FASTMath Direct Solver Technologies

FASTMath Team Members: Jim Demmel (UC Berkeley), Sherry Li, Francois-Henry Rouet (Lawrence Berkeley National Laboratory)

We are developing scalable sparse direct linear solvers and effective preconditioners for the most challenging linear systems, which are often too difficult for iterative methods. Our focal effort is the development of three types of linear solvers: The first is a pure direct solver, encapsulated in SuperLU_DIST software. The second is a Schur complement-based hybrid solver, , encapsulated in PSDLin software. The third type is the nearly-optimal preconditioners using low-rank approximate factorization of the dense submatrices. We are also developing communication-avoiding linear algebra algorithms which have the potential to be used in the above sparse linear solvers.

SuperLU Communication and Multicore Optimization

Objectives

- Develop scalable sparse direct linear solvers that are essential for simulations of numerically challenging problems, e.g., accelerator, fusion, quantum chemistry, and fluid mechanics.
- Enhance performance of SuperLU with better utilization of new hardware resources, such as systems with many nodes and many cores per node.

Recent Accomplishments

- Implemented new static scheduling and flexible look-ahead algorithms that reduced the processors' idle time and shortened the length of the critical path. The parallel factorization is improved nearly 3x on thousands of cores.
- Introduced light-weight OpenMP threads in MPI processes. Significantly reduced memory footprint (up to 2x), and enabled use of all cores per NUMA node.
- In SuperLU_DIST 3.3 release, March 2013

Factorization time improvement of V3.0 over V2.5. Hopper at NERSC, Cray XE6

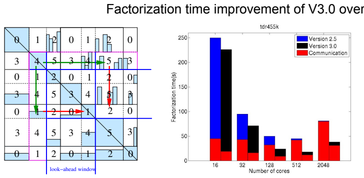


Fig.1. Accelerator Omega3P, dimension 2.7 M

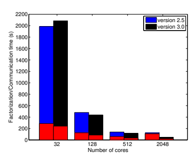


Fig.2. DNA, dimension 445 K

Impact

- There were over 24,000 downloads of SuperLU in FY2012. It is an indispensable kernel solver in DOE simulation codes, such as M3D-C1, NIMROD, and Omega3P. It is adopted in many commercial mathematical libraries and simulation software, including AMD (circuit simulation), Boeing (aircraft design), Chevron (geology), Cray's LibSci, FEMLAB, HP's MathLib, IMSL, NAG, OptimaNumerics, Python (SciPy), Walt Disney Feature Animation. Performance-enhanced SuperLU will be leveraged by large base of simulation codes.
- Explore general optimization techniques, challenged by irregular data access, high degree of task & data dependency (DAG), and complex communication patterns. Techniques are applicable to other irregular applications with similar DAG features.

Future Plans

- Aggregate blocks to enable use of multithreaded BLAS on multicore node and batch GEMMs on GPU
- More OpenMP enhancement for the remaining code

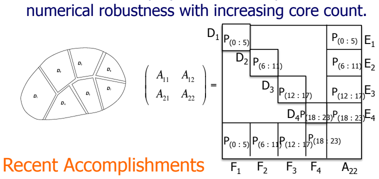
References

- I. Yamazaki and X.S. Li, "New Scheduling Strategies for a Parallel Right-looking Sparse LU Factorization Algorithm on Multicore Clusters", IPDPS 2012, Shanghai, May 21-25, 2012.

PDSLin Hybrid Solver

Objectives

- Develop a hybrid direct+iterative solver, Parallel Domain decomposition Schur complement based linear solver, which achieves better scalability than direct solver and more robust than iterative solver.
  - graph partitioning is used to obtain subdomains
  - subdomains are solved by a parallel direct solver
  - Schur complement system is solved by a preconditioned iteration solver. No need to form Schur complement explicitly
- It is crucial to employ two levels of parallelism to maintain scalability as well as numerical robustness with increasing core count.



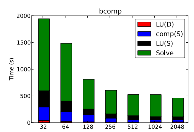


Fig.3. Fusion: magnetic reconnection Dimension 4.2 M

Recent Accomplishments

- Made the first release (version 1.0) in May 2012, has been stabilizing the code.
- Investigated some combinatorial algorithms to enhance PDSLin's performance. These have led to 30-60% reduction in runtime.
  - Used graph and hypergraph algorithms for the multi-constraint partitioning problem to balance the workload while computing the preconditioner in parallel.
  - Developed a new ordering algorithm to reorder the sparse right-hand side vectors to improve the data access locality during the parallel solution of a sparse triangular system with many right-hand sides.
- Applied PDSLin to the linear systems from the magnetic reconnection problem in the plasma fusion simulation and demonstrated that PDSLin is scalable to thousands of processors while maintaining the same robustness as a direct solver (see Fig. 3)

Future Plans

- Investigate various parallel preconditioners for the Schur complement systems
- Multicore enhancement

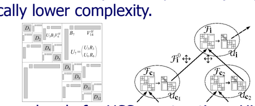
References

- I. Yamazaki and X.S. Li, "On techniques to improve robustness and scalability of the Schur complement method", VECPAR'10, June 22-25, 2010, Berkeley.
- X. Yuan, X.S. Li, I. Yamazaki, S.C. Jardin, A.E. Koniges, and D.E. Keyes, "Application of PDSLin to the magnetic reconnection problem", IOP Journal of Computational Science & Discovery, Vol. 6, No. 1, 2013.
- I. Yamazaki, X.S. Li, F.-H. Rouet, and B. Ucar, "On partitioning and reordering problems in a hierarchically parallel hybrid linear solver", PDSEC Workshop at IPDPS, 2013.

Parallel HSS Low Rank Factorization

Objectives

- Develop a new class of sparse approximate factorization algorithms that exploit the hidden data-sparseness via Hierarchically Semiseparable (HSS) representation, which has asymptotically lower complexity.



Recent Accomplishments

- Developed the first parallel algorithms and code for HSS construction, ULV HSS factorization, and HSS solution.
- Developed the first parallel HSS-structured sparse multifrontal code, demonstrated 2-3x faster than pure multifrontal for the Helmholtz equations on 16000+ cores.

References

- S. Wang, X.S. Li, J. Xia, and M.V. de Hoop, "Efficient algorithms for solving linear systems with hierarchically semiseparable matrices", SIAM SISC, Nov. 2012. (revised)
- S. Wang, X.S. Li, F.-H. Rouet, J. Xia, and M. de Hoop, "A Parallel Geometric Multifrontal Solver Using Hierarchically Semiseparable Structure", ACM TOMS, June 2013. (submitted)

Communication Avoiding Linear Algebra

Objectives

- Develop a new class of parallel factorization algorithms that have asymptotically lower communication complexity than conventional ones.

Recent Accomplishments

- New symmetric indefinite factorization LDL^T, based on blocked Aasen, achieved up to 2.8x speedup over MKL on 48 core AMD Opteron.
- LU factorization with panel rank revealing pivoting. Improve stability of partial pivoting while minimizing communication. As stable as GEPP in practice, more resistant to pathological cases.
- New rank revealing QR with column pivoting and minimize communication. Uses tournament pivoting (originally invented for TSLU) to select pivots.
- Develop arithmetic lower bounds for non-Strassen-like 1-sided factorizations, which in turn give communication lower bounds. New lower bounds apply to general sparse and structured matrices, not just dense.

References

- D. Becker, G. Ballard, J. Demmel, J. Dongarra, A. Druinsky, I. Peled, O. Schwartz, S. Toledo, I. Yamazaki, "Implementing a blocked Aasen's algorithm with a dynamic scheduler on multicore architectures", IPDPS'13. (best paper)
- A. Khabou, J. Demmel, L. Grigori, M. Gu, "LU factorization with panel rank-revealing pivoting and its communication avoiding version", to appear in SIMAX.
- J. Demmel, L. Grigori, M. Gu, H. Xiang, "Communication Avoiding Rank Revealing QR Factorization with Column Pivoting", submitted to SIMAX.
- J. Demmel, "An arithmetic complexity lower bound for computing rational functions, with applications to linear algebra", submitted to SIMAX.

More Information: http://www.fastmath-scidac.org or contact Lori Diachin, LLNL, diachin2@llnl.gov, 925-422-7130