Algebraic and Differential Topology in Data Analysis (Math 595 ADT)

Yuliy Baryshnikov

The course will cover some recent applications of topology and differential geometry in data analysis.

Tools of differential and algebraic topology are starting to make noticeable impact in the area of data sciences, where the mathematical apparatus thus far was dominated by the ideas from statistical learning, computational linear algebra and high-dimensional normed space theories.

While the research community around ADT topics in data analysis is lively and fast-growing, the area is somewhat sparsely represented in campus syllabi. I wanted to test ground whether there is a constituency for a course covering some of it (with the choice of topics, of course, tilted towards what I find interesting).

There are several courses on Data Analysis on campus (IE 531, CSE 448, ECE 566); however they have thematically virtually no overlap with the proposed syllabus.

I expect besides Math students interested in applications of geometric and topological tools, some participation from CoE students.

Syllabus [43][1]

1. Tools from algebraic topology [4]
   - 1.1. Homotopy equivalence
   - 1.2. Simplicial homology
   - 1.3. Nerve lemma, Dowker's theorem
   - 1.4. **Applications**:
     Sketches: Merge trees; Reeb graphs; skeletonizations
2. Tools from Hodge theory [4]
   - 2.1. Laplacians, eigenvalues, eigenfunctions
   - 2.2. Interplay between geometry and spectral properties. Cheeger inequality
   - 2.3. Expanders
   - 2.4. **Applications**:
     Synchronization, Clustering
3. Tools from differential topology [7]
   - 3.1. Basic notions: topological spaces, manifolds, simplicial complexes.
   - 3.2. Transversality.
   - 3.3. Sard's theorem.

---

[1] [*] indicates hours allocated

- 3.4. Whitney's embedding theorem.
- 3.5. **Applications**:
  Dimensionality reduction: embedding based tools (PCA, Nonlinear PCA; Random Projections; Multidimensional Scaling); embedding generating tools (Isomap; Eigenmap; Diffusion Maps; LLE).

4. Topological Approximations [7]
   - 4.1. Vietoris-Rips and Čech complexes.
   - 4.2. Topology reconstruction from dense samples: Hausmann-Latscher
   - 4.3. Topology reconstruction from random samples: Niyogi-Smale-Weinberger
   - 4.4. **Applications**:
     Netflix problem complexes
     Cech complexes and their topology in robotic and neurophysiology.
     Merge trees in time series analysis
5. Euler calculus [6]
   - 5.1. Set algebras and valuations
   - 5.2. Hadwiger and McMullen's theorems
   - 5.3. Integration with respect to Euler characteristics
   - 5.4. Fundamental Kinematic Formula
   - 5.5. **Applications**:
     Gaussianity tests for random fields
     Topological Sensor Network
6. Topological Inference [6]
   - 6.1. Persistent Homology
   - 6.2. Algorithms
   - 6.3. Stability
   - 6.4. **Applications**:
     Image patches spaces
     Textures and characterization of materials
7. Aggregation [4]
   - 7.1. Spaces with averagings
   - 7.2. Arrow theorem and Topological Social Choice
   - 7.3. Aggregation in CAT(0) spaces
   - 7.4. **Applications**:
     Consensus in phylogenetic analysis
8. Clustering [3]
   - 8.1. Basic clustering tools
   - 8.2. Kleinberg's Impossibility theorem.
   - 8.3. Carlsson-Memoli functorial approach to clustering.
9. Students presentations [2]

With exception of background material, covered in the books below, the course will rely mainly on the recent papers.

R Ghrist, Elementary Applied Topology, Createspace, 2014
A Hatcher, Algebraic Topology, CUP, 2002
V Prasolov, Elements of Combinatorial and Differential Topology, AMS, 2006

The students will be given 4-5 homework assignments, to practice for the final.
Each student will be expected to either present a paper (from a sample provided by me), or to run a computational project (also designed by me). The presentations account for 40% of the overall score; the final exam, 60%.