Applying topological data analysis to local search problems

Erik Carlsson[*], John Gunnar Carlsson[†], and Shannon Sweitzer[‡]

June 18, 2021

Abstract

We present an application of topological data analysis (TDA) to discrete optimization problems, which we use to improve the performance of the 2-opt local search method for the traveling salesman problem. We then construct a filtered simplicial complex associated to every a stochastic matrix with a steady state vector (P, π), which is tailored to optimization problems through Markov chain Monte-Carlo algorithms. When P is induced from a random walk on a finite metric space, this complex exhibits similarities with standard constructions such as Vietoris-Rips, but is not sensitive to outliers, as sparsity is a natural feature of the construction. We interpret the persistent homology groups in several examples coming from random walks and discrete optimization.

[*]Department of Mathematics, UC Davis
[†]Department of Industrial and Systems Engineering, University of Southern California
[‡]Department of Industrial and Systems Engineering, University of Southern California

1 Introduction

The concepts of *local search* and *local optimality* are foundational to the field of optimization. For continuous problems, the most famous examples are the Lagrange multiplier theorem and the KKT conditions, which establish necessary conditions for local optimality for constrained problems. Unconstrained problems are almost universally solved using *descent methods* in which one constructs a sequence of iterates whose objective values are consistently improving, and which converge to a local minimizer provided that appropriate regularity conditions are satisfied.

Local optimality is equally ubiquitous in the design of heuristics for discrete optimization problems. The most common manifestation is in the notion of *k-optimality*, which is commonly applied to (for example) the travelling salesman problem [5] and the quadratic assignment problem [8]: a candidate solution is said to be k-optimal if it is at least as good as all other solutions that differ by at most k terms. Another prominent example is *Lloyd's location-allocation* algorithm, which finds evenly spaced points in subsets of Euclidean spaces, and forms the basis for the so-called *naïve k-means* algorithm in machine learning [6].

Not surprisingly, a major drawback of local search methods is the possibility of being stuck at a locally optimal solution that is of poor overall quality. A second well-known drawback of local search schemes arises in multi-criteria decision making: in practice, when one has multiple (possibly competing) objectives that must be simultaneously reconciled, it may be useful to study the "landscape" consisting of all plausible solutions to the problem (perhaps in an *ad hoc* fashion), such as the *Pareto frontier*.

In this paper we develop an approach to these problems using topological methods, specifically topological data analysis (TDA) and persistent homology.

This happens in two parts. First, in Section 2, we apply TDA to a data set whose points are local minimizers of the travelling salesman problem (TSP). We devise a scheme for incorporating the persistent zero-th Betti number (i.e. connected components) of the Vietoris-Rips construction into a local search procedure, and show that it leads to improved performance. We then outline a general basic pipeline for applying persistent homology to (discrete) optimization problems.

In Section 3, we present a new filtered simplicial complex motivated by statistical mechanics. This construction associates to every pair of a stochastic matrix (i.e. a Markov chain) and a steady state row vector (P, π), a filtered simplicial complex (X, f), which is particularly tailored for use in Monte Carlo methods in optimization, such as simulated annealing. Unlike the previous construction, which regards the simulated outputs as a data set, this new construction takes advantage of available structure that the data set comes from simulating a random process. One feature of this complex is that unlike Vietoris-Rips, zero simplices have nontrivial persistence values, which measure a form of density.

We then illustrate Construction 1 in several specific examples. Our first example shows that the first persistent homology groups of a simple Markov chain corresponding to a random walk in a "Jeu-de-Taquin" game have a natural interpretation as in terms of the configuration space of three ordered points in the plane, showing that (X, f) is a viable answer to the "landscape" issue described earlier. Next, we present an application of (X, f) in the case when (P, π) represents a random walk on a finite metric space. We find several advantages even in this scenario over standard constructions such as Vietoris-Rips, for instance insensitivity to outliers, which is typically handled by directly removing all points with low density values. Finally, we give an application which shows how higher homology groups may be applied to study connected components in high-dimensional optimization problems.

1.1 Related work

To the best of our knowledge, the fields of optimization and TDA have had very little cross-pollination. The best example to our knowledge is [9], which notes that certain procedures in TDA have natural interpretations as optimization problems:

> Finding good representatives for qualitative features often turns out to be a case of searching within such a class for an optimal member.... Another area where homology shows up as a tool for optimization is inevaluating coverage for sensor agents – such as ensembles of robots, or antenna configurations.

The paper [4] does not apply TDA per se, but uses sheave theory and Poincaré duality to generalize the famous *max flow-min cut* theorem; their generalization lends itself to applications in multi-commodity flows, among others. Very recently, the paper [3] gives an application of persistent homology to the *image correspondence* problem, which is used to produce 3D reconstructions of scenery from two or more cameras. They produce a complex whose nontrivial homology groups correspond to recognizable anomalies in image pairs, such as repeated patterns, which contribute to nonconvexity of the objective function.

2 Local search with TDA

This section describes a simple metaheuristic that uses persistence barcodes to identify features that we deem likely to be present in good local minimizers. In order to compare our approach with related work, we briefly describe other metaheuristics that arise commonly in optimization literature:

Guided local search augments the objective function with a penalty term that causes the local search procedure to avoid local minimizers that it has

already encountered; the idea is to make previous local minimizers more costly than the surrounding search space. Given an objective function $f(x)$ and a collection of *features* indexed by a term i, guided local search seeks to minimize the augmented objective function

$$g(x) := f(x) + \lambda \sum_i I_i(x) p_i \,;$$

here λ is an intensity parameter, $I_i(x)$ is a binary function that indicates whether or not feature i is present in candidate solution x, and p_i is a penalty parameter for that feature. A simple example of a feature would be "$I_i(x) = 1$ if $x_i = 1$", for instance. The penalty parameters p_i are initially all set to zero, and increase as the metaheuristic finds more and more local minimizers.

Tabu search avoids previous local minimizers by maintaining a *tabu list* of local search moves that are prohibited. In general, a local search move is prohibited if it takes the search back to a state that it has recently visited.

Simulated annealing decides probabilistically whether to move from a state x to a new state x', even if x' has a worse objective value. For example, if $f(x') > f(x)$, one might move from x to x' with probability $ae^{b(f(x)-f(x'))}$, for suitable constants a and b. This moves the search away from poor local minimizers by allowing the possibilities of climbing out of a basin of attraction.

Our proposed *local search with TDA* is similar to the above approaches; its main distinction is that, rather than forbidding certain moves, the goal is to identify features that are present in many disparate local minimizers, and to encourage their presence in subsequent runs.

2.1 A pipeline for local search with TDA

The following is an informal outline for using TDA to guide a local search procedure, for fixed positive integers M and N and a given optimization problem:

1. Perform M local searches with randomly seeded inputs and record the local minimizers $\hat{x}^{(1)}, \ldots, \hat{x}^{(M)}$ that are obtained.
2. Compute persistence barcodes for data points $\hat{x}^{(1)}, \ldots, \hat{x}^{(M)}$, using (for instance) a Čech complex or a Vietoris-Rips complex.
3. Determine a representative clustering by looking at the barcodes corresponding to β_0, the 0-th Betti number. Let $C_1, \ldots, C_m$ denote these clusters.
4. Identify features that are common to the clusters $C_1, \ldots, C_m$.
5. Perform N local searches with randomly seeded inputs, but require that the common features from step 4 be present.
6. Return the best solution obtained over all $M + N$ iterations.

This description is purposefully vague and we show a concrete example applied to the travelling salesman problem in the next section.

2.2 The traveling salesman problem

Here we describe an implementation and show an example of the algorithm from Section 2 that applies TDA to the well-known *2-opt* heuristic [5] for the travelling salesman problem (TSP) with a symmetric and complete distance matrix $D = [d_{ij}]$. 2-opt is a simple strategy that works as follows: given an existing tour of the input points, take 2 edges (a, b) and (c, d) from the tour and remove them. Then, form a new tour by inserting edges (a, d) and (c, b). Update the tour if the objective value decreases, which occurs if and only if

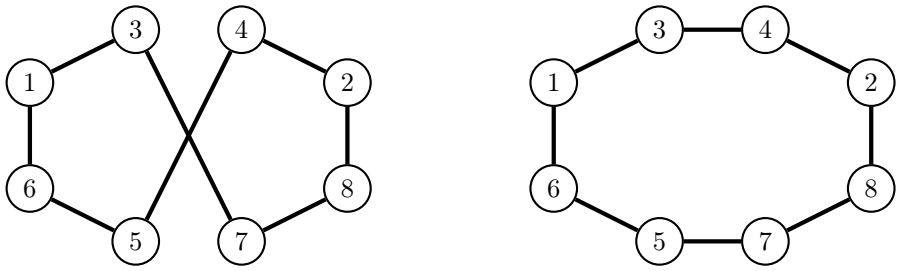


Figure 1: A 2-opt exchange on a tour of 8 points. The edges $(3, 7)$ and $(4, 5)$ in (1a) are replaced with $(3, 4)$ and $(7, 5)$ in (1b).

$d_{ab} + d_{cd} > d_{ad} + d_{cb}$. This is illustrated in Figure 1. The search completes when there are no pairs of edges that result in a cost reduction.

In our experiment, steps 1-6 from Section 2 now take the following form:

1. Perform M local search instances with randomly seeded inputs and record the local minimizing tours $\hat{x}^{(1)}, \ldots, \hat{x}^{(M)}$ that are obtained. Each local search instance consists of taking a random permutation of the input points, and selecting the 2-opt move that reduces the tour cost as much as possible (this means that we enumerate through all $\binom{n}{2}$ possible pairs of edges, where n is the number of points). For our example, we collected a set of 1000 local minimizing tours on a TSP instance of 100 points.

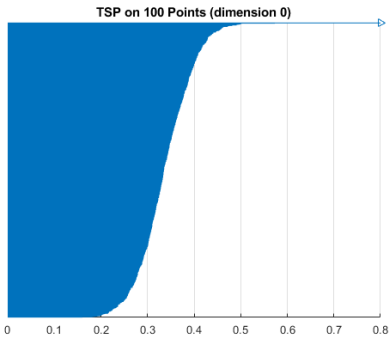
2. Compute persistence barcodes for data points $\hat{x}^{(1)}, \ldots, \hat{x}^{(M)}$ using a Čech complex, where the metric is the Hamming distance: the distance between two tours $\hat{x}^{(i)}, \hat{x}^{(j)}$ is

$$d(\hat{x}^{(i)}, \hat{x}^{(j)}) = \#(\text{edges belonging to only one of } \hat{x}^{(i)} \text{ and } \hat{x}^{(j)})$$

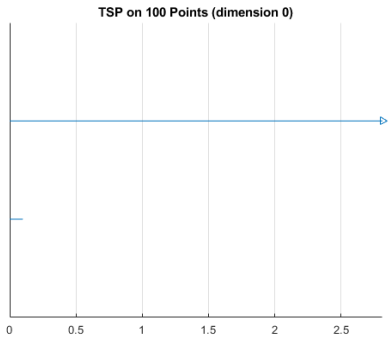
   In our example, we construct a Vietoris-Rips complex $VR(X, \epsilon)$ on the metric space X of the local minimizing tours equipped with the Hamming distance as the metric. This simplicial complex is a variant of

the Čech complex where $\{\hat{x}^{(1)}, \ldots, \hat{x}^{(k)}\}$ spans a k-simplex if and only if $d(\hat{x}^{(i)}, \hat{x}^{(j)}) \leq \epsilon$ for all $0 \leq i, j \leq k$. The β_0 persistence barcodes resulting from one such complex are shown in Figure 2a. The figure indicates that there are two persistent (long) barcodes.

3. Let $C_1, \ldots, C_m$ be the m longest barcodes corresponding to the 0-th Betti number. While m can be determined from the β_0 persistence barcodes generated by the Čech/Vietoris-Rips complex, in practice it can be difficult to identify m visually due to the large vertex set. In this case, a strong witness complex may provide a clearer visual aid. Using a strong witness complex can also reduce the computational expense of the Vietoris-Rips complex, which utilizes all of X as the vertex set. Given a set of landmark points $L \subseteq X$, we define for every point $\hat{x}^{(i)} \in X$ a distance $d_{\hat{x}^{(i)}} = \min_{l \in L} d(\hat{x}^{(i)}, l)$, i.e. its minimum distance to the set L. The strong witness complex $W(X, L, \epsilon)$ is then constructed with vertex set L with the following rule: $\{l_1, \ldots, l_k\}$ spans a k-complex if and only if there exists an $\hat{x}^{(i)} \in X$ such that $d(\hat{x}^{(i)}, l_j) \leq d_{\hat{x}^{(i)}} + \epsilon$ for all $1 \leq j \leq k$. The landmark points L can be chosen via a variety of methods, including random selection. Here we chose a set of 50 landmark points using sequential maxmin, a greedy selection process. The β_0 persistence barcodes for this example are shown in Figure 2b. This figure more clearly identifies the presence of the two persistent intervals first identified in the Vietoris-Rips complex, and confirms our choice of the $m = 2$ longest barcodes for this step. Furthermore, now that we have detected two connected components within the set X of local minimizing tours, we can examine a sample of tours from each of these components. Figure 3 shows a sample of 20 local minimizing tours from each of the two connected components.

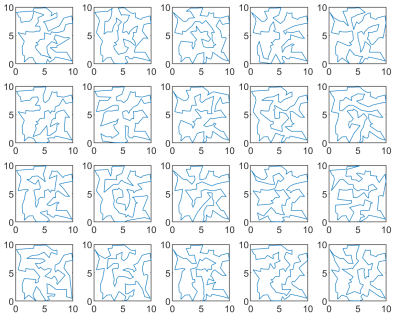


(a) β_0 persistence barcodes, $VR(X, \epsilon)$

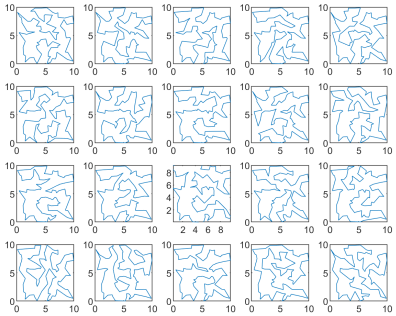


(b) β_0 persistence barcodes, $W(X, L, \epsilon)$

Figure 2: The β_0 persistence barcodes for a set of 1000 local minimizing tours from a TSP instance on 100 points. One can see that the two longest barcodes are more evident in the witness complex.



(a) Sample of size 20 from Component 1



(b) Sample of size 20 from Component 2

Figure 3: Local minimizing tours from a TSP instance on 100 points

4. For each edge e, define probability weights

$$w_e = \prod_{i=1}^{m} \Pr(e \text{ appears in } X | X \text{ is uniformly sampled from } C_i)\,.$$

When w_e is high, it means that e appears frequently in all of the clusters C_i, and is therefore more likely to be part of the optimal solution.

5. Perform N local searches with randomly seeded inputs, subject to the constraint that all edges e such that $w_e \geq \alpha$ are present (here α is specified beforehand).

6. Return the best solution obtained over all $M + N$ iterations.

2.3 Computational experiments

We implemented the procedure from section 2.2 and applied it to 41 benchmark problem instances from the TSPLIB library [7]: for efficiency's sake, we restricted ourselves only to those instances with up to 300 nodes, and set $M = N = 500$, $m = 3$, and $\alpha = 10^{-3}$. Our comparison procedure is as follows: after computing the first M minimizers, we then compute one additional set of N minimizers with TDA information, and another set of N minimizers without TDA information (that is, they are obtained using the same procedure as the original M). Table 1 shows the best objective values obtained with and without TDA. Out of the 41 instances, there are 30 for which local search with TDA is superior, 7 where the two are tied, and 4 where local search with TDA is inferior to vanilla local search.

Instance name	TDA	No TDA
a280	**2729**	2751
att48	**33556**	33601
berlin52	7545	7545
bier127	**119282**	119353
ch130	**6133**	6236
ch150	**6705**	6790
d198	16010	16010
eil101	**659**	660
eil51	434	**431**
eil76	**555**	558
gil262	**2489**	2501
gr137	**710**	711
gr202	498	498
gr229	**1682**	1683
gr96	**517**	520
kroA100	**21311**	21357
kroA150	26991	26991
kroA200	**30165**	30193
kroB100	**22322**	22441
kroB150	26716	**26582**
kroB200	**30351**	30419
kroC100	**20770**	20944
kroD100	**21538**	21559
kroE100	**22226**	22334
lin105	**14383**	14496
pr107	**44463**	44495
pr124	59031	59031
pr136	**98280**	98698
pr144	58535	58535
pr152	**73822**	74221
pr226	80883	**80729**
pr264	50984	**50871**
pr299	**49417**	49889
pr76	108305	108305
rat195	**2473**	2477
rat99	**1234**	1264
rd100	**7953**	7979
st70	**678**	684
ts225	**126905**	127316
tsp225	**4034**	4041
u159	**42768**	42781

Table 1: Best objective values obtained on TSPLIB benchmark problem instances, with and without TDA. A bold and underlined number indicates the superior (i.e. smaller) objective value.

3 A filtered complex for simulated annealing

We now present our main construction, which is a filtered complex associated to a stochastic matrix (in other words, a Markov chain), together with a steady state vector. We then consider several examples, with applications to discrete optimization and TDA. We find for instance in Section 3.2 that when the Markov chain is defined as a random walk on a finite metric space, we obtain a new construction for measuring the persistent homology of a data set. An advantage this has over Vietoris-Rips for instance, is that the zero-simplices now have nontrivial persistence value, making the common practice of removing sparse points unnecessary, as sparsity is a natural consequence of the construction.

By a filtered topological space, we mean a pair $(\mathcal{X}, F)$ where $\mathcal{X}$ is a topological space, and $F : \mathcal{X} \to \mathbb{R}$ is a continuous function. We then have the sublevel set persistent homology, defined as the image $H_*(\mathcal{X}^a) \to H_*(\mathcal{X}^b)$ for $a < b$, where $X^a = F^{-1}(-\infty, a]$ is the sublevel set. Similarly, a filtered complex is a pair (X, f) where X is an abstract simplicial complex, and $f : X \to \mathbb{R}$ has the property that $f(\Delta') < f(\Delta)$ whenever Δ' is a face of Δ. We then have a similar definition of the persistent homology group $H_*^{a,b}(X)$ as the image $H_*(X^a) \to H_*(X^b)$. Rather than considering specific values of a, b, one usually considers the persistence diagram or barcode diagram, which contains all of those groups at once, and has longer bar for more robust homology classes.

Let P be an $n \times n$ transition matrix satisfying

$$P_{i,j} \geq 0, \quad \sum_{j=1}^{n} P_{i,j} = 1 \text{ for all } i,$$

and let $\pi \in \Omega$ be a steady state row vector, which is an element of

$$\Omega = \left\{(p_1, ..., p_n) \in \mathbb{R}^n_{\geq 0} : p_1 + \cdots + p_n = 1\right\}$$

satisfying $\pi P = \pi$. Notice that Ω can be thought of as the geometric realization space of the complete $(n-1)$-dimensional simplicial complex, $\Omega = |K_n|$. We construct a filtration function $F : \Omega \to \mathbb{R}$, which favors distributions supported on denser points, and also favors distributions whose support is localized at points that are close together in a Markovian sense.

Construction 1. Let (P, π) be an $n \times n$ Markov chain together with a steady state. Define a second $n \times n$ matrix Q

$$Q_{i,j} = \sup_{m \geq 1} \left((P^m)_{i,j} - \pi_j \right), \tag{1}$$

noticing that $Q_{i,j} \geq 0$ for all i, j. We define a pair $(\mathcal{X}, F)$ by

$$F : \Omega \to \mathbb{R} \cup \{\infty\}, \quad F(p_1, ..., p_n) = -\log \left(\sum_{j=1}^{n} \pi_i \prod_{i=1}^{n} Q_{i,j}^{p_i} \right) \tag{2}$$

where $0^0 = 1$, and let $\mathcal{X} \subset \Omega$ be the subset on which F is finite. We have the discretized version, which is the filtered complex (X, f) given by

$$f : X \to \mathbb{R}, \quad f(\Delta) = \min_{p \in |\Delta|} F(p), \tag{3}$$

where $X \subset K_n$ is the subcomplex of the complete simplicial complex K_n of all simplices for which $f(\Delta)$ is not infinite. In other words, $f(\Delta)$ is the min of $F(p)$ over all $p \in \Omega$ for which $p_j = 0$ for $j \notin \Delta$.

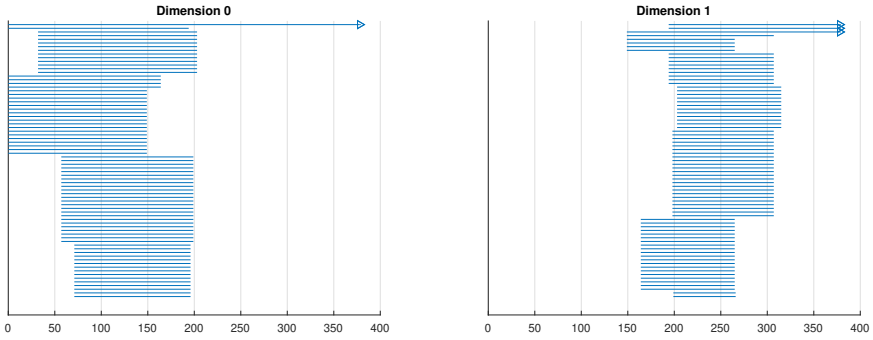
The motivation is to extend a natural measure of sparsity which is $-\log(\pi_j)$, to a continuous function on Ω, in a way that captures geometric features of the Markov chain. We find that the weighted geometric mean in (2) will be small between two points that are far apart in a Markovian sense, and interpolates between points that are close together. The discrete complex (X, f) is just the natural one to consider given $(\mathcal{X}, F)$.

The density measure on zero simplices is not exactly the same as the one mentioned in the previous paragraph. However, if Q is replaced by P^m for any power m in equation (2), then the value of F at any vertex i would simply be $-\log(\pi_i)$, since π is a steady state. We found this to produce interesting results, but it has the undesirable property that one has to make a particular choice of m. Low values could lead to the space being disconnected, whereas larger values would create too much blending, by filling in features. Considering the matrix Q as the max in (1) resolves this issue in a way that is independent of the time-scale coordinate. Notice that the effect of subtracting π_j becomes negligible if π_j is small, but makes a difference when there are a small number of states.

In practice, we have found the min in (3) may be taken by discretizing the simplex $|\Delta|$, taking the min only over those points whose coordinates are integer multiples of $1/k$. In most of our examples below, we have simply taken $k = 2$, which corresponds to taking the endpoints in Barycentric subdivision. We may also replace the supremum over all m in (2) with the max over a finite set $m \in S$, provided we replace negative values of $Q_{i,j}$ with 0, so that the weighted geometric mean is real-valued. Choosing $S = \{m\}$ to have one element is the same as just setting $Q = P^m$, as in the last paragraph.

3.1 Jeu de Taquin and configuration space

Consider a simple Markov chain whose states are the $9 \cdot 8 \cdot 7 = 504$ ways of placing the numbers 1,2,3 into distinct boxes in the 3×3 grid, and whose transitions are all ways of moving any number into an adjacent square, without moving into the space of another number, and with no self-transitions. Every legal move is then given the same probability. One can easily obtain a steady state exactly in this case.



(a) β_0-diagram (b) β_1-diagram

Figure 4: The persistence diagrams for Construction 1 applied to the "jeu de taquin" Markov chain. The 100 longest bars are shown. Despite having many shorter ones, one can see three lasting bars in the β_1 diagram.

The zeroth and first persistence diagrams of (X, f) are shown in Figure 4. One can see that there are three lasting β_1-bars. To interpret these classes, consider the (ordered) configuration space of three points in the plane:

$$\mathcal{X} = \left\{(p_1, p_2, p_2) : p_i \in \mathbb{R}^2,\ p_1 \neq p_2 \neq p_2\right\}.$$

Its homology groups are well-known, and in fact $H_1(\mathcal{X}, \mathbb{Q}) = \mathbb{Q}^3$. We interpret the three longest bars as being a discretized version of this homology group.

To justify this interpretation, we have shown a particular cycle representative in Table 2, which shows a loop in the space, in which the three numbers move around each other and return to their original position. Other cycles involve a single number being fixed, while the remaining two encircle each other. The natural continuous versions of such loops define nonzero classes in $H_1(\mathcal{X}, \mathbb{Q})$. Note that Construction 1 does not require edges to connect only adjacent vertices; they simply turn out to be the ones with the lowest persistence values in this example.

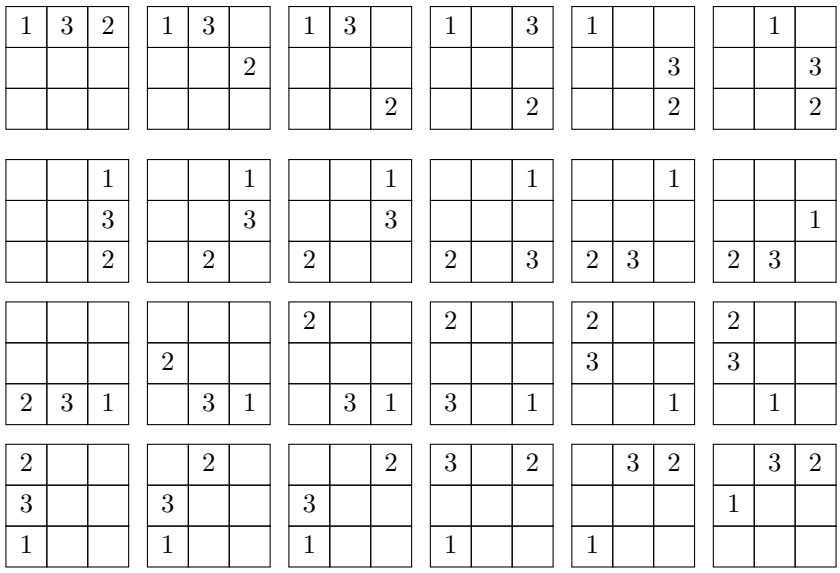
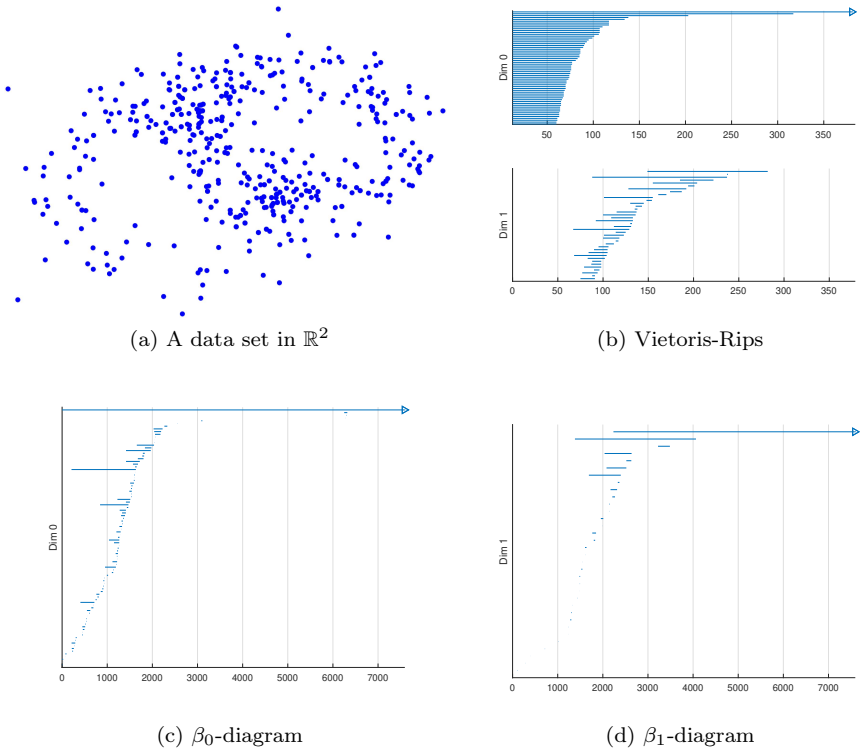


Table 2: The sequence of vertices in a representative cycle for the jeu de taquin game, generated by Javaplex.

3.2 Data sets and sparse points

In this example, we apply our complex to the standard application of persistent homology, which is to measure homological features of a point-cloud contained $\mathbb{R}^n$, typically using the Euclidean metric. To apply the construction, we simply replace the data set by the Markov chain of a random walk on it. We give an example which shows more prominent features when compared with Vietoris-Rips, and additionally does not require the usual preliminary step of removing outliers, as sparsity is already part of the construction.

Let $D \subset \mathbb{R}^2$ be the data set in Figure 5, which is sampled from a distribution with two visible modes (dense regions), and two more prominently visible voids. We have generated a stochastic matrix P whose states are in correspondence with points $p \in D$, and the transitions are determined by taking the 5 nearest neighbors to p, and then choosing one at random (no self-transitions). We then



(a) A data set in $\mathbb{R}^2$

(b) Vietoris-Rips

(c) β_0-diagram

(d) β_1-diagram

Figure 5: Persistence diagrams built from the data set in Figure 5a. Figure 5b shows the results of applying Vietoris-Rips, which shows some β_1 features reflecting the two voids in the data set. Figures 5c and 5d are built using a random walk, and show more distinguishable features, because outliers produce high persistence values.

took the unique steady state π, and built the pair (X, f) from Construction 1.

The resulting persistence diagrams are shown in Figures 5c and 5d. One can see a prominent bar in the the β_0 diagram, and a smaller one reflecting the two denser regions. In the β_1-diagram, there are two fairly clear one-dimensional bars. The Vietoris-Rips construction is less clear due to the presence of outliers, which are given high persistence values in Construction 1. Perhaps more importantly, the results are harder to predict, as adding more sampled points also produces more outliers.

3.3 Minimizing paths

The next example illustrates an application of higher-rank Betti numbers to optimization problems in high-dimensional state spaces. A similar idea here has been carried out in two dimensions with an application to the "stereo vision" problem in [3], for continuous optimization.

Let $F(x, y)$ be the two-variable Beale function [1] on the domain $\mathcal{X} = [0, 1] \times [0, 1]$, shown in Figure 6a. Consider the optimization problem whose states Ω are continuous functions $\varphi : [0, 1] \to [0, 1]$, and whose cost function is given by

$$E(\varphi) = \max_{x \in [0,1]} F(x, \varphi(x)).$$

This problem is nonconvex, and indeed has visible local optimizers given by the four paths described in Figure 6. In the context of Section 2, the next step would be to to study the zeroth persistent homology for the sublevel set filtration on $\mathrm{Map}([0, 1], [0, 1])$. It is unclear how to implement this, given the infinite-dimensional nature of the domain.

However, we also have a lower-dimensional space, which has interesting β_1-features as follows: let $\mathcal{A} = \{0, 1\} \times [0, 1] \subset \mathcal{X}$ denote the union of the left and right boundary segments of $\mathcal{X}$, corresponding to the values $x = 0, 1$. Then each path φ with $E(\varphi) \leq a$ represents a cycle which defines a *relative* homology class in $H_1^{a,b}(\mathcal{X}, \mathcal{A})$ for any $b > a$. Thus, we may divide the connected components of $E^{-1}(-\infty, a]$ into groups by associating to each $\varphi \in E^{-1}(-\infty, a]$ the induced class $\varphi_*(x)$, where

$$\varphi_* : H_1([0, 1], \{0, 1\}) \to H_1^{a,b}(\mathcal{X}, \mathcal{A}),$$

and x is a generator of $H_1([0, 1], \{0, 1\}) \cong H_1(S^1)$.

To implement this, we created a Markov chain whose states are the dis-

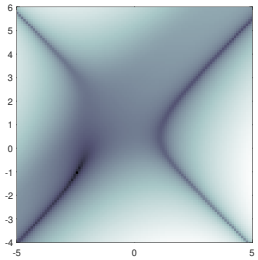
cretized points of $\mathcal{X}$ using a 35×35 grid. The transition matrix is determined be the simulated annealing algorithm, to favor low values of $F(x, y)$, with diagonal steps allowed so that there is a unique steady state. Again there are no self-transitions. To represent the relative homology groups, we define X to be the complex from Construction 1, together with a "point at infinity," and we add the complete simplicial complex connecting this point with each of the 70 points on the left and right boundary. We see in Figures 6e and 6f that we have four homology classes, as expected, as well as a nontrivial β_0 feature, indicating that there are two modes. We have used $m \in \{1, 2\}$ for the supremum in (2) instead of all numbers $m \geq 1$. The filtration function on 0-simplices $f(p)$ is shown for various other ranges $\{1, ..., k\}$ for comparison.

4 Conclusions and future directions

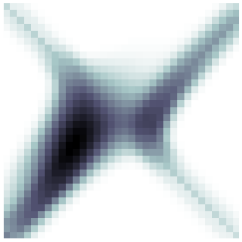
Construction 1 provides a viable definition of a filtered topological space associated to a stochastic matrix P (in other words a Markov chain), and a steady state vector π. For more involved applications to discrete optimization, computing its persistent homology directly from definition may not be practical if the number of simplices is prohibitively large. For instance, the traveling salesman example of Section 2 is itself an example of a Markov chain using the simulated annealing procedure, but the number of states of that Markov chain is $2^{|E|}$, where E is the edge set of the graph.

One direction is to approximate complex to (X, f) using "landmark points" in place of the entire state space. In an upcoming article [2], the first two authors have studied one such construction in the case of continuous optimization, which is described here:

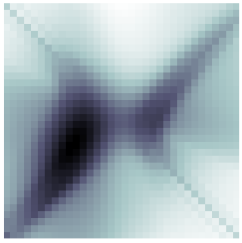
Construction 2. Suppose $\{x_1, ..., x_n\}$ are sampled from $\mathcal{X}$, and $F : \mathcal{X} \to \mathbb{R}$ is bounded below. We define a filtered complex (X, f) by



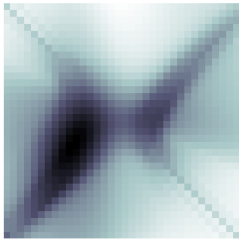
(a) Beale function



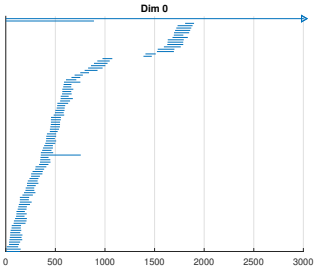
(b) $k = 2$



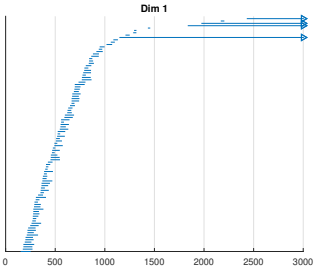
(c) $k = 8$



(d) $k = \infty$



(e) β_0 persistence diagram, $k = 2$



(f) β_1 for $k = 2$

Figure 6: A Beale function $F(x, y)$ on a region shown in 6a, then modeled on a 35×35 grid. There are 4 types of path $\varphi(x)$ with low values of $F(x, \varphi(x))$, upper left/lower left through upper right/lower right. The values of $f(p)$ on 0-simplices $p \in X$ for the complex from Section 3.3 are shown with $m \in \{1, ..., k\}$, for various values of k. Notice that the diagram for $k = 8$ is already quite similar to the $k = \infty$ case, which is the default in Construction 1. The major bars for the relative first Betti numbers correspond to the four paths in 6f. Notice there is a second bar in 6e corresponding to two dense regions.

1. The vertices of X are the numbers $\{1, ..., n\}$.

2. We have a k-dimensional simplex Δ with endpoints $\{i_0, ..., i_k\}$ if the region

$$\Omega_J = \left\{x \in \mathbb{R}^d : x_{j_0}, ..., x_{j_k} \text{ are the closest points to } x, \text{ in order}\right\}$$

   is nonempty for all $(k+1)!$ orderings $J = (j_0, ..., j_k)$ of the indices $\{i_0, ..., i_k\}$.

3. The filtration of a simplex $\Delta = \{i_0, ..., i_k\}$ is given by

$$f(\Delta) = \max_J \inf_{x \in \Omega_J} F(x), \tag{4}$$

   where the max is taken over the orderings J from item 2.

Optimizing a complex such as this one for discrete optimization problems is an interesting question which is left for future papers.

A second direction is to interpret the states of discrete optimization problems such as TSP as spaces of functions, and develop an approach along the lines of Section 3.3. The viability of this method is supported by the fact that only β_0-features, i.e. connected components were applied in Section 2, and that an analogous version was implemented in [3], in the case of the stereo correspondence problem from computer vision. To extend this method to TSP, the underlying graph must be endowed with some relevant geometry, which should most likely depend on the annealing procedure, such as 2-opt, or some other local search algorithm.

References

[1] EML Beale and John Forrest. Global optimization using special ordered sets. *Mathematical programming*, 10(1):52–69, 1976.

[2] E. Carlsson and J. Carlsson. A persistence complex for sublevel set filtration. *In preparation*, 2021.

[3] Erik Carlsson and John Gunnar Carlsson. Topology and local optima in computer vision. Technical report, UC Davis, 2021.

[4] Robert Ghrist and Sanjeevi Krishnan. A topological max-flow-min-cut theorem. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 815–818. IEEE, 2013.

[5] Keld Helsgaun. General k-opt submoves for the lin–kernighan tsp heuristic. *Mathematical Programming Computation*, 1(2-3):119–163, 2009.

[6] Tapas Kanungo, David M Mount, Nathan S Netanyahu, Christine D Piatko, Ruth Silverman, and Angela Y Wu. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE transactions on pattern analysis and machine intelligence*, 24(7):881–892, 2002.

[7] Gerhard Reinelt. Tsplib - a traveling salesman problem library. *ORSA journal on computing*, 3(4):376–384, 1991.

[8] Thomas Stützle. Iterated local search for the quadratic assignment problem. *European Journal of Operational Research*, 174(3):1519–1539, 2006.

[9] Mikael Vejdemo-Johansson and Primoz Skraba. Topology, big data and optimization. In *Big data optimization: recent developments and challenges*, pages 147–176. Springer, 2016.