MATH 581: Image and Data Analysis

Instructors: Kevin R. Vixie, Thomas J. Asaki
and Bala Krishnamoorthy
http://datachallengecooperative.org/math581/
Cross-listed as **UI Math 541**

MWF 2:10–3:00 pm, Sloan 46, Washington State University

1 Introduction

As scientists and citizens, we are being overwhelmed by an exponentially increasing flood of data. From video and hyperspectral streams, distributed sensor networks with thousands of nodes, and network monitoring to automated astronomical surveys and astrophysical space instruments, the need for cleverness and innovation in data analysis and information extraction only increases with time. How can we extract useful information in real time from streaming data, so that we store only the useful bits extracted as the stream flies by us? How do we optimize the use of sparse measurements from extremely high dimensional systems? How can we characterize with precision the propagation of uncertainty in the high or infinite dimensional systems that are typical of many data processing pipelines? Questions like these are far from settled and will require advances on all fronts, theoretical, experimental, and computational.

The prerequisites for this course are an excellent grasp of calculus (including vector calculus), linear algebra and a willingness to work. Projects will involve programming in octave or Matlab, which the student is expected to have or obtain (octave is free, Matlab is not free). We now turn to a brief description of the contents of the course.

2 Contents, Briefly

We cannot hope to even begin to cover all of image and data analysis, so we have chosen one particularly rich path to explore. Our decisions of what to cover was largely determined by our own research threads: teaching what inspires us is most likely to inspire you. And focusing on the (rather big) neighborhood of our research makes it possible for us to provide a sense for where the current cutting edge is.

In the sequence that we will follow:

{image denoising} → {image segmentation} → {shape characterization and data signatures} → {learning from data (classification, detection, recognition)}

we are (roughly) moving up in the level of task from low level processing to high(er) level understanding tasks.

We will find that the coexistence of low and high dimensional geometry of image and image-like data brings an incredible richness to the study, opening up vast areas of mathematics as potential sources of insights for us to exploit.

The philosophy behind our approach to research combines both an "applied", *problem-driven perspective* and a "pure", *pursuit of deeper mathematical questions*.

While the *problem-driven* perspective places a high premium on working algorithms in response to real data challenges, the work on *deeper theoretical questions* means that motivated students can work on *both* theoretical and applied research through the open problems and projects we will introduce from time to time.

Here is a list of the 43 lectures:

- **Introduction and Overview:** a panorama and a bit of philosophy
    - Noisy images: Rudin-Osher-Fatemi (ROF) as a case study – Bayes law, negative log-likelihood, dimension reduction, regularization
    - Noisy images: L1TV – L1 vs L2 makes a big difference.
    - Noisy images: L1TV and the flat norm: – a spin off and deeper theory
    - Noisy images: Graph denoising – Another approach that asks why do it this way?
    - Noisy Images: Graph denoising – Asaki et al.
    - Noisy images: Computation – gradient descent
    - Noisy images: Computation – graph cuts, min cut/max flow
- **Transition:** from noise removal to shape extraction
    - Shapes from images: Segmentation
    - Shapes from images: The Mumford-Shah Functional
    - Shapes from images: Computation via Chan-Vese and Esedoglu-Tsai
    - Shapes from images: Level Set methods
    - Shapes from images: Tubular neighborhoods and curvature measures
    - Shapes from images: Shapes from data – intro and overview of computational topology
- **Transition:** From shape extraction to shape comparison
    - Topology of Shapes and Data: Voronoi diagrams and Delaunay tessellation
    - Topology of Shapes and Data: Union of balls, dual complex, Alpha Shapes, Rips complex
    - Topology of Shapes and Data: Filtration, "birth" and "death" of features
    - Topology of Shapes and Data: Algebraic topology: homeomorphism, homotopy
    - Topology of Shapes and Data: Simplicial complexes, Homology groups, Betti numbers
    - Topology of Shapes and Data: Computation: Incremental and matrix algorithms

- Topology of Shapes and Data: Persistent homology groups
  - Topology of Shapes and Data: Persistence barcodes for shapes
  - Topology of Shapes and Data: Examples
  - Shape Measures and Data Signatures: Geomeasure signatures
  - Shape Measures and Data Signatures: Flat norm signature

- **Transition:** from shape and data comparison to data classification
  - Learning from data: **Probabilistic view of data analysis**
  - Learning from data: PCA and SVD – eigenfaces
  - Learning from data: CMODI – invariance to transformations
  - Learning from data: Diffusion geometry and Kernel PCA
  - Learning from data: Clustering – intro, overview
  - Learning from data: Clustering – muon data
  - Learning from data: FLD, regression, etc
  - Learning from data: SVMs
  - Learning from data: **High dimensions and Concentration of measure**
  - Learning from data: Random Projections and Compressive sensing
  - Learning from data: Finding transformations – Registration, TSWARP
  - Learning from data: Finding transformations – Map Seeking Circuit
  - Learning from data: mixed variable optimization I
  - Learning from data: mixed variable optimization II

- **Summary and links to further studies:** Research teasers, IPAM, IMA, SIAM SIAG, workshops

3 Projects

There will three projects that must be completed for credit in the course. Students will choose three of the following to complete.

Image denoising: implement and test denoising methods on images with different levels and types of of noise.

Image segmentation: implement a segmentation method and extract shapes from a standard test set that we will hand out.

Computational Topology: implement code to calculate that various invariants described in class. Use this to generate "bar codes".

Shape signatures: use algorithms to study what information can be extracted from images using geomeasures or the flat norm signatures (or any other signature that you invent or find).

Dimension reduction: study the dimensionality of several sets of data that we will distribute using various algorithms we make available or give details on how to implement.

Pattern Classification: build pattern classification algorithms using tools learned about in course.

Map seeking circuits: implement and use the map seeking circuit to find objects in images.

4 References and Annotated Bibliography

The course will, for the most part, be based on notes. The part of the course on computational topology will use Zomorodian's book (see below). Below we give a list of references and resources that are useful for deeper study.

Image Analysis While there are many books on image analysis and processing, we have chosen to recommend *Image Processing and Analysis* by Tony Chan and Jackie Shen to those that want to buy one book on imaging. It covers the three main branches - PDE/variational, wavelet, and statistical - of mathematical image analysis quite well. While not perfect (no book is) it is the most comprehensive book available. The authors are very creative and prolific contributers to the field, which enabled them to write a book that presents intuitions as well as techniques and algorithms. Perusing paper repositories like the CAM web site at UCLA is also quite useful.

Functions of Bounded Variation To understand total variation regularized image methods, one must understand functions of bounded variation. For this purpose, we recommend *Measure Theory and Fine Properties of Functions* by Evans and Gariepy. While this book is quite advanced, Chapter 5 is the best coverage of BV functions we know of. Another good reference is Giusti's *Minimal Surfaces and Functions of Bounded Variation.* For a deeper understanding of the Mumford-Shah functional, there are a few references that one can consult. First there is the monograph by Morel and Solimini *Variational Methods in Image Segmentation.* Then there is the book by Ambrosio, Fusco and Pallara – *Functions of Bounded Variation and Free Discontinuity Problem* – in which they present methods allowing them to prove existence of minimizers for the Mumford-Shah functional.

Geometric Measure Theory To penetrate the depths involving the geometric and analytic nature of sets, measures and functions, one can should geometric measure theory. Instead of immediately diving into Federer's treatise *Geometric Measure Theory*, we recommend first studying Frank Morgan's *Geometric Measure Theory: a beginners Guide, 4th Edition* and then Leon Simon's *Geometric Measure Theory* (you can get this directly from Australia by Calling or emailing Canbera). While studying these two books, one can begin reading pieces of Federer's book. There is also the nice new notes of Camillo De Lellis on rectifiable sets, *Rectifiable Sets, Densities and Tangent Measures* but this is more specialized.

Topological Data Analysis Afra J. Zomorodian's book, *Topology for Computing* is the reference that will be used for this course. There are others like *Computational Homology* by Tomasz Kaczynski, Konstantin Mischaikow, and Marian Mrozek as well as

Herbert Edelsbrunner's book *Geometry and Topology for Mesh Generation* that are also useful.

Information Theoretic and Statistical methods No library is complete without a copy of Cover and Thomas' *Information Theory*. It is simply the best book on information theory and the book to learn it from. And if you are interested in data of any kind, you should have gone through this book at some point in your education. The statistical/stochastic viewpoint is a prevalent viewpoint. But this has not led to expositions that we find appealing or particularly enlightening or in line with how we think about these subjects. We will update you on what we find, but for now plan on asking us for more references for particular bits and pieces in this area.

Dimension Reduction and High Dimensional Geometry We recommend Michael Kirby's book *Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns*. Another, that we will be reviewing and that looks interesting is *Nonlinear Dimensionality Reduction* by John A. Lee and Michel Verleysen.

Concentration of Measure For this subject we recommend Michel Ledoux's (thin!) AMS monograph *The Concentration of Measure Phenomenon*.

Computational Methods Curt Vogel's book *Computational Methods for Inverse Problems*, is a highly recommended. One can find algorithms and explanations as well as code for the Book on Vogel's website.

Optimization Good texts on various optimization topics include *Introduction to Linear Optimization* by Bertsimas and Tsitsiklis and *Nonlinear Optimization* by Bertsekas.

5 Schedule In Detail

Lecture 1: Monday, Jan 12 **Introduction and Overview:** a panorama and a bit of philosophy **[Vixie]**

Lecture 2: Wednesday, Jan 14 Noisy images: ROF as a case study – Bayes law, negative log-likelihood, dimension reduction, regularization **[Vixie]**

Lecture 3: Friday, Jan 16 Noisy images: L1TV – L1 vs L2 makes a big difference. **[Vixie]**

Holiday: Monday, Jan 19 Martin Luther King Day

Lecture 4: Wednesday, Jan 21 Noisy images: L1TV and the flat norm: – a spin off and deeper theory **[Vixie]**

Lecture 5: Friday, Jan 23 Noisy images: Graph denoising – Another approach that asks why do it this way? **[Asaki]**

Lecture 6: Monday, Jan 26 Noisy Images: Graph denoising – Asaki et al. **[Asaki]**

Lecture 7: Wednesday, Jan 28 Noisy images: Computation – gradient descent **[Krishnamoorthy]**

Lecture 8: Friday, Jan 30 Noisy images: Computation – graph cuts, min cut/max flow **[Krishnamoorthy]**

Lecture 9: Monday, February 2 **Transition:** from noise removal to shape extraction **[Asaki]**

Lecture 10: Wednesday, February 4 Shapes from images: Segmentation **[Asaki]**

Lecture 11: Friday, February 6 Shapes from images: Mumford Shah **[Vixie]**

Lecture 12: Monday, February 9 Shapes from images: Chan-Vese and Esedoglu-Tsai M-S **[Vixie]**

Lecture 13: Wednesday, February 11 Shapes from images: Level Set methods **[Vixie]**

Lecture 14: Friday, February 13 Shapes from images: Tubular neighborhoods and curvature measures **[Vixie]**

Holiday: Monday, February 16 Presidents Day

Lecture 15: Wednesday, February 18 **Transition:** From shape extraction to shape comparison **[Vixie]**

Lecture 16: Friday, February 20 Topological methods for data: – motivating examples, Voronoi diagram and Delaunay tessellation **[Krishnamoorthy]**

Lecture 17: Monday, February 23 Topological methods: Homeomorphism, homotopy, deformation retraction **[Krishnamoorthy]**

Lecture 18: Wednesday, February 25 Topological methods: Union of balls, power Voronoi diagram, dual complex **[Krishnamoorthy]**

Lecture 19: Friday, February 27 Topological methods: Filtration, demo of Plex, simplicial complexes, triangulation **[Krishnamoorthy]**

Lecture 20: Monday, March 2 Topological methods: Abelian groups, chains, cycles, and boundaries **[Krishnamoorthy]**

Lecture 21: Wednesday, March 4 Topological methods: Boundary homomorphism, "size" of k-th homology group, examples **[Krishnamoorthy]**

Lecture 22: Friday, March 6 Topological methods: Betti numbers, Euler characteristic, incremental algorithm **[Krishnamoorthy]**

Lecture 23: Monday, March 9 Topological methods: Persistent homology, computation **[Krishnamoorthy]**

Lecture 24: Wednesday, March 11 Topological methods: Barcode shape descriptors, computation in Plex **[Krishnamoorthy]**

Lecture 25: Friday, March 13 Topological methods: Barcodes, examples **[Krishnamoorthy]**

Holiday: Monday, March 16 Spring Break

Holiday: Wednesday, March 18 Spring Break

Holiday: Friday, March 20 Spring Break

Lecture 26: Monday, March 23 Shape Measures and Data Signatures: Geomeasure signatures **[Asaki]**

Lecture 27: Wednesday, March 25 Shape Measures and Data Signatures: Flat norm signature **[Asaki]**

Lecture 28: Friday, March 27 Transition: from shape and data comparison to data classification **[Asaki]**

Lecture 29: Monday, March 30 Learning from data: **Probabilistic view of data analysis** **[Vixie]**

Lecture 30: Wednesday, April 1 Learning from data: PCA and SVD – eigenfaces **[Vixie]**

Lecture 31: Friday, April 3 Learning from data: CMODI – invariance to transformations **[Vixie]**

Lecture 32: Monday, April 6 Learning from data: Diffusion geometry and Kernel PCA **[Vixie]**

Lecture 33: Wednesday, April 8 Learning from data: Clustering – intro, overview **[Asaki]**

Lecture 34: Friday, April 10 Learning from data: Clustering – muon data **[Asaki]**

Lecture 35: Monday, April 13 Learning from data: **Concentration of measure** **[Vixie]**

Lecture 36: Wednesday, April 15 Learning from data: FLD, regression **[Krishnamoorthy]**

Lecture 37: Friday, April 17 Learning from data: SVMs **[Krishnamoorthy]**

Lecture 38: Monday, April 20 Learning from data: Random Projections and Compressive sensing **[Asaki]**

Lecture 39: Wednesday, April 22 Learning from data: Finding transformations – Registration, TSWARP **[Asaki]**

Lecture 40: Friday, April 24 Learning from data: Finding transformations – Map Seeking Circuit **[Asaki]**

Lecture 41: Monday, April 27 Learning from data: mixed variable optimization I **[Asaki]**

Lecture 42: Wednesday, April 29 Learning from data: mixed variable optimization II **[Asaki]**

Lecture 43: Friday, May 1 Summary and links to further studies: Research teasers, IPAM, IMA, SIAM SIAG, workshops **[Vixie]**