# EN.553.662: Optimization for Data Science (Spring 2026)

(syllabus)

—----------------------------

**Instructor:** Nicolas Loizou (nloizou@jhu.edu)

**Lecture Times:** Monday, Wednesday 12:00 pm - 01:15 pm (from 01/20 to 27/04)

**Location:** Homewood Campus, Bloomberg 274

**Teaching Assistants:**

- Ben Weinberg (bweinbe5@jhu.edu ). Office Hours: Tuesday 3:00 pm - 5:00 pm
- Yuxin Ma (yma93@jh.edu ). Office Hours: Tuesday 10:30 am - 12:30 pm
- Qi Li (qli112@jh.edu). Office Hours: Thursday 3:00 pm - 5:00 pm

**Course Description :**

Optimization formulations and algorithms have long played a central role in data analysis and machine learning. In the era of big data, the need to solve large-scale optimization problems is ubiquitous in essentially all quantitative areas of human endeavor, including industry and science.
This course is a mathematically rigorous and comprehensive introduction to the field of large-scale optimization for data science and machine learning and is based on the latest results and insights.
We discuss the most important algorithms in the area, with an analysis of their convergence and complexity properties, as well as their practical implementations. Applications of the methods covered in the course can be found virtually in all fields of data science, including text analysis, page ranking, speech recognition, image classification, finance, and decision sciences.

**Prerequisites/corequisites:**

It is recommended that students have familiarity with fundamental concepts in

- linear algebra (abstract vector spaces, linear independence, basis, linear operators, quadratic forms, Euclidean spaces, inner product, norm, etc.),
- multivariate calculus (gradient, Hessian, Taylor approximation, chain rule, etc.),
- probability theory (probability spaces, expectation, law of large numbers, tower property of expectation, etc.) and
- Numerical programming will be required for this course, so familiarity with MATLAB, Python, Julia, or an equivalent will be necessary.

Some degree of mathematical maturity (ii.e., ability to comprehend and generate proofs) is also required. Coursework or background in optimization theory is highly recommended.

**Course Material:**

- Detailed slides (these will be handed out before each lecture)
- Textbooks and resources:
  - Optimization for Data Analysis, Stephen J. Wright, Benjamin Recht (Chapters 1-5)
  - First-order and Stochastic Optimization Methods for Machine Learning, Guanghui Lan (Chapters 1-6)
  - Advances and Open Problems in Federated Learning, Kairouz et al. (Chapters 1-3)

**Further useful textbooks and resources:**

- Convex optimization: Algorithms and complexity, Sébastien Bubeck
- First-order Methods in Optimization, Amir Beck
- Understanding Machine Learning: From Theory to Algorithms, S Shalev-Schwartz, S Ben-David

**Learning outcomes:**

The material of the course is designed for students wishing to enter the big data industry / MSc students in quantitative disciplines (e.g., optimization, informatics, data science, mathematics, operations research, machine learning, engineering).

With the completion of this course, the students will be able to:

- Analyze popular scalable nonlinear optimization algorithms appearing in ML and Data Science tasks.
- Execute high-performance computing (HPC) implementations of optimization algorithms to a range of selected applications.
- Have the ability to understand which optimization algorithm should be selected based on the structure of the problem of interest (convex or non-convex, smooth or non-smooth, many data points vs. high dimensions, problem fits into the memory of a single computer or not).

**Course Syllabus :**

**Part I: Stochastic Optimization**

1. Foundations of Large-Scale Optimization.
   - Introduction: Stochastic Optimization for Machine Learning
   - Introduction: Basic Tools from Convex Analysis, Optimization and Probability
   - Deterministic Gradient Descent

2. Stochastic Gradient Descent and its Popular Variants.
   - Basic Assumptions
   - Constant Step-size: Convergence guarantees for Strongly Convex, Convex, and Non-convex Problems

- Decreasing Step-size: Convergence guarantees for Strongly Convex, Convex and Non-convex Problems
- General Analysis Results: Mini-batch Size, Sampling Selection (uniform vs. importance Sampling)

3. Variance Reduced Methods (Fast Convergence to the exact solution).
    - SVRG
    - L-SVRG
    - SAGA
    - Unified Analysis of Stochastic Gradient Methods

4. Linear System Solvers:
    - Sketch and Project Methods
    - Randomized Kaczmarz method and Randomized Coordinate Descent (Gauss-Seidel Method)
    - Stochastic reformulations of Linear Systems and connections to SGD

5. Acceleration using Momentum:
    - Heavy Ball Method (Polyak's Momentum): Update rule and convergence guarantees
    - Nesterov's acceleration: Update rule and convergence guarantees
    - Momentum Variants in the Stochastic Setting

6. Adaptive/ Parameter-Free Optimization Algorithms:
    - Adagrad, RMSProp and Adam
    - Stochastic Polyak Step-size

**Part II: Variational Inequalities and min-max Optimization**

7. Background
    - Formulation and connections to optimization
    - Variational Inequalities, min-max optimization, smooth games, and saddle points
    - Applications in ML
    - Important Assumptions

8. Popular Algorithms: Convergence guarantees and practical applications
    - Gradient Descent Ascent and stochastic variants
    - Extragradient methods and stochastic variants
    - Hamiltonian Gradient Methods and Consensus Optimization

**Part III: Distributed/Decentralized/Federated Optimization**

9. Main Algorithms
    - Motivation & Applications
    - Distributed & Decentralized Methods

10. Rules for Improved Communication Complexity
    - Compressed Operators
    - Local Updates
    - Network Topology

**Evaluation:** Assignments (50%), Exams (50%)