Electrical and Computer Engineering
Northeastern University
Fall 2021

EECE7345 Big Data and Sparsity in Control, Machine Learning and Optimization

Description: Arguably, one of the hardest challenges faced now by the systems and machine learning communities stems from the exponential explosion in the availability of data. Simply stated, classical techniques are ill equipped to analyze very large volumes of (heterogeneous) data. To overcome these shortcomings, during the past few years a large research effort has been devoted to developing computationally tractable methods that seek to mitigate the "curse of dimensionality" by exploiting the self-similarity and sparsity properties underlying many problems. The goal of this course is twofold: (1) provide an introduction to the subject for people in the systems, machine learning and computer vision communities faced with "big data" and scaling problems, and (2) serve as a "quick reference" guide, summarizing the state of the art as of today and provide a comprehensive set of references. Part I of the course covers the issue of handling large data sets and sparsity priors, presenting very recently developed techniques that exploit a deep connection to semi-algebraic geometry, rank minimization and matrix completion. Part II of the course continues this theme, but focusing on applications, including control and filter design subject to information flow constraints, subspace clustering and classification on Riemannian manifolds, and activity recognition and classification and anomaly detection from video sequences.

Prerequisites: EECE 5644 or EECE 7323. By topic: Basic knowledge of Matrix Algebra and Optimization. Basic knowledge of programming in Matlab.

Class Schedule: Tu. 11:45-1:25, Th. 2:50–4:30, Room 309, Kariotis Hall.

Instructor: Professor Mario Sznaier, Room 414, Dana Bldg., 373–5364,
email: msznaier@coe.neu.edu. Office Hours: Tu & Th 9:00–10:00 (Zoom) or by appointment.

Teaching Assistant: Jared Miller, Office Hours T.B.D

Grading: will be based on paper presentations, paper reviews and a short project involving application of the ideas discussed in class to a problem, as follows:

Presentations **(40% of total grade)** Teams of 2 students will give a 20 minutes presentation on a paper related to the course. This presentation should roughly cover the following points:

(a) Statement of the problem

(b) Motivation: who cares? Why is the problem important, interesting, difficult to solve?

(c) Description of key contributions and technical ideas.
(d) Relation to the material in the course.
(e) Impact and follow-ups.
(f) Open research questions, possible applications, extensions, etc.

You are encouraged to search material from the authors to enhance your presentation. Look for their project page, check if they make the software available. It is ok to borrow slides. However, each of the slides should acknowledge the source. **Deadline for forming the teams is September 28** (earlier is better).

Paper Reviews **(20% of total grade)** Each student is required to write a brief (one page) critical review of the paper being presented each week. The review should address the following points:

(a) Give a paragraph with a brief summary of the paper.
(b) State what is the main contribution of the paper.
(c) Briefly cover the main technical points.
(d) Additional comments, including unclear points, open research questions, and applications.

Final Project **(40% of total grade)** Teams of two students will complete a term project and write a report. The project can be either mostly theoretical (applications of the ideas discussed in class to a new problem, requiring substantial extension of these ideas and development of new algorithms) or mostly experimental (application of an existing algorithm/technique to a new problem). Theoretical papers need to present a proof-of-concept example, applied ones need to have a comprehensive experimental protocol and extensive testing. The final projects will be graded as follows:

Proposal (**5% of grade**) A two page proposal describing the project, **due on 11/4/2021**.

Report (**20% of grade**) The report should have no more than 8 pages, double column, single spaced plus up to 2 pages with references. It should be written in a format suitable for submission to a major conference in your field. Each group is expected to submit a a zip file with the presentation, code and additional material such as video demos. A draft paper is due on **11/23/2021** and the final version on **12/3/2021**.

Presentation (**15% of grade**) Each group will give a 20 minutes presentation of the project during the last two weeks of classes, following the same guidelines used for paper presentations.

Some References:

Convex Optimization Boyd and Vandenberghe (http://stanford.edu/ boyd/cvxbook/)

Polynomial Optimization J. B. Lasserre, Moments, Positive Polynomials and Their Applications, Imperial College Press, 2009. ISBN 978-1-84816-445-1. Check also Lasserre's home page (http://homepages.laas.fr/lasserre/) and Monique Laurent's monograph (http://homepages.cwi.nl/ monique/files/moment-ima-update-new.pdf)

Other useful resources Pablo Parrilo's homepage (http://www.mit.edu/ parrilo/), Maryam Fazel's homepage (faculty.washington.edu/mfazel/) and Ashkan Jasour robust optimization course homepage (https://rarnop.mit.edu/risk-aware-and-robust-nonlinear-planning)

Syllabus (by topic)

- Mathematical Foundations
  - Review of convex optimization and Linear Matrix Inequalities
  - Promoting sparsity via optimization. Convex surrogates for cardinality and rank.
  - Unconstrained and constrained optimization.
  - Fast algorithms for convex optimization.
  - Review of semi-algebraic geometry
  - Polynomial optimization: Sum-of-squares and moments based approaches.
  - Exploiting sparsity in polynomial optimization
- Sparsity in Systems Identification
  - Identification of LTI systems with missing data and outliers
  - Identification of sparse graphical models
  - Identification of Wiener systems
  - Semi-supervised identification of switched affine systems
  - Model invalidation.
- Sparsity in Control and Estimation
  - Synthesis of controllers subject to sparsity constraints: the convex case
  - Synthesis of controllers subject to sparsity flow constraints: the general case
  - Sparse filter design for LTI systems
  - Worst case optimal filters for switched systems
- Connections to Machine Learning
  - Robust regression and subspace clustering
  - Benign overfitting and generalization
  - Manifold embedding as a Wiener identification problem
- Examples of Applications
  - Actionable information extraction as a SysId problem
  - Finding causal interactions as a sparse graphical model identification
  - Recovering 3D geometry from 2D data as a Wiener identification
  - Anomaly detection as a model (in)validation problem.
  - Multitarget tracking as a rank-minimizing assignment problem.

Syllabus (tentative schedule)

- Introduction
- Review of convexity: convex sets, operations that preserve convexity, separating hyperplanes
- Properties of convex functions: first and second order convexity conditions, Jensen's inequality, the conjugate function, convex envelope, connections to promoting sparsity.
- Convex optimization problems: Linear Programs, Second Order Cone Programs, Semi-Definite Programs, Linear Matrix Inequalities.
- Duality: weak and strong duality, geometric interpretation, KKT conditions, duals of some important problems, Theorem of Alternatives and applications.
- Promoting sparsity via optimization: convex surrogates for cardinality and rank, atomic norms, applications to robust regression, matrix completion, identification with missing data, recovering 3D geometry from 2D data, multi-target tracking as a rank minimization problem.
- Solving unconstrained optimization problems: Newton's method, stochastic gradient descent, machine learning applications.
- Solving convex constrained optimization problems via interior point methods.
- First order methods: dual ascent, ADMM, Frank Wolfe.
- Non differentiable functions: subgradients and subdifferentials, computing proximal operators via Moreau decomposition, Moreau-Yoshida envelope and proximal gradients methods.
- Applications of first order methods: consensus optimization problems, distributed model identification, support vector machines, finding sparse solutions to systems of linear equations, outlier rejection, matrix decomposition with applications to image segmentation and subspace clustering, structured matrix decomposition problems.
- Sparsity in Control and Estimation: synthesis of controllers subject to information flow constraints (convex case), sparse filter design for LTI systems.
- Sparsity in Systems Identification: identification of linear systems with missing data and outliers, identification of sparse graphical models, identification of Wiener systems, learning Koopman operators from data.

- Sparsity in Machine Learning: robust regression and subspace clustering, manifold discovery and non-linear embedding of dynamic data, actionable information extraction as a sparse identification problem, finding causal interactions as a sparse graphical model identification, outlier detection.

- Polynomial optimization: Sum-of-squares and moments based approaches.

- Exploiting sparsity in polynomial optimization

- Applications of polynomial optimization: robust subspace clustering, time series segmentation, anomaly detection in video data, synthesis of controllers subject to sparsity flow constraints (the general case), worst case optimal filters for switched systems, semi-supervised and unsupervised identification of switched affine models.