Randomized Algorithms for Big Data Matrices
Short Description

Course Web Page:	Monitor `http://drineas.org/` for updates
Instructor:	Prof. Petros Drineas
Required textbook:	We will cover material from review articles and research papers

Course Description

Matrices are a popular way to model data (e.g., term-document data, population genetics data, social network data, recommendation systems data, etc.), but the size, noise properties, and diversity of modern data presents serious challenges for many traditional deterministic matrix algorithms. The course will cover the theory and practice of **Rand**omized algorithms for **N**umerical **L**inear **A**lgebra problems (or **RandNLA** for short). Such algorithms are particularly suitable when dealing with large-scale matrix problems arising in modern, massive, data analytics. Topics to be covered include: underlying theory, including the Johnson-Lindenstrauss lemma, random sampling and projection algorithms, and connections between representative problems such as matrix multiplication, least-squares regression, low-rank matrix approximation, etc.; numerical and computational issues that arise in practice in implementing algorithms in different computational environments; as well as recent work that builds on the basic methods.

The course is appropriate for advanced undergraduate and graduate students in computer science, statistics, and mathematics, as well as computationally-inclined students from application domains.

Prerequisites and course requirements

Prerequisites: general mathematical sophistication; a solid understanding of Algorithms, Linear Algebra, and Probability Theory, at the advanced undergraduate or beginning graduate level, or equivalent.

Course requirements: three or four homeworks, one or more presentations, and a final project.

Representative topics

The following is a non-exhaustive list of representative topics that will be covered in the course.

- Randomized Matrix Multiplication: Building Blocks and Establishing Measure Concentration
- Random Projections: Slow, Fast, and Subspace
- Least-squares Regression: Sampling versus Projections, Low versus High Precision
- Low-rank Matrix Approximation: Additive-error, Relative-error, and Fewer Samples
- Element-wise Sampling and Applications
- Solving Systems of Linear Equations with Randomized Algorithms
- Extensions and Applications: Kernel-based Learning; Matrix Completion; Graph Sparsification; ℓ_p Regression and Convex Optimization; Parallel and Distributed Environments; etc.