Topological Descriptors for Geometric Data

Steve Oudot

Resources:

- http://geometrica.saclay.inria.fr/team/Steve.Oudot/courses/TUM/
- S.O. *Persistence Theory: from Quiver Representations to Data Analysis*. AMS Mathematical Surveys and Monographs (209), 2015.
### Geometric Data

**Input:** point cloud equipped with a metric or (dis-)similarity measure

**data point ≡** image/patch, geometric shape, protein conformation, patient, LinkedIn user...
Geometric Data

**Input:** point cloud equipped with a metric or (dis-)similarity measure

data point ≡ image/patch, geometric shape, protein conformation, patient, LinkedIn user...

**Goal:** describe the structure of the geometry underlying the data, for interpretation or summary
Context: the data deluge

Data are becoming more and more massive and complex:

- academia
- industry
- general public
Context: the data deluge

Data are becoming more and more massive and complex:

- academia
- industry
- general public

Need scalable and robust methods to analyze and classify these data.
Challenges

Scale

Noise

Dimensionality
Challenges

4 million data points in $\mathbb{R}^9$
(source: [Lee, Pederson, Mumford 2003])

Motivation: study cognitive representation of space of images

(source: [Carlsson, Ishkhanov, de Silva, Zomorodian 2008])

Topology
Challenges

4 million data points in $\mathbb{R}^9$

(source: [Lee, Pederson, Mumford 2003])

Motivation: study cognitive representation of space of images

Topology
Topological Data Analysis (TDA)

topological invariants for classification

\[ \beta_0 = \beta_2 = 1 \]
\[ \beta_1 = 2 \]

Algebraic topology in the 20th century
Algebraic topology in the 21st century

compact set

point cloud
Topological Data Analysis (TDA)

Properties of topological descriptors:
- invariant under coordinate changes
- stable with respect to perturbations
- informative

compact set

Algebraic topology in the 21st century

topological descriptors for inference and comparison

\[ \beta_0 \quad \beta_1 \quad \beta_2 \]

point cloud
The TDA community (as of 2002)

- 2 research groups (5-10 researchers)
This is our goal at large. To achieve it, we use concepts and tools from algebraic topology (A.T.).

The TDA community (as of 2016)

- 50-100 researchers working on theoretical foundations
- 200-300 researchers at the interface with applications
- Very successful applications and company (Ayasdi)
Some applications

- analysis of random, modular and non-modular scale-free networks and networks with exponential connectivity distribution,
- analysis of social and spatial networks like neurons, genes, online messages, air passengers, Twitter, face-to-face contact, etc.,
- coverage and hole detection in wireless sensor fields,
- multiple hypothesis tracking on urban vehicular data,
- analysis of the statistics of high-contrast image patches,
- image segmentation,
- 1d signal denoising,
- 3d shape classification/segmentation/matching,
- clustering of protein conformations,
- measurement of protein compressibility,
Some applications

- analysis of random, modular and non-modular scale-free networks and networks with exponential connectivity distribution,
- analysis of social and spatial networks like neurons, genes, online messages, air passengers, Twitter, face-to-face contact, etc.,
- coverage and hole detection in wireless sensor fields,
- multiple hypothesis tracking on urban vehicular data,
- analysis of the statistics of high-contrast image patches,
- image segmentation,
- 1d signal denoising,
- 3d shape classification/segmentation/matching,
- clustering of protein conformations,
- measurement of protein compressibility,
- identification of breast cancer subtypes,
- analysis of activity patterns in the primary visual cortex,
- identification of hidden patterns in the U.S. house of representatives,
- analysis of force networks in granular matter,
- analysis of regimes in dynamical systems.
Some applications

- analysis of random, modular and non-modular scale-free networks and networks with exponential connectivity distribution,
- analysis of social and spatial networks like neurons, genes, online messages, air passengers, Twitter, face-to-face contact, etc.,
- coverage and hole detection in wireless sensor fields,
- multiple hypothesis tracking on urban vehicular data,
- analysis of the statistics of high-contrast image patches,
- image segmentation,
- 1d signal denoising,
- 3d shape classification/segmentation/matching,
- clustering of protein conformations,
- measurement of protein compressibility,
- identification of breast cancer subtypes,
- analysis of activity patterns in the primary visual cortex,
- identification of hidden networks in the U.S. house of representatives,
- analysis of 2d cortical thickness data,
- time series analysis,
Some applications

- analysis of random, modular and non-modular scale-free networks and networks with exponential connectivity distribution,
- analysis of social and spatial networks like neurons, genes, online messages, air passengers, Twitter, face-to-face contact, etc.,
- coverage and hole detection in wireless sensor fields,
- multiple hypothesis tracking on urban vehicular data,
- analysis of the statistics of high-contrast image patches,
- image segmentation,
- 1d signal denoising,
- 3d shape classification/segmentation/matching,
- clustering of protein conformations,
- measurement of protein compressibility,
- identification of breast cancer subtypes,
- analysis of activity patterns in the primary visual cortex,
- identification of hidden networks in the U.S. house of representatives,
- analysis of 2d cortical thickness data,
- time series analysis,
- refinement of the classification of NBA players,
- discrimination of electroencephalogram signals recorded before and during epileptic seizures,
- statistical analysis of orthodontic data,
- measurement of structural changes during lipid vesicle fusion,
- characterization of the frequency and scale of lateral gene transfer in pathogenic bacteria,
- pattern detection in gene expression data,
- study of the cosmic web and its filamentary structure,
The TDA pipeline in a nutshell

- Point cloud $P$
- Proximity rule
- Filter $P \rightarrow \mathbb{R}$
- Filtration $\mathcal{F}$
- Homology
- Persistence
- Barcode (descriptor)
The TDA pipeline in a nutshell

3 pillars to the theory (topological persistence):
- decomposition theorems (\( \exists \) barcodes)
- algorithms (computation of barcodes)
- stability theorems (barcodes as stable descriptors)
The TDA pipeline in a nutshell

3 pillars to the theory (topological persistence):

- decomposition theorems (∃ barcodes)
- algorithms (computation of barcodes)
- stability theorems (barcodes as stable descriptors)

menu for today