Matrix Computations & Scientific Computing Seminar

Organizer(s): James Demmel & Ming Gu

Wednesday, 11:00am–12:00pm, 380 Soda

March 31 **Ken Clarkson**, IBM Almaden Research

Numerical Linear Algebra in the Streaming Model

We give near-optimal space bounds in the streaming model for linear algebra problems that include estimation of matrix products, linear regression, low-rank approximation, and approximation of matrix rank.

In the streaming model, we take one pass over the matrix entries, in an arbitrary order (that is not up to us). We wish to store a compressed version of the matrix, a *sketch*, and then use the sketch for the various problems.

For matrices A and B, our sketches are simply $S'A$ and $S'B$, where S is a random sign matrix, with each entry being $+1$ or -1 with equal probability. (Here A, B, and S all have the same number of rows, and S' denotes the transpose of S.) We sharpen prior quality guarantees for estimating $A'B$ as $(S'A)'S'B = A'SS'B$, and prove novel lower bounds for the number of bits needed for sketches such that such guarantees can hold.

We also show related results for best rank-k approximation; for example, that the sketches $S'A$ and AR, where R is another sign matrix, can be used to approximate A by a matrix $\widehat{A} = AR(S'AR)^{-}S'A$ of rank at most k/ϵ, so that with probability at least $1 - \delta$,

$$\|A - \widehat{A}\| \leq (1 + \epsilon)\|A - A_k\|$$

where A_k is the best approximation to A of rank k, $(S'AR)^{-}$ is the Moore-Penrose pseudo-inverse of $S'AR$. Here $\|.\|$ denotes the Frobenius norm, and the number of columns of S and R depend on ϵ, δ, and k.

This talk will focus on our upper bounds. This work was joint with David Woodruff.