First-Order Methods in Optimization with Applications to Machine Learning

Credit: 7.5 ECTS

Course coordinator:
Alp Yurtsever, alp.yurtsever@umu.se
Dept. of Mathematics and Mathematical Statistics, Umeå University

Course Period:
May – August 2023

Main field of study and progress level:
Computational Science and Engineering/ Mathematics/ Mathematical Statistics, PhD

Prerequisites:
To succeed in this course, students should have a strong foundation in linear algebra, multivariable calculus, and probability theory. Ideally, students should have completed courses in these subjects at the second-cycle level or higher. Students who feel that they lack sufficient preparation in any of these areas are encouraged to review the relevant material before beginning the course.

Objective:
The objective of this course is to provide students with a solid understanding of first-order methods in optimization that are commonly used for solving large-scale problems that arise in machine learning and other data science applications. The expected learning outcomes of the course are as follows. After completing this course,

- Students should have a thorough understanding of the basic notions of optimization, including decision variables, objective function, constraints, convexity, strong convexity, smoothness, subgradients, proximal mappings, first-order optimality conditions, Lagrange duality, and strong duality.
- Students should have a thorough understanding of various first-order optimization algorithms, including gradient descent, accelerated gradient method, Frank-Wolfe method, subgradient method, proximal gradient method, stochastic gradient method, coordinate descent method, dual subgradient method, augmented Lagrangian method, and alternating direction method of multipliers.
- Students should be able to formulate optimization problems, such as least squares regression, matrix factorization, support vector machines, logistic regression classification, and neural networks.
- Students should be able to analyze and evaluate the performance of optimization techniques and algorithms in various data analysis scenarios and effectively communicate the results of their optimization analysis and experiments.

Contents:
First-order optimization methods are commonly used to solve large-scale problems in machine learning and other data science applications. These methods search for a solution iteratively by using gradient information (or sub-gradient information, if the loss function is non-differentiable) but not higher-order derivatives like the Hessian. Although they often have slow iterative progress and sublinear convergence rates, they offer a desirable trade-off in data science and machine learning

applications, where models often incorporate large amounts of data and high-dimensional optimization variables.

Motivated by the renewed interest in first-order optimization methods, this course will cover the following topics:

- Foundations of nonlinear optimization
- Gradient descent and accelerated gradient methods
- Stochastic gradient methods
- Coordinate descent methods
- Sub-gradient methods
- Projected gradient methods for constrained optimization
- Projection-free first-order methods for constrained optimization
- Proximal gradient methods and operator splitting
- Duality in constrained optimization

Form of instruction:
The course will be taught using a combination of self-study and exercises. Students are expected to read and comprehend the course material independently and can seek clarification or assistance from the instructor during office hours or online forums. The primary reading material for the course comprises sections 1-11 in [1], supplemented by sections 3, 4, 6, 7, and Appendix A in [2].

To reinforce learning, the instructor will provide two exercise sets: one at the end of section 6 in [1] (Coordinate Descent Methods), and the other at the end of section 10 in [1] (Proximal Gradient Methods and Operator Splitting). Although these exercises will not be graded, students are highly encouraged to work on them in groups. Some of the exercises may be included in the oral exam.

Examination:
The examination will be an oral exam, which will assess the understanding and mastery of the course material. At least 50% of the questions in the exam will be taken from the provided exercise sets. In addition, open discussion-type questions will be included to evaluate the students' ability to apply the concepts and techniques learned in the course.

Literature:
The primary course literature will be

[1] Wright, S. and Recht, B. *Optimization for Data Analysis*, Cambridge University Press, 2022
[2] Beck, A. *First-Order Methods in Optimization*, Society for Industrial and Applied Mathematics, 2017