RANDOMIZED ALGORITHMS FOR MATRICES

Computer Science

Spring 2021

Instructor: Professor Victor Pan

Course Rationale: Matrices are a popular way of modeling data, such as term-document data, social network data, machine learning kernels, and so on. The size of real-world data sets is very large and grows fast, becoming unfeasible for processing with classical deterministic algorithms. A common way out is to apply efficient randomized methods: they output approximate solutions that are accurate with a high probability. This course gives an introduction to such modern algorithmic techniques for efficient large-scale matrix computations.

Course Description: This course will cover the theory and practice of randomized algorithms for large-scale matrix problems arising in modern massive data analysis.

Prerequisites: General mathematical sophistication and a solid understanding of algorithms. The students will get the most of the course if they are familiar with linear algebra and probability theory at the advanced undergraduate level, although the instructor will cover the relevant material in these areas as necessary.

Topic List:

- Subspace embedding
- Approximating matrix multiplication
- Scalar and matrix concentration
- Random projections and Johnson-Lindenstrauss lemmas
- Least squares regression
- Least absolute deviations regression
- Additive-error low rank approximation
- Relative-error low rank approximation
- Distributed low rank approximation
- CUR approximation
- Tensor decomposition

Learning Objectives: Learn the mathematical theory of the methodology and algorithms of matrix processing techniques. Understand the numerical and computational issues that arise in implementing algorithms in different environments. Gain experience in applying randomized algorithms to large-scale data applications.

Assessment: Problem sets will be assigned on most weeks. Each student will present a critical review, summarizing the problems and solutions given in a selected research paper.

Grading policy: Homework – 50, Final project – 50