STAT 9910-303: Large-Scale Optimization for Data Science

Fall 2023

Instructor:	Yuxin Chen	**Time:**	Tue & Thu, 3:30pm - 5:00pm
Email:	yuxinc@wharton.upenn.edu	**Location:**	TBA

Course Description. This graduate-level course introduces optimization methods that are suitable for large-scale problems arising in data science and machine learning applications. We will explore several algorithms that are efficient for both smooth and nonsmooth problems, including gradient methods, proximal methods, ADMM, quasi-Newton methods, extragradient methods, as well as large-scale numerical linear algebra. We will also discuss the efficacy of these methods in concrete data science problems.

Instructor. Yuxin Chen (email: yuxinc@wharton.upenn.edu; office: 313 Academic Research Building). Office hours: by appointment.

Prerequisites. Students should have backgrounds in basic linear algebra and in basic probability (measure-theoretic probability is not needed), as well as knowledge of a programming language like MATLAB or Python to conduct simulation exercises. Students should also know the basic notion of convexity and optimization; a course such as STAT4810/5810 (Convex optimization for statistics and data science) would be beneficial.

Tentative Topics

- Week 1: gradient methods
- Week 2: subgradient methods
- Week 3: projected (sub)-gradient methods, Frank-Wolfe algorithm
- Week 4: mirror descent methods
- Week 5: block coordinate descent
- Week 6: proximal methods
- Week 7: Nesterov's accelerated methods
- Week 8: proximal dual methods, alternating direction methods of multiplier (ADMM)
- Week 9: Douglas-Rachford splitting
- Week 10: Quasi-Newton methods / BFGS
- Week 11: stochastic gradient descent (SGD), variance-reduced SGD
- Week 12: distributed optimization, saddle-escaping algorithms (e.g. cubic regularization, perturbed GD)
- Week 13: extragradient methods
- Week 14: TBD

Grading. The grade will be based heavily on the quality of the team project. The grading breakdown is as follows:

- *Homework problems (40%).* There will be 4 homework assignments. No late homeworks are accepted. You are encouraged to use LaTeX to typeset your homeworks.
- *Final project (60%).* See the description below.

Review sessions. There will be a few review sessions given at the beginning of the semester. The purpose of these sessions is to give students a quick review of basic linear algebra and probability. Attendance is optional.

Project. The term project can either be a literature review or include original research:

(i) *Literature review.* We will provide a list of related papers not covered in the lectures, and the literature review should involve in-depth summaries and exposition of one of these papers.

(ii) *Original research.* It can be either theoretic or experimental (ideally a mix of the two). If you choose this option, you can do it either individually or in groups of 2. You are encouraged to combine your current research with your term project.

There are 3 milestones / deliverables to help you through the process.

- *Proposal.* Submit a short report (no more than 1 page) stating the papers you plan to survey or the research problems that you plan to work on. Describe why they are important or interesting, and provide some appropriate references. If you elect to do original research, please do not propose an overly ambitious project that cannot be completed by the end of the semester, and do not be too lured by generality. Focus on the simplest scenarios that can capture the issues you'd like to address.

- *A written report.* You are expected to submit a final project report summarizing your findings / contributions. You must turn in a hard copy of your report to my office (you can slip it under the door of my office), as well as an electronic copy to my email for our records.

Collaboration policy. We encourage you to work on homework problems in study groups. However, you must write up and submit your own solutions and code without reading or copying the solutions of other students or other online resources.

Textbooks. We recommend the following two books, although we will not follow them closely.

- *Convex Optimization: Algorithms and Complexity*, Sebastien Bubeck, *Foundations and Trends in Optimization*, 2015.

- *First-Order Methods in Optimization*, Amir Beck, *MOS-SIAM Series on Optimization*, 2017.

- *Large-Scale Convex Optimization: Algorithms & Analyses via Monotone Operators*, Ernest K. Ryu, Wotao Yin, 2022.

- *Optimization for Data Analysis*, Stephen Wright, Benjamin Recht, 2022.

References. The following references also contain topics relevant to this course, and you might want to consult them.

- *Nonlinear programming (3rd Edition)*, Dimitri Bertsekas, *Athena scientific*, 2016.

- *Optimization methods for large-scale machine learning*, Léon Bottou, Frank Curtis, and Jorge Nocedal, https://arxiv.org/abs/1606.04838, 2016.

- *Proximal Algorithms*, Neal Parikh and Stephen Boyd, http://web.stanford.edu/~boyd/papers/pdf/prox_algs.pdf, *Foundations and Trends in Optimization*, 2013.

- *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, http://web.stanford.edu/~boyd/papers/admm_distr_stats.html, Stephen Boyd, Neal Parikh, Eric Chu, *Foundations and Trends in Machine Learning*, 2011.

- *Numerical Optimization*, Jorge Nocedal and Stephen Wright, *Springer*, 2006.

- *Numerical Linear Algebra*, Lloyd Trefethen and David Bau, Vol. 50, *SIAM*, 1997.

- *Convex Optimization*, Stephen Boyd, and Lieven Vandenberghe, http://stanford.edu/~boyd/cvxbook/, *Cambridge University Press*, 2004.

- *Lectures on Modern Convex Optimization*, Aharon Ben-Tal and Arkadi Nemirovski, Vol. 2, *SIAM*, 2001.