Memory Accessor for Mixed-precision GMRES-based iterative refinement

Patrick Amestoy[1], Antoine Jego[*2], Jean-Yves L'excellent[1], and Theo Mary[3]

[1]Mumps Technologies – Mumps Technologies – France
[2]LIP6 – Sorbonne Université – France
[3]LIP6 – Sorbonne Université – France

Abstract

Mixed-precision iterative refinement algorithms have been designed to provide high precision solution to well-conditioned problems while achieving high performance by relying on low precision factorizations. However both the factorization step and the iterative correction step of these algorithms may use multiple arithmetics. This mixture of precisions requires from the developper that they either convert the factorized system or use flexible iterative correction steps. Such modifications limit the problem scalability as the memory footprint would grow. Instead we propose to use mixed precision memory accessor approaches that decouple storage and compute precisions (data is stored and accessed in low precision, but computations are kept in higher precision) and reduce data accesses, improve accuracy, and simplify programming. In this work, we present experimental results of mixed-precision GMRES-based iterative refinement. We leverage the block low-rank structures and the mixed-precision storage of the sparse direct solver MUMPS to achieve low memory footprint. We also present the BLAS-based block memory accessor of MUMPS that is leveraged during the solve step to reach high performance. We discuss the importance of the memory accessor to reach better problem scalability compared with a flexible GMRES-based iterative refinement.

[*]Speaker