Abstract simplicial complex

Simplicial complex

A simplicial complex $\mathcal{C}$ is a set of finite sets such that if $\sigma \in \mathcal{C}$ and $\tau \subseteq \sigma$ then $\tau \in \mathcal{C}$. For every $\tau \subseteq \sigma \in \mathcal{C}$, the set $\tau$ is a face of $\sigma$, whereas $\sigma$ is a coface of $\tau$.

$k$-simplex

$\sigma \in \mathcal{C}$ with $|\sigma| = k + 1$ is called a $k$-simplex.

Orientation

An orientation of $k$-simplices is an equivalence class of orderings where two simplices are considered equal if the permutation has a sign of 1.
Geometric realization

- Realize a $k$-simplex as the *convex hull* of $k + 1$ affinely independent points in some $\mathbb{R}^d$, with $d \geq k$.
- Need to ensure that the simplicess only intersect along *shared faces*.
- Geometric intuition:
  - 0-simplices: vertices
  - 1-simplices: edges
  - 2-simplices: triangles
  - 3-simplices: tetrahedra

Not interested in that.
The \( k \)th chain group \( C_k \) of \( C \) is the free abelian group on the set of oriented \( k \)-simplices. The group contains all abstract combinations of oriented \( k \)-simplices with coefficients from either a field or a principal ideal domain.

\[ c \in C_k \text{ is a } k\text{-chain, i.e.} \quad c = \sum_i \lambda_i [\sigma_i], \]

with \( \lambda_i \in \mathbb{Z} \), for example, and \( \sigma_i \in C \).
The $k$th boundary operator $\partial_k : C_k \rightarrow C_{k-1}$ is a homomorphism whose action on a chain $c$ is defined on a simplex $\sigma = [v_0, v_1, \ldots, v_k]$ by

$$\partial_k \sigma = \sum_i (-1)^i [v_0, v_1, \ldots, \hat{v}_i, \ldots, v_k],$$

where $\hat{v}_i$ signifies that the $i$th vertex is removed from the chain.
Mathematical background

Chain complex and subgroups

The boundary operators connect the chain groups of different dimensions. This forms a *chain complex*, i.e.

\[ \cdots \rightarrow C_{k+1} \rightarrow C_k \rightarrow C_{k-1} \rightarrow \cdots \]

Subgroups of $C_k$

We have the *cycle group* $Z_k = \ker \partial_k$ (mnemonic: “Zykel”) and the *boundary group* $B_k = \text{im} \partial_{k+1}$. Since $\partial_k \partial_{k+1} = 0$, the subgroups are nested:

\[ B_k \subseteq Z_k \subseteq C_k \]
Homology

$k$th homology group

\[ H_k = Z_k / B_k \]

This is well-defined because the subgroups are nested. The elements of the $k$th homology group are classes of *homologous cycles*. If the coefficients are taken from a field $\mathbb{F}$ then $H_k$ becomes a *vector space*.
**Betti numbers**

\[ \beta_k = \text{rank } H_k \]

- \( \beta_0 \) is the number of *connected components*
- \( \beta_1 \) is the number of 2-dimensional holes (circles)
- \( \beta_2 \) is the number of 3-dimensional holes (voids)
- \( \ldots \)
Homology is useful

- Invariants of topological spaces
- "Homology googles" to distinguish different spaces from one another

Classical example

<table>
<thead>
<tr>
<th></th>
<th>Möbius strip</th>
<th>Torus</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>$\mathbb{Z}$</td>
<td>$\mathbb{Z}$</td>
</tr>
<tr>
<td>$H_1$</td>
<td>$\mathbb{Z}$</td>
<td>$\mathbb{Z} \times \mathbb{Z}$</td>
</tr>
<tr>
<td>$H_2$</td>
<td>0</td>
<td>$\mathbb{Z}$</td>
</tr>
</tbody>
</table>
Applications

Input data

Assumption

The input data is given as a high-dimensional *point cloud*. There is some kind of *metric*, i.e. Euclidean distance.
Goal

Identify “interesting” topological structures in the data—especially relevant for time series data.
How to obtain a simplicial complex?

- Use points in point cloud as vertices of a graph
- Determine edges by *proximity*, i.e. take all vertices situated within a distance of $\epsilon$
- This yields the *neighbourhood graph* $N_\epsilon$
- Expand the graph afterwards

**Čech complex**

Topologically faithful but very hard to compute. Relies on *precise* distances.

**Vietoris-Rips complex**

Less expensive calculation but possibly different homotopy type, i.e. we may not “see” what we want to see.
How to choose $\epsilon$?

Figure: $\epsilon = 0.013$

Figure: $\epsilon = 0.019$
Persistent homology

- Need to distinguish between “essential” and “non-essential” holes
- Question of “optimal” values for $\epsilon$ is a mistake

**Idea**

Do *all* computations for a *large range* of parameter values for $\epsilon$. Features that *persist* over the course of varying the parameter are likely to be “real” topological features.
Figure: Default “barcode” visualization taken from [1].
 Workflow (so far)

- **Point cloud data** $\mathbb{R}^n$
- **Neighbourhood graph**
- **Simplicial complex** $\{\beta_0, \beta_1, \ldots, \beta_n\}$
- **Homology calculation**

Incremental/inductive expansion

Bastian Rieck (IWR)
Applied algebraic topology
September 7, 2012
Current status of my work

- Literature survey; we need to know the state of the art
- Implemented algorithms for constructing the Vietoris-Rips complex \([2]\)
- Started working on implementation of persistent homology calculation \([3]\)

Problems

- Complexes are very large
- Calculations are slow
- Not many applications out there (this may be a good thing)
Roadmap

- Even more literature survey
- Examination of some data sets—how can we profit from these methods?
- Try *approximations* to topology (sometimes we know the topology of the underlying space)
- Rather vague: Use *domain knowledge*

Possible applications

- Time-series data
- Clustered data
- ?
Robert Ghrist.
Barcodes: The persistent topology of data.

Afra Zomorodian.
Fast construction of the Vietoris-Rips complex.

Afra Zomorodian and Gunnar Carlsson.
Computing persistent homology.