3, Pages 263-271, ISSN 0097-8493, DOI: https://doi.org/10.1016/j.cag.2010.03.007

[17] Iacopini I, Petri G, Barrat A and Latora V (2019). Simplicial Models of Social Contagion. Nature communications, 10(1), 2485.

[18] Rouvreau V (2015). Čech complex. GUDHI User and Reference Manual, GUDHI Editorial Board, URL: https://gudhi.inria.fr/doc/3.0.0/group__Čech__complex.html