M 462: Theoretical Basics of Big Data Analytics and Real Time Computation Algorithms (Spring 2016)

Syllabus

Instructor: Professor Peter Golubtsov (visiting from Moscow State University, Russia).
Course Format: 3 lectures; one meeting per week: Th 2:10 PM-5:00 PM (+ online option).

Course Objectives: The main goal of this course is to provide students with a unique opportunity to acquire conceptual knowledge and theoretical background behind mathematical tools applicable to Big Data Analytics and Real Time Computations.

Course Content: The course will review specific challenges of Big Data Analytics, such as problems of extracting, unifying, updating, and merging information, and specific needs in processing data, which could be highly parallel and distributed. A number of mathematical tools for Big Data analytics, such as regression analysis, linear estimation, calibration problems, real time processing of incoming (potentially infinite) data, will be studied in more detail. It will be shown how these approaches can be transformed to conform to the Big Data demands. It will be discussed why most of the widely used algorithmic languages are not quite appropriate for solving Big Data problems and alternative approaches will be outlined.

Principal Topics:

1. The notion of canonical information: basic properties; extraction of canonical information from raw data and manipulations with canonical information.
2. Linear experiment and optimal estimation problem; canonical information for linear experiments; optimal estimation with prior information.
3. Manipulating information in different forms: raw vs. explicit vs. canonical; transforming one form into another.
4. Gauss-Markov theorem.
5. Calibration problem; canonical calibration information.
6. Real time signal processing with finite and infinite field of view.
7. Time series processing; balancing estimation accuracy, delay and computational demands.
8. Image processing with infinite field of view; parallel processing of signals and images.

Learning Outcomes: Upon completion of this course the students will be able to

- understand theoretical concepts lying at the core of Big Data Analytics algorithms;
- apply theoretical concepts they learned in class to solving practical Big Data problems;
- use elementary programming tools for practical implementations of theoretical ideas related to fining solutions of Big Data problems.

Prerequisites: M221 Introduction to Linear Algebra and two other MATH / STAT classes at 200-level or above; or consent of instructor.