Inferring Underlying Manifold of Low Density Data using Adaptive Interpolation

Noritaka Yamada[1][a] and Takeshi Shibuya[2][b]
[1]*Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Japan*
[2]*Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Japan*

Keywords: Topological Data Analysis, Persistent Homology, Underlying Manifold, Topological Feature, Interpolation.

Abstract: Inferring the topological shape of an underlying manifold of data is efficient for point cloud data analysis. This is accomplished by estimating the Betti numbers of the underlying manifold in each dimension from point cloud data. Futagami et al. proposed a method to automatically estimate the Betti numbers of the underlying manifold using persistent homology. However, this method estimates 2nd the Betti numbers of the underlying manifold less accurately as data density decreases. The low accuracy of estimating 2nd the Betti numbers is caused by the difficulty of detecting 2-dimensional holes. In this study, we propose a method to estimate 2nd the Betti number of the underlying manifold of low density data accurately. Concretely, we increase the density of data using interpolation that adds temporary points close to the underlying manifold. Then, we calculate persistent homology of data whose density has been increased and estimate 2nd Betti numbers from the calculation results. We confirm that our proposed method is effective to estimate 2nd the Betti numbers of the underlying manifold.

1 INTRODUCTION

According to Bishop (2006), many data sets have the property that the data points lie close to a manifold. We call the manifold that the data points lie on the "underlying manifold."

Inferring the topological shape of the underlying manifold of a data set is efficient for point cloud data analysis. For example, ensuring that the topology of a graph for a self-organizing map (SOM) is the same as that of the underlying manifold of the data set is critical because it enable the SOM to preserve topological relationships among data points (Futagami and Shibuya, 2016).

Inferring the underlying manifold of a data set is accomplished by estimating the number and dimension of the "holes" in the underlying manifold and then defining its topological shape based on the same number and dimensions of "holes." A "hole" is a topological feature such as the loop in a donut or the void in an empty sphere. A loop is a 1-dimensional hole, and an enclosed solid void is a 2-dimensional hole. The number of holes in a given shape is known as the "Betti number." For example, if the underlying manifold has one 1-dimensional hole and two 2-dimensional holes, the topological shape of the underlying manifold is the same as that of a torus. Table 1 shows a few examples of topological shapes and the number of holes in each shape.

Calculating persistent homology derives the size, number and dimension of holes that composed of data points (Zomorodian and Carlsson, 2005; Edelsbrunner and Harer, 2008). Some holes derived by calculating persistent homology correspond those in the underlying manifold of a data set. On the other hand, other holes derived by calculating persistent homology are simply topological noises produced by gaps among points on the surface of the underlying manifold. We call the former "cycle" and the latter "noise." The number of n-dimensional cycles is equivalent the n-th Betti number of the underlying manifold. Estimating the number of cycles from the calculation result of persistent homology of a data set gives an estimate of the Betti numbers of the underlying manifold of that data set.

Futagami et al. (2019) proposed a method to estimate the number of cycles from the calculation result of persistent homology of a data set. In this paper, we call this method the "conventional method." The Betti

[a] https://orcid.org/0000-0003-2862-490X
[b] https://orcid.org/0000-0003-4645-5898

Yamada, N. and Shibuya, T.
Inferring Underlying Manifold of Low Density Data using Adaptive Interpolation.
DOI: 10.5220/0008915803950402
In *Proceedings of the 12th International Conference on Agents and Artificial Intelligence (ICAART 2020) - Volume 2*, pages 395-402
ISBN: 978-989-758-395-7; ISSN: 2184-433X

numbers of the underlying manifold are estimated automatically using this method when data density is high enough, that is the number of point in an N-dimensional unit space composing an N-dimensional underlying manifold is high. However, when data density is low, using the conventional method often yields an incorrect 2nd Betti number. Detecting cycles is difficult when data points are few.

We cannot always obtain high density data in practice. Accurately estimating the 2nd Betti number of the underlying manifold is necessary even if data density is low. In this study, we propose a method to estimate the 2nd Betti number of the underlying manifold accurately even when data density is for a range in which the conventional method often estimates the 2nd Betti number incorrectly.

This study targets the range of data density for which the estimation success rate ranges from approximately 30% to 60% when using the conventional method. Data points are too few to infer the underlying manifold when data density is lower than it is in this range.

2 PERSISTENT HOMOLOGY

Persistent homology is one of the tools of topological data analysis. Calculating persistent homology determines the size, number and dimension of holes in a point cloud data set (Zomorodian and Carlsson, 2005; Edelsbrunner and Harer, 2008). A hole is an area surrounded by data points but itself containing no data points.

Let us suppose that $(n+1)$-dimensional balls with radius r centering on each data point in a data set have been drawn and r increases, as shown in Figure 3. As r increases, the $(n+1)$-dimensional balls touch and cross each other and n-dimensional holes surrounded by $(n+1)$-dimensional balls are born. Let $r = b$ when a hole is born. The time when a hole is born is known as “birth time.” As r increase more, n-dimensional holes disappears. Let put $r = d$ when a hole disappear. The time when a hole disappear is known as “death time.” Figure 3 shows an example of a 1-dimensional hole birth and death. The 1-dimensional hole surrounded by 2-dimensional balls (disks) is born at $r = b$, and the hole disappears at $r = d$ when filled with disks.

A persistence diagram is a graph that maps (b,d) as a coordinate (Cohen-Steiner et al., 2007). Figure 2 is the persistence diagram representing the calculation result of persistent homology of the torus shape data shown in Figure 1. The difference between birth and death times $(d-b)$ is called “persistence.” Persistence represents the size of holes. In the graph, distances between each point and the diagonal represents persistence. The larger the persistence is, the larger the hole is. Holes with large persistence are usually considered to be cycles.

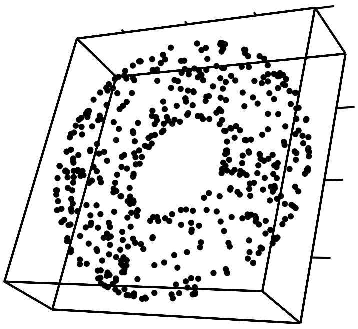

Figure 1: Torus shape data.

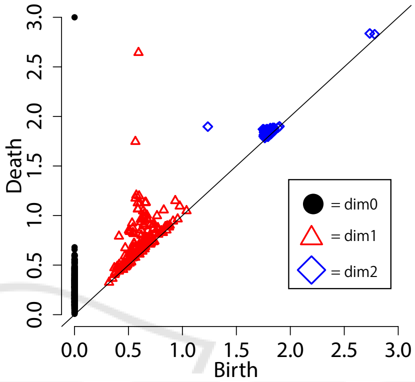

Figure 2: Persistence diagram of torus shape data.

The red triangles in Figure 2 indicate 1-dimensional holes and the blue squares in Figure 2 indicate 2-dimensional holes. The torus that is the underlying manifold of Figure 1 has two 1-dimensional holes and one 2-dimensional hole. However, many more holes appear in Figure 2 than there actually are in the underlying manifold. Noises produced from gaps among points on the underlying manifold cause this problem. The method to estimate the number of cycles in the calculation result of persistent homology of the data set is required in order to estimate the Betti numbers of the underlying manifold.

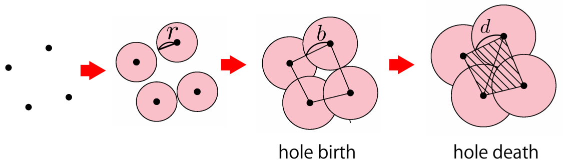

Figure 3: Radius increasing, hole birth and hole death.

Table 1: The number of holes in topological shapes.

Topological shape	1-dimensional holes	2-dimensional holes
Circle	1	0
Sphere ($\mathbb{S}^2$)	0	1
Torus ($\mathbb{S}^1 \times \mathbb{S}^1$)	2	1

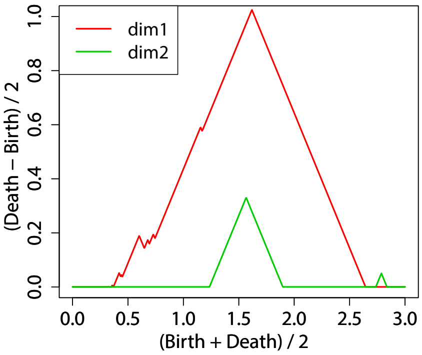

Figure 4: Persistence landscape of torus shape data.

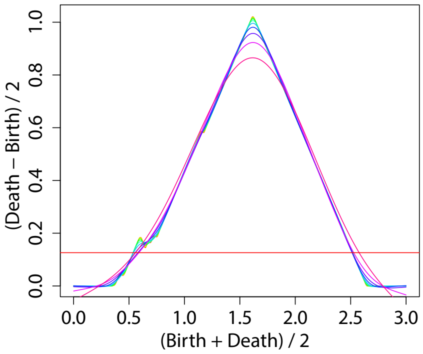

Figure 5: Smoothed persistence landscape of torus shape data.

2.1 Estimating the Betti Numbers by the Use of Persistent Homology Analysis

When a data density is high enough, the Betti numbers of an underlying manifold is estimated by using the method proposed by Futagami et al. (2019). In this section, we describe this conventional method briefly.

First, to reduce noises, a persistence landscape is calculated from a persistence diagram. A persistence landscape is given by mapping each coordinate $p = (b,d)$ in a persistence diagram D to a piecewise linear function (Bubenik, 2015), such that

$$\Lambda_p(t) = \begin{cases} t-b & t \in [b, \frac{b+d}{2}] \\ d-t & t \in [\frac{b+d}{2}, d] \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In general, the collection of the largest functions in Eq.1 is used, such that,

$$\lambda(t) = \max_{p \in D} \Lambda_p(t). \tag{2}$$

Figure 4 is the persistence landscape derived from Figure 2.

Second, the persistence landscape is smoothed to determine whether small local maxima on the side of large local maxima is cycles or noises. Local maxima that exist even after high smoothing are considered to be cycles. Concretely, a persistence landscape is smoothed by repeatedly fitting cubic smoothing splines based on B-splines while using various smoothing parameters. Figure 5 shows the 1-degree persistence landscapes in Figure 4 smoothed with various parameters. A red horizontal line in Figure 5 represents the mean of persistence used as a threshold to discriminate between cycles and noises. The larger smoothing parameter is, the smoother a persistence landscape is.

Third, counting local maxima above the threshold in each smoothed persistence landscape. The mean of the number of local maxima above the threshold in each smoothed persistence landscape is considered to be the number of cycles in a given data set.

The processes mentioned above are applied to some subsamples of a given data set. The mean of the number of cycles in the subsamples is then rounded off and is considered to be the estimated Betti numbers.

2.2 Limit of the Conventional Method

The estimation accuracy of the 2nd Betti numbers using the method proposed by Futagami et al. (2019) gets worse as data density decreases. Figure 6 shows the relationship between data density and the accuracy of estimating 2nd Betti numbers. Success rate represents the rates of data sets whose 2nd Betti numbers of their underlying manifold are estimated correctly among all data sets when using the conventional method. Data density is the number of points per unit area of the surface of the underlying manifold. The vertical and horizontal axes in Figure 6 represent the success rate and data density, respectively. We estimated the 2nd Betti number of data sampled from the uniform distribution on the torus, as shown

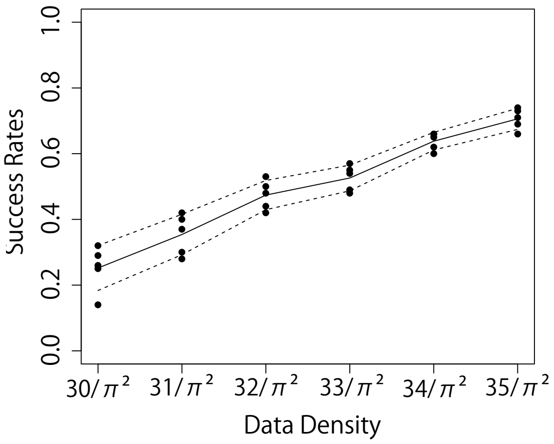

Figure 6: Relationship between data density and success rates of estimating 2nd Betti number.

in Figure 1, using the conventional method. The major radius and the minor radius of the torus were 2.5 and 1, respectively. Then, we calculated the rates of data sets of which the conventional method estimated 2nd Betti numbers of their underlying manifold correctly among 100 data sets at once trial. While changing the number of data points in one data set from 300 to 350 by 10, we conducted this trial five times in each data density. Each black circle in Figure 6 indicates a success rate for each trial. For example, when data density is $31/\pi^2$, that is 310 data points lie on the torus, the conventional method estimated correctly for 32% of 100 data sets in first trial, and for 35% of 100 data sets in second trial, and for 38% of 100 data sets in third trial, and for 40% of 100 data sets in fourth trial, and for 39% of 100 data sets in fifth trial. A black line and black dashed lines are straight lines connecting the mean of success rates and the sum or difference between the mean and standard deviation of success rates in each data density, respectively.

Figure 6 shows that estimation accuracy gets worse as data density decreases. When data density is low, data points is distributed sparsely. Therefore, being crossed 3-dimensional balls each other take more time, that is, birth times of 2-dimensional holes get later, when data density is low. On the other hand, death times of 2-dimensional holes are about the same when data density is high. When birth times is delayed, the persistence of 2-dimensional holes is smaller and detecting 2-dimensional cycles becomes more difficult.

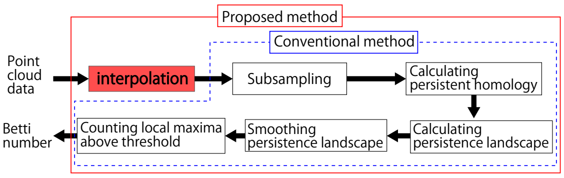
Proposed method
Conventional method
Point cloud data
interpolation
Subsampling
Calculating persistent homology
Betti number
Counting local maxima above threshold
Smoothing persistence landscape
Calculating persistence landscape

Figure 7: Sequence of proposed method.

3 PROPOSED METHOD

3.1 Method to Estimate the Betti Number using Interpolation

The estimation accuracy of 2nd Betti numbers using the conventional method gets worse as data density decreases. Late birth times of 2-dimensional holes make the persistence of 2-dimensional cycles smaller and discriminating between cycles and noises more difficult. Difficulty of detecting cycles causes the poor estimation accuracy of 2nd Betti numbers. To solve this problem, we propose a method to estimate 2nd Betti numbers using interpolation that add points in sparse areas on an underlying manifold. Using interpolation to make the detection of cycles easier improves the estimation accuracy of 2nd Betti numbers.

We add points close to a tangent space and a point of tangency in the underlying manifold in our proposed method. According to Zomorodian (2005), intuitively, a manifold is a topological space that locally looks like $\mathbb{R}^n$. We approximate a tangent space using this property that N-dimensional manifold is locally similar to N-dimensional Euclidean space. Points in a tiny range on an N-dimensional underlying manifold are considered to be in an N-dimensional space. Additionally, this space is considered to be approximate to a tangent space. Therefore, we approximate tangent spaces using points in a tiny range on an underlying manifold. Adding points in approximated tangent spaces increases data density while retaining the topological feature of the underlying manifold. In our proposed method, we use principal component analysis (PCA) to approximate tangent spaces.

However, if too many points are added, the computational time of persistent homology will be too long to estimate the Betti numbers of an underlying manifold in most practical applications. Therefore, we add points only in comparatively sparse areas on an underlying manifold as much as possible.

The proposed method employs the conventional method to analyze data set whose density is increased. Figure 7 shows the sequence of the proposed method.

We describe an interpolation method employed in the proposed method in Sec.3.2.

Algorithm 1: Interpolating near on Underlying Manifold.

```
1: Inputs:
       $\mathcal{X} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\} \subset \mathbb{R}^D$ : a input data
       $\mathcal{M} = \{1, \cdots, m\}$ : indices of $\mathcal{X}$
       $N$ : the dimension of underlying manifold of $\mathcal{X}$
       $K$ : the number of neighbors
2: Outputs:
       $\hat{\mathcal{X}}$ : a interpolated point set
3: $I \leftarrow \emptyset$
4: $\hat{\mathcal{X}} \leftarrow \emptyset$
5: for $l \leftarrow 1$ to $m$ do
6:     if $l \notin I$ then
7:         $I \leftarrow I \cup \{l\}$
8:         $\{\mu_1, \cdots, \mu_K\}_{\mu_i \in \mathcal{M}} \leftarrow$ the indices of the $K$ nearest neighbors of $\boldsymbol{x_l} \in \mathcal{X}$
9:         $I \leftarrow I \cup \{\mu_1, \cdots, \mu_K\}$
10:        $\{\boldsymbol{x}'_0, \boldsymbol{x}'_1, \cdots, \boldsymbol{x}'_K\} \leftarrow$ project $\{\boldsymbol{x}_l, \boldsymbol{x}_{\mu_1}, \cdots, \boldsymbol{x}_{\mu_K}\}$ into $N$-dimensional space using PCA
11:        $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p\} \leftarrow$ the vertexes of the Voronoi region of $\boldsymbol{x}'_0$
12:        $\{\hat{\boldsymbol{x}}_1, \cdots, \hat{\boldsymbol{x}}_p\} \leftarrow \{\hat{\boldsymbol{x}}_i = \boldsymbol{U}\boldsymbol{v}_i + \boldsymbol{x}_l\}_{i=1,\cdots,p}$ ($\boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_N]$ are first to $N$-th principal vectors derived in line 10)
13:        $\hat{\mathcal{X}} \leftarrow \hat{\mathcal{X}} \cup \{\hat{\boldsymbol{x}}_1, \cdots, \hat{\boldsymbol{x}}_p\}$
14:    end if
15: end for
```

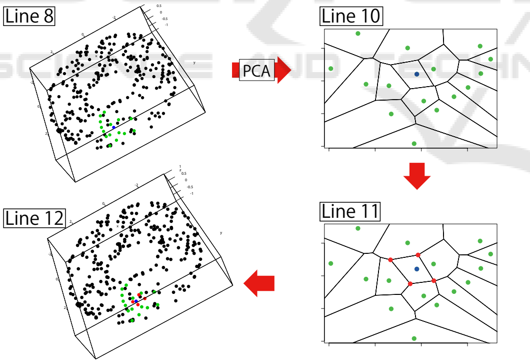

Figure 8: Interpolation that add points close to underlying manifold. "Line X" framed by rectangle indicate line in Algorithm 1. **Top Left:** Select one point and K nearest neighbor points. **Top Right:** Project those points onto an approximated tangent space using PCA. **Bottom Right:** Add points at the vertexes of a Voronoi region. **Bottom Left:** Map added points using PCA reconstruction into $\mathbb{R}^D$, a space which a given data set belongs to.

3.2 Adding Points into Sparse Areas on the Underlying Manifold

We describe the interpolation method employed in the proposed method. Algorithm 1 shows the algorithm of the interpolation method.

Let $\mathcal{X}$ be a point cloud data set of which we want to estimate the Betti numbers of its underlying manifold. The dimension N of the underlying manifold of $\mathcal{X}$ is determined using the method to estimate the intrinsic dimension of $\mathcal{X}$ proposed by Hein et al. (Hein and Audibert, 2005).

Figure 8 show processes in one loop from line 5 to line 15 in Algorithm 1. "Line X" framed by rectangle indicate the line in Algorithm 1. We detail processes at line 8, 10, 11 and 12 in Algorithm 1 as follows.

Line 8. Select one point $\boldsymbol{x}_l$ and K nearest neighbor points $\{\boldsymbol{x}_{\mu_1}, \cdots, \boldsymbol{x}_{\mu_K}\}$ of $\boldsymbol{x}_l$. Blue and green points shown in Figure 8 indicate $\boldsymbol{x}_l$ and $\{\boldsymbol{x}_{\mu_1}, \cdots, \boldsymbol{x}_{\mu_K}\}$, respectively. Those points are considered to lie close to a tangent space.

Line 10. Project those points onto an approximated tangent space using PCA. Let $\mathcal{X}' = \{\boldsymbol{x}'_0, \boldsymbol{x}'_1, \cdots, \boldsymbol{x}'_k\}$ be the projected $\{\boldsymbol{x}_l, \boldsymbol{x}_{\mu_1}, \cdots, \boldsymbol{x}_{\mu_K}\}$. We put a region for each points of $\mathcal{X}'$, such that

$$V_i = \{\boldsymbol{x}' \in \mathbb{R}^N \mid \|\boldsymbol{x}' - \boldsymbol{x}'_i\| \le \|\boldsymbol{x}' - \boldsymbol{x}'_j\|,\ 0 \le j \le k, j \neq i\} \quad (3)$$

Those regions V_i assigned for each $\boldsymbol{x}'_i$ are known as "Voronoi regions," and a partitioning of a space into Voronoi regions is known as a “Voronoi partition.”

Line 11. Add then points at the vertexes $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p\}$ of a Voronoi region having $\boldsymbol{x}'_0$.

Line 12. Lastly, map added points $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p\}$ using PCA reconstruction with $\boldsymbol{x}_l$ and principal vectors $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_N\}$ into $\mathbb{R}^D$, a space which a given data set belongs to.

In line 12, we use the point $\boldsymbol{x}_l$ selected in line 8 in stead of the mean $\bar{\boldsymbol{x}}$ of $\boldsymbol{x}_l$ and K nearest neighbor points $\{\boldsymbol{x}_{\mu_1}, \cdots, \boldsymbol{x}_{\mu_K}\}$ for PCA reconstruction. With $\boldsymbol{x}_l$ and principal vectors $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_N\}$, added points $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p\}$ are reconstructed into the space that is spanned by $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_N\}$ and has $\boldsymbol{x}_l$. Reconstruction of $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p\}$ with $\boldsymbol{x}_l$ adds points closer to the tangent space that contact with the underlying manifold at $\boldsymbol{x}_l$ than with $\bar{\boldsymbol{x}}$.

Red points in Figure 8 shows added points $\{\hat{\boldsymbol{x}}_1, \cdots, \hat{\boldsymbol{x}}_p\}$ using this method. The proposed method repeats those processes while ensuring that points already selected as neighbors are not selected again.

Putting points at the vertexes of a Voronoi region achieves the requisite interpolation that adds points close to comparatively sparse areas on the underlying manifold. Additionally, once a point has been selected as a neighbor, we do not select it as a candidate to be the center of neighborhood in order to reduce the total number of points and any subsequent computational complexity.

4 EXPERIMENT

We estimated the 2nd Betti number of an underlying manifold of data using the proposed and conventional methods to confirm that the proposed method estimates more accurately than the conventional method. Table 2 shows the spec of computer used in the experiment.

Table 2: The spec of computer used in the experiment.

OS	Windows Server 2019 Standard 64bit ver.1809
CPU	Intel(R) Xeon(R) Gold 5122 CPU @ 3.60GHz
RAM	256GB

4.1 Experiment Settings

We used a point cloud sampled from the uniform distribution on a torus as one data set in an experiment. The major radius and the minor radius of the torus were 2.5 and 1, respectively. We estimated the 2nd Betti number of the underlying manifold, which is the torus, using the proposed method. We calculated the rates of data sets of which the proposed method estimated 2nd Betti numbers of their underlying manifold correctly among 100 data sets at once trial. While changing the number of data points in one data set from 300 to 350 by 10, we conducted this trial five times in each data density. Then, we compared the success rates of estimates given by the proposed and conventional methods.

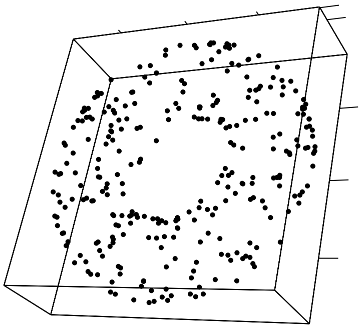

Figure 9: Torus shape data before interpolating.

4.2 Experiment Results

Figure 9 shows one example of the torus shape data used in the experiment. Additionally, Figure 10 shows torus shape data after applying the interpolation method to the data shown in Figure 9. Red points in Figure 10 are interpolated points. Comparing Figure 9 with Figure 10, we find that points are added close to sparse areas on the torus as intended.

Figure 11 shows the success rates of estimating 2nd Betti numbers using the proposed and conventional methods. Each red and black circle in Figure 11 indicates a success rates of estimating using the proposed and conventional methods for each trial, respectively. Red and black lines are straight lines con-

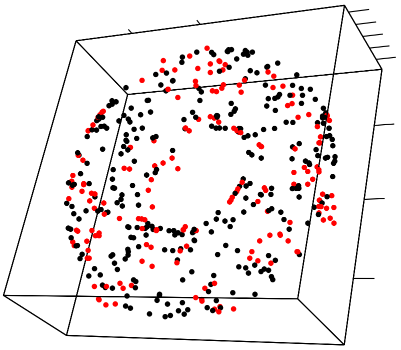

Figure 10: Torus shape data after interpolating.

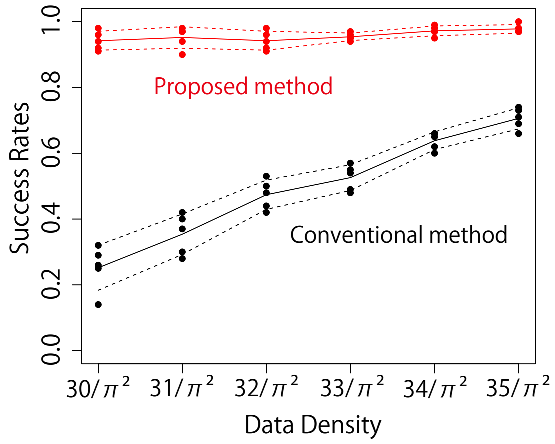

Figure 11: Success rates of estimating 2nd Betti number.

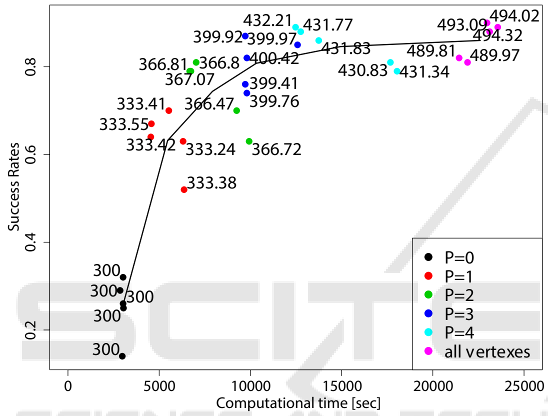

Figure 12: Relationship between computational time and success rate.

necting the mean of success rates by the proposed and conventional methods, respectively. Red and black dashed lines are also straight lines connecting the sum or difference between the mean and standard deviation of success rates by the proposed and conventional methods in each data density, respectively. Figure 11 shows that the proposed method estimated much more accurately than the conventional method.

We confirmed that the proposed method estimated 2nd Betti number much more accurately than the conventional method based on the experimental result. The proposed method is effective to estimate 2nd Betti numbers of the underlying manifold of low density data.

4.3 Trade-off between Accuracy and Computational Time

We confirm relationship between the accuracy of estimating 2nd Betti numbers and computational times. We examined estimation accuracy and computational times when changing the number of points added to toruses that were used in the experiment and had 300 points. In order to change the number of added points, we added points at P vertexes of Voronoi region having x_l in descending order of distance between x_l and vertexes in the interpolation method employed in the proposed method. We used 1, 2, 3 and 4 for P. We calculated the rates of data sets of which the proposed method estimated 2nd Betti numbers of their underlying manifold correctly among 100 data sets at once trial. While changing P, we conducted this trial five times in each P.

Figure 12 shows relationship between the accuracy of estimating 2nd Betti numbers and computational time when changing the number of added points. The vertical and horizontal axes in Figure 12 represent the rates of data sets estimated correctly among 100 data sets and computational times to estimate 2nd Betti numbers of 100 data sets, respectively. Red, green, blue and light blue circles in Figure 12 indicate the results when P is 1, 2, 3 and 4, respectively. Black circles indicate the results when $P = 0$, that is when using the conventional method. Purple circles indicate the results when added points at all vertexes of Voronoi region. The numbers beside circles indicate the mean of the numbers of points of data sets in each trial. Black straight line conects the mean of the results in each P.

Figure 12 shows that the computational times increase as added points increase. The proposed method estimated 2nd Betti numbers more accurately than the conventional method. On the other hand, the proposed method costed more computational time to estimate than the conventional method. The trade-off between accuracy and computational times as shown in Figure 12 occurs when using the proposed method.

5 CONCLUSION

In this study, we propose the method to estimate the 2nd Betti numbers of the underlying manifold using the interpolation method that adds points close to comparatively sparse areas in the underlying manifold. Then, we confirm that the proposed method estimated 2nd Betti number of the underlying manifold more accurately than the conventional method. Consequently, we confirm that our proposed method is effective for estimating 2nd Betti numbers of the underlying manifold even when data density is in the range in which the conventional method often estimates 2nd Betti number incorrectly. Adding points close to the underlying manifold enables the proposed method to estimate 2nd Betti number of the underlying manifold with greater accuracy.

We would like to confirm the effectiveness of the proposed method for other manifolds in future research. We also would like to create a method to interpolate with fewer errors.

REFERENCES

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.

Bubenik, P. (2015). Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16:77–102.

Cohen-Steiner, D., Edelsbrunner, H., and Harer, J. (2007). Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120.

Edelsbrunner, H. and Harer, J. (2008). Persistent homology - a survey. *Surveys on Discrete and Computational Geometry*, 453:257–282.

Futagami, R. and Shibuya, T. (2016). A method deciding topological relationship for self-organizing maps by persistent homology analysis. In *2016 55th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, pages 1064–1069.

Futagami, R., Yamada, N., and Shibuya, T. (2019). Inferring underlying manifold of data by the use of persistent homology analysis. In Marfil, R., Calderón, M., Díaz del Río, F., Real, P., and Bandera, A., editors, *Computational Topology in Image Context*, pages 40–53, Cham. Springer International Publishing.

Hein, M. and Audibert, J.-Y. (2005). Intrinsic dimensionality estimation of submanifolds in rd. In *Proceedings of the 22nd International Conference on Machine Learning*, ICML '05, pages 289–296, New York, NY, USA. ACM.

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

Zomorodian, A. J. (2005). *Topology for Computing*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press.