Course: Distributed optimization

Course instructors: Dusan Jakovetic

Course type: elective

Credit points ECTS: 12

Prerequisites: -

Course objectives:

Introducing a wide range of modern optimization methods for large scale, parallel, and distributed optimization; their convergence analysis; and guidelines towards practical implementations.

Learning outcomes:

Outcomes:

- Ability and experience in applying the taught algorithms on real-world problems;
- Ability to apply the taught algorithms on research problems from a wide variety of application areas;
- Ability to synthesize and analyze efficient distributed algorithms for a given application.

Course description (outline):

Theoretical classes

Modern first-order methods for large-scale optimization: proximal gradient; accelerated Nesterov gradient; accelerated gradient for non-smooth optimization; Randomized methods: randomized coordinate gradient; stochastic/online gradient; Parallel and distributed methods: primal decomposition; dual decomposition; augmented Lagrangian; ADMM; average consensus methods; distributed gradient; distributed dual averaging; distributed approximate Newton; distributed primal-dual methods; analysis under various network settings (static/time-varying; undirected/directed); convergence and convergence rate analysis under various function settings (smooth/nonsmooth; convex/nonconvex; strongly convex/Lipschitz gradient); connection with training machine learning models; application examples of selected methods on real-world problems.

Practice classes

Implementation of the theoretically analysed methods.

References:

Main:

Selected papers in the field of distributed optimization

Textbooks (additional)

1. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Foundations and Trends in Machine Learning, Vol. 3, No. 1, pp. 1–122, 2011.
2. S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.
3. D. Bertsekas, Nonlinear Programming, Athena Scientific, 2004/

D. Bertsekas and J. Tsitsiklis, Parallel and Distributed Computation: Numerical Methods, Prentice-Hall, 1989

Active teaching hours: 5	Theoretical classes: 5	Practice classes: 0

Methods of teaching:

Theoretical lectures and independent work of students during practical hours.

Grading structure (100 points)

50 points on pre-exam and 50 points on oral exam