Annotated Bibliography

Neil Lindquist

May 24, 2017

References

[1] Marc Baboulin, Alfredo Buttari, Jack Dongarra, Jakub Kurzak, Julie Langou, Julien Langou, Piotr Luszczek, and Stanimire Tomov. Accelerating scientific computations with mixed precision algorithms. *CoRR*, abs/0808.2794, 2008.

> The authors studied performance of iterative linear algebra solvers by doing the first passes with single precision floats, before using double precision to get the final awnser. They found that doing the first passes with single precision floats was significantly faster than using only double precision.

[2] Alfredo Buttari, Jack Dongarra, Jakub Kurzak, Piotr Luszczek, and Stanimir Tomov. Using mixed precision for sparse matrix computations to enhance the performance while achieving 64-bit accuracy. *ACM Trans. Math. Softw.*, 34(4):17:1–17:22, July 2008.

> The authors tried using both single and double precision in an iterative refinement algorithm to try to increase speed while retaining accuracy.

[3] Alfredo Buttari, Jack Dongarra, Julie Langou, Julien Langou, Piotr Luszczek, and Jakub Kurzak. Exploiting the performance of 32 bit floating point arithmetic in obtaining 64 bit accuracy (revisiting iterative refinement for linear systems). In *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, SC '06, New York, NY, USA, 2006. ACM.

> Single precision floats are able to be processed faster and the communication is much cheaper that working with double precision floats. The authors mixed single and double precision in each refinment step. They found that if factorization, forward substitution and backward substitution are do in single precision while the residual and update to the solution are done in double precision then the iterative refinement will produce the same accuracy than if double precision was used exclusively (as long as the matrix is "not too badly conditioned").

[4] Alfredo Buttari, Jack Dongarra, Julie Langou, Julien Langou, Piotr Luszczek, and Jakub Kurzak. Mixed precision iterative refinement techniques for the solution of dense linear systems. *Int. J. High Perform. Comput. Appl.*, 21(4):457–466, November 2007.

The authors compared using strictly double precision to using mixed single and double precision in iterative linear equation solvers with dense matrices. Except for small problems, the mixed precision solver was able to solve faster.

[5] Wei-Fan Chiang, Mark Baranowski, Ian Briggs, Alexey Solovyev, Ganesh Gopalakrishnan, and Zvonimir Rakamarić. Rigorous floating-point mixed-precision tuning. *SIGPLAN Not.*, 52(1):300–315, January 2017.

The authors address general improvements for mixing precision of floats in algorithms.

[6] James Demmel, Yozo Hida, William Kahan, Xiaoye S. Li, Sonil Mukherjee, and E. Jason Riedy. Error bounds from extra-precise iterative refinement. *ACM Trans. Math. Softw.*, 32(2):325–351, June 2006.

The authors address calculating the error bounds when the precision used to calculate the residual is higher than the precision used for the rest of the calculations in an iterative refinement algorithm.

[7] Nicholas J. Higham. *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1996.

Addresses the math supporting different numerical algorithms.

[8] J. D. Hogg and J. A. Scott. A fast and robust mixed-precision solver for the solution of sparse symmetric linear systems. *ACM Trans. Math. Softw.*, 37(2):17:1–17:24, April 2010.

The author anaylizes uses a mixed precision strategy utilizing fallbacks to more accurate algorithms to reach the user's requested precision. It starts with a single precision solve of $Ax = b$, then applies mixed-precision iterative refinement, then falls back to mixed-precision FGMRES, if target precision is not yet met then it uses a double precision solver on $Ax = b$ followed by double precision iterative refinement, then uses double-precision FGMRES. If target precision is not met a this point, then algorithm results in an error.