COMP 545: Advanced topics in optimization - From simple to complex ML systems

Anastasios Kyrillidis

Spring 2019

E-mail (instructor): `anastasios@rice.edu`

Web: **http://akyrillidis.github.io/comp545/**
E-mail (course): `RiceCOMP545@gmail.com`

Office Hours: By appointment
Office: DH 3119

Class Hours: TR 2:30pm - 3:45pm
Class Room: AEL B209

Course Description

Courses on machine learning (ML), artificial intelligence (AI), and signal processing (SP) abide by the following recipe: first, they introduce models that well-describe a task at hand; then, rigorously back up these models with intuition, and finally propose basic algorithms that solve such objectives for learning, inference and estimation. And in most cases, a gradient-based algorithm is the solution to our problems that saves the day!

Nevertheless, there are many papers that might have come to your attention, but go beyond plain gradient descent: papers on momentum and acceleration; papers with algorithms other than gradient descent, such as AdaGrad, Adam and RMSProp; research on variance-reduced techniques and hyper-parameter tuning such as learning rate, mini batch size and regularization; papers on gradient-variants that promote structures such as sparsity and low-rankness; literature that discusses the matter of "convex vs. non-convex optimization" and how gradient descent behaves on each case; papers that study the objective landscapes for saddle points and local minima; papers that study how easy it is to distribute gradient descent computations; and so on... You might wonder "*what is so different and significant in all these cases, since gradient descent is what is used in each case after all?*"

COMP XXX is a graduate-level course on optimization techniques and algorithms, as these are used in modern ML/AI/SP tasks. During this course, we will learn and study the above topics (both in depth and breadth). The course (*i*) will focus on different objective classes (convex vs. non-convex objectives, with constraints or not, etc.), (*ii*) will cover different optimization strategies within each class, (*iii*) will study algorithmic choices based on computational resources (*e.g.*, use of low-dimensional structures (when/why), asynchronous vs. synchronous algorithms, etc.) and (*iv*) lastly, will study schemes that handle some specific, but well-studied optimization constraints (sparsity, low-rankness).

The main objective of the course is to highlight optimization as a vital part of contemporary research in ML/AI/SP, and draw the attention of students to open-questions in related topics. In particular, the aim for students is to (*i*) learn how to distinguish differences in research papers of related fields, (*ii*) understand the connection between them and how researchers advance each area, and (*iii*) be able to consider possible extensions of these works, as part of the final (open-ended) project of the course.

Textbook

There is no textbook for the class. The class will be a collection of lectures, prepared by the instructor, as well as presentations of research papers. Links to resources will be provided during the course.

Prerequisites

Basics of calculus, linear algebra and basic knowledge of machine learning topics. Programming skills are not necessary, but might be required, depending on the project selected at the end of the course. In the latter case, programming in Python/Matlab could be sufficient.

During the first class, a quiz will be given to the students to assess their background; this quiz is going to be used for course content assessment (edits in the syllabus will possibly occur).

Course outcomes

After successful attendance, students are expected to:

(i) have a good understanding of the differences / difficulties of convex and non-convex optimization.

(ii) have a good comprehension how optimization plays a key role in different areas of ML/AI/SP.

(iii) have a first touch with various optimization-driven applications in ML/AI/SP.

(iv) be able to read and review advanced papers on similar subjects, as well as present the papers in front of an audience.

Registration / Communication / Attendance

Please send an email to the course email address to set up a time to meet and discuss your taking the course (after the quiz is preferable).

The instructor will be available for discussion after an appointment is set up; email communication is sufficient also, if preferable by the student. You are highly encouraged to attend and participate in class (see Grading and Evaluation), even if you are just auditing.

Course Format and Structure

There will be a traditionally formatted series of lectures. For each lecture, a different student (or set of students collaboratively) will scribe and take notes. During lectures, participation with questions / comments is encouraged.

For selected sessions, students will do presentations of papers (maybe in groups depending on the attendance), selected from a pile of papers that are related to the topics of the course and will be provided by the instructor.

Class Structure (tentative)

- *Overview of the course + Logistics.* Describe the main path to be followed.
- *Interlude #1*: Linear algebra and optimization background (mostly convex optimization).
- *What kind of convergence can we achieve?* Basics of convergence theory in convex optimization: zeroth-, first-, and second-order methods: gradient methods, L-BFGS, SR1, etc.
- *How can we make first-order methods faster?* Acceleration techniques, stochastic methods, variance-reduced methods.
- *What about hyper-parameters? (1/2)* Step size selection, mini-batch, momentum, regularization, initialization, etc, for simplified scenaria such as convex optimization.

- *Beyond convex optimization: Sparse recovery.* Applications in linear regression and graphical model selection, difference between convex and non-convex methods, non-convex algorithms, hyper-parameter tuning.
- *Interlude* #2: Concentration inequalities + RIP, restricted smoothness and strong convexity.
- *Beyond convex optimization: Low-rank recovery (1/2).* Applications in recommendation systems, quantum state tomography, etc., difference between convex and non-convex methods, non-convex algorithms, hyper-parameter tuning, matrix factorization machines.
- *Beyond convex optimization: Low-rank recovery (2/2).* Geometry of low-rank problems, local minima, constrained low-rank recovery, etc.
- *When data size makes optimization not easy to handle.* online vs. offline optimization, stochastic gradient descent, hyper-parameter tuning (2/2), ML & parallelism (synchronous vs. asynchronous optimization), HOGWILD!.
- *Non-convex empirical risk minimization:* Weakness of theoretical guarantees compared to convex optimization (driving example: neural networks), geometry (*e.g.*,saddle points and how to escape them), stochastic methods for generic non-convex methods and state of the art results.
- *Algorithms for deep learning:* stochastic variants (AdaGrad, ADAM, RMSProp), the marginal (?) value of adaptive methods.
- *(If time allows) Special topics in ML*: Implicit regularization, acceleration in non-convex settings, non-convex analysis of EM, inference using SGD, latest advances in theoretical guarantees of algorithms for deep learning.

The class will have intercalary sessions where *i*) students will present papers (2 sessions anticipated), *ii*) we will discuss preliminary descriptions of proposals for projects (around mid-way through the course), and *iii*) students will present the progress of their project towards the end of the class.

Grading Policy

The grade is based on the following factors:

- **5%** participation and attendance.
- **10%** scribing of notes.
- **70%** final project.
- **15%** paper reviews - students will submit a review almost each week.

The instructor reserves the right to curve the scale dependent on overall class scores at the end of the semester. The course has the format of "Satisfactory/Unsatisfactory", but the instructor will probably provide unofficially letter grades. Any curve will only ever make it easier to obtain a certain letter grade.

Scribing notes

Every week, a different student (or group of students) will volunteer (otherwise, will be randomly selected) to take notes and prepare a short –but consistent– note on the material presented each week. In case the attendance is low, this 10% of the final grade will be transferred to paper reviews (summing to 25%). A latex template will be shared by the instructor.

Reviews

Key outcome of this course is be able to read, comprehend and (healthy) criticize research papers. Further, as a bonus, reviewing could potentially lead to final project topic suggestions (after discussion with the instructor). A successful review should be concise (maximum 2 pages, single column, 11pt): beyond main comments, it should include the summary of the paper, discuss its importance, novelty, clarity, and reveal strengths/weaknesses.

Final project logistics

Students can team up (up to 3 members per group). *The goal of the project is to engage students to research related topics, even beyond the timeframe of the course*. I.e., there can be topics for a project that focus on simpler scenaria (say convex optimization), and topics that consider some harder non-convex questions. While the former could be potentially finished during the timeframe of the course, the latter could continue after the end of the course (this is the meaning of a *open-ended* project), and the instructor "bets" (and believes) on the self-motivation of the students to continue working on it, after the end of the course.

A project must include:

- The study of at least 3 research papers, on which the project is based on.
- The proposal of at least one "open" question: this includes the theoretical analysis of a specific scenario, or the implementation of a ML/AI system for some task in an interesting scenario, or a survey comparison of several algorithms on an interesting task.

The instructor will provide feedback to the students (by appointments + electronic communication). There will be a "midterm" 5-minute (tentative) pitch from each group (a session will be booked for this purpose - dates are tentative). After the discussion with the instructor, each group should prepare a four-paged description of the project with:

- Abstract and Introduction.
- Description of state of the art (summary and connection of the 3 papers selected, description of strengths and weaknesses and how these have led to the open question).
- Any preliminary results you have, and what is the plan from now on.

The project will culminate in a final project report of at least six pages, not including references, in NIPS/ICML format. At the end of the course, the group will prepare a 10-15 minute presentation, describing the background and the results they obtained. Final report dates will be available towards the end of the semester.

Course Policies

During Class

The electronic recording of notes will be important for class and so computers will be allowed in class. Please refrain from using computers for anything but activities related to the class. Drinking (coffee, tea, water) is allowed in class. Try not to eat your lunch in class as the classes are typically active.

Policies on Late Assignments

Assignments (scribing, reviews, project) should be turned on time. I don't like penalties but you will receive a 10% penalty for each day of delay. No submissions after a 2 day grace period. Exceptions will be given to special circumstances, with proper documentations.

Academic Integrity and Honesty

Students are required to comply with the university policy on academic integrity found in the Honor System Handbook http://honor.rice.edu/honor-system-handbook/.

Accommodations for Disabilities

If you have a documented disability that may affect academic performance, you should: 1) make sure this documentation is on file with Disability Resource Center (Allen Center, Room 111 / adarice@rice.edu / x5841) to determine the accommodations you need; and 2) meet with me to discuss your accommodation needs.