SYLLABUS FOR IE 690000: TOPICS IN MODERN CONTINUOUS OPTIMIZATION ALGORITHMS

> "Nothing in the World takes place without Optimization, and there is no doubt that all aspects of the World, that have a rational basis, can be explained by Optimization Methods."
> – Leonard Euler

1. Overview

Optimization is ubiquitous, from fitting the GPT model, figuring out an optimal investment portfolio, to training autonomous UAVs and many more. Central to all these exciting applications are the increasingly sophisticated models (e.g. the objective functions for the deep neural network model is non-convex and non-smooth) and their enormous scale (e.g. GPT has over 100 billion parameters to tune and its corpus consists of the entire internet). The speed with which we could solve the underlying optimization problems is a major bottleneck to their success. This class seeks to illustrate this issue by developing theoretical understandings for the convergence speed and presenting the recent advances in algorithm design.

Naturally, one would desire one general algorithm to solve all problems effectively. However, the seminar works of Nemirovsky, Nesterov and Yudin refuted the existence of such an algorithm, i.e., there is some hard problem for which no algorithm could find a good solution in finite time. Instead, one should consider some narrower problem class, and exploit the inherent problem structure to design effcient algorithms with sound theoretical convergence guarantees. In this class, we focus on the optimization problem classes fundemental to machine learning and control, that is, stochastic optimization, non-convex optimization, non-smooth optimization, and variational inequality.

2. Course Outline

Specifically, we plan to cover the following topics.

- Quick review of the optimization complexity theory (3 weeks)
  - The basic terminology: convexity, Lipschitz continuity, smoothness, black box model, oracle complexity, the impossibility result.
  - The Legendre dual function, Bregman distance function, the mirror descent algorithm.
  - The accelerated gradient descent method and lower complexity bounds.
- Stochastic optimization for statistics and machine learning (3 weeks)
  - The stochastic gradient descent method and the lower complexity bound.
  - The stochastic block coordinate descent algorithms.
  - The finite sum problem, the variance reduction technique.
- Smooth non-convex optimization (3 weeks)
  - The stationarity criterion. The gradient descent method.
  - The inexact proximal point method & the ADAM method.

    - AC-SA method for stochastic nonconvex smooth optimization and some zeroth order method.
- Non-smooth non-convex optimization (3 weeks)
    - The descent principle.
    - The Clark sub-differential. The weakly convex problem (DC problem).
    - The piecewise smooth problem.
    - The Goldstein sub-differential. The INGD method, its stochastic variants and lower complexity bound.
- Variational Inequality (3 weeks)
    - Min-max optimization, connection to games, the OGDA algorithm
    - The operator extrapolation algorithm and its application to policy evaluation in reinforcement learning.
    - The policy gradient method.

3. Pre-Requisite

The class is intended for 2nd or 3rd year PhD students who have some exposure to continuous optimization. There is no hard prerequisite for the class. However, the student is expected have some mathematical maturity (comfortable with analysis (at least on the level of baby Rudin), linear algebra and statistics, and have taken at least one proof-based class at the graduate level).

4. Course Outcome

By the end of the course, students should have developed some appreciation for the cutting edge of theoretical continuous optimization research and mastered the key technical tools to carry out independent research in the area.

5. Grading

The course grade is determined by assignments and the final presentation. There would be no exams.

(1) Assignment (10%): the goal is to reinforce key technical ideas which would be useful later in the class. So the workload is light (requires at most 2 hours per week in the first half of the course).
(2) Attendence(20%): you are expected to attend every class. If you have to miss some class, please email the instructor in advance.
(3) Final presentation (90 %): every student would give a 25 minutes presentation towards the end of the semester. It should be some interesting theoretical result (either from some published work or a student's own work) that broadly relates to the topics we discuss (stochastic optimization, non-convex optimization, non-smooth optimization, and variation inequality). The emphasis is on depth and clarity, rather than breath. The research problem needs to be well motivated (what was the bottleneck in previous results?), and one key technical innovation should be clearly illustrated (what's the key new insight, and how that could resolve the previous bottleneck? Please make enough simplications to "show rather than tell" the key new insight).

6. Instructor

Zhe (Jimmy) Zhang
email: zhan5111@purdue.edu
website: https://sites.google.com/view/jimmy-zhe-zhang/home
office: Grimssom 245