Rapid and Reliable Randomized Algorithms

Communication Avoiding and Communication Hiding at Exascale

Ming Gu
Department of Mathematics
University of California at Berkeley
Berkeley, CA 94720

Abstract

As part of the CACHE effort, we have been in the process of developing superfast parallel direct solvers for large sparse linear systems of equations. One critical component of this effort is the rapid and reliable compression of dense matrices into lower ranked ones, and randomized sampling techniques seem to be an ideal technique for this purpose: they are very efficient and require minimum communication on high performance architectures.

In this talk, we report a number of surprising theoretical as well as numerical results on randomized algorithms. First of all, we develop a novel approach to analyzing randomized algorithms which has lead to error bounds that are very different and drastically sharper than those in the literature. Secondly, we have developed new randomized algorithms based on this analysis. The matrices which are of most relevance to our superfast direct solver are those with fast-decaying singular values. For such matrices, our randomized algorithms can rapidly compute highly accurate low-rank approximations at costs much lower than more traditional methods such as the Lanczos algorithm and subspace iteration methods for the same accuracy. This is in sharp contrast to just order-of-magnitude approximations typical of randomized algorithms. Thirdly, we establish that the traditional subspace iteration methods with a random start matrix can be viewed as randomized algorithms, and as such satisfy our error bounds. These bounds differ from traditional singular value error bounds for subspace iteration methods in subtle but important ways that speak strongly in support of a randomized start matrix. Finally, we present numerical experimental results that are in agreement with our conclusions.