Proposal for a PhD. Thesis in Computer Science.

Subject: Randomization techniques in large scale matrix computations.

Scientific context and subject description: The advent of new processor architectures (e.g. multicore, GPUs) requires the rethinking of most of the numerical linear algebra algorithms which are at the heart of many scientific applications. As a result, innovative methods must be proposed in order to take full advantage of current supercomputers. Recent efforts have been made by the high-performance computing community in adapting numerical libraries in the area of linear systems, linear least-squares, or eigenvalue problems (e.g. PLASMA[1], MAGMA[2]). However, these libraries show some limitation on large numbers of processors or multiple GPUs when the volume of communication in the algorithm is significant.

A new class of algorithms, often referred to as "communication avoiding algorithms", has been recently studied [1] in order to minimize the number of messages exchanged during the algorithm. To go further in minimizing communication, we propose an approach based on randomization techniques that enable us to address larger problems with less scalability constraints. We presented recently in [2] performance results of a linear system solver in which random transformations are used to avoid pivoting in the Gaussian elimination. In these random transformations, we precondition the initial matrix with special random matrices [3] and, for sake of stability, we add some iterative refinement steps in the work precision.

The aim of this PhD thesis consists of generalizing this approach to other linear system solvers but also to extend it to applications that require large computing capabilities like error analysis in parameter estimation problems [4]. Applications are numerous in particular for dense and sparse linear systems and optimization problems (symmetric indefinite systems). The impact on scientific libraries would also be very important since this approach would provide very fast solvers with various physical applications. These solvers will be also used as kernels for iterative methods and we propose to validate our method on real applications coming from CEA.

The machines which are targeted in our simulations are Petascale supercomputers or large clusters of GPUs.

This PhD thesis will take part in a multidisciplinary collaboration between the researchers from CEA and INRIA Saclay. Moreover, for sake of visibility of the results, one of our aim is that the software developed during this thesis would be part of a reference public domain library. This is why it would also include a collaboration with the University of Tennessee (Innovative Computing Laboratory, contact: Jack Dongarra). The nature of this thesis which handles new mathematical results as well as innovative parallel algorithms should enable us to publish our results in both computer science conference proceedings and applied mathematics journals.

Advisor: Marc Baboulin, INRIA Saclay/University Paris-Sud (80%), in collaboration with Laura Grigori, HDR, INRIA Saclay (20%).
Contact: `marc.baboulin@inria.fr`

[1] Parallel Linear Algebra for Scalable Multi-core Architectures, `http://icl.cs.utk.edu/plasma/`
[2] Matrix Algebra on GPU and Multicore Architectures, `http://icl.cs.utk.edu/magma/`

References

[1] L. Grigori, J. Demmel, and H. Xiang, **Communication avoiding Gaussian elimination**, *Proceedings of the IEEE/ACM SuperComputing SC08 Conference*, November 2008.

[2] S. Tomov, J. Dongarra, M. Baboulin, **Towards dense linear algebra for hybrid GPU accelerated manycore systems**, *Parallel Computing*, Vol. 36, No 5&6, pp. 232-240 (2010).

[3] D. Parker, **Random butterfly transformations with applications in com putational linear algebra**, Technical Report CSD-950023, Computer Science Department, UCLA (1995).

[4] M. Arioli, M. Baboulin, S. Gratton, **A partial condition number for linear least-squares problems**, *SIAM J. Matrix Analysis and Appl.*, Vol. 29, No 2, pp. 413-433 (2007).