Balancing Inexactness in Matrix Computations

Erin C. Carson[1]

[1] *Department of Numerical Mathematics, Faculty of Mathematics and Physics, Charles University, Czech Republic*
E-mail: carson@karlin.mff.cuni.cz

Abstract

On supercomputers that exist today, achieving even close to the peak performance is incredibly difficult if not impossible for many applications. Techniques designed to improve the performance of matrix computations - making computations less expensive by reorganizing an algorithm, making intentional approximations, and using lower precision - all introduce what we can generally call "inexactness". The questions to ask are then:

1. With all these various sources of inexactness involved, does a given algorithm still get close enough to the right answer?
2. Given a user constraint on required accuracy, how can we best exploit and balance different types of inexactness to improve performance?

Studying the combination of different sources of inexactness can thus reveal not only limitations, but also new opportunities for developing algorithms for matrix computations that are both fast and provably accurate. We present few recent results toward this goal, involving mixed precision randomized decompositions and mixed precision sparse approximate inverse preconditioners.

Acknowledgements: Work partially supported by ERC Starting Grant No. 101075632 and the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.