Scalable Hybrid Solvers for Large Sparse Linear Systems of Equations on Petascale Computing Architectures

S. X. Li, E. G. Ng (LBNL, Berkeley), L. Giraud, J. Roman (HiePACS, INRIA)

March 2011

Solving large sparse linear systems $Ax = b$, where A is a given matrix, b is a given vector, and x is an unknown vector to be computed, appears often in the inner-most loops of intensive simulation codes, and is consequently the most time-consuming computation in many large-scale computer simulations in science and engineering. Over the past decade or so, the French team (HiePACS INRIA) and the Berkeley team (Scientific Computing Group) involved in this France-Berkeley Fund project have been developing innovative numerical algorithms to exploit advanced high performance, large-scale parallel computers to solve these equations efficiently.

There are two basic approaches for solving linear systems of equations: sparse direct methods and iterative methods. Sparse direct solvers have been for years the methods of choice for solving linear systems of equations because of their reliable numerical behavior. However, it is nowadays admitted that such approaches are not scalable in terms of computational complexity or memory for large problems such as those arising from the discretization of large 3D partial differential equations (PDEs). There are on-going efforts in further improving existing parallel packages, but those efforts are mainly related to advanced software engineering. Although we will not contribute directly to such activities, we will use parallel sparse direct solvers as building blocks in the hybrid approaches we consider in this research project. Iterative methods, on the other hand, generate sequences of approximations to the solution. These methods have the advantage that the memory requirements are small. Also, they tend to be easier to be parallelized than direct methods. However, the main problem with this class of methods is the rate of convergence, which depends on the properties of the matrix. One way to improve the convergence rate is through preconditioning, which is another difficult problem.

Our approach to high-performance, scalable solution of large sparse linear systems in parallel scientific computing is to combine direct and iterative methods. Such a hybrid approach exploits the advantages of both direct and iterative methods. The iterative component allows us to use a small amount of memory and provides a natural way for parallelization. The direct part provides its favorable numerical properties. The general underlying ideas are not new. They have been used to design domain decomposition techniques for the numerical solution of PDEs. Domain decomposition refers to the splitting of the computational domain into sub-domains with or without overlap. The splitting strategies are generally governed by various constraints/objectives but the main one is to enhance parallelism. The numerical properties of the PDEs to be solved are usually extensively exploited at the continuous or discrete levels to design the numerical algorithms. Consequently, the resulting specialized technique will only work for the class of linear systems associated with the targeted PDEs.

In our work, we develop domain decomposition techniques for general unstructured linear systems. More precisely, we consider numerical techniques based on a non-overlapping decomposition of the graph associated with the sparse matrices. The vertex separator, constructed using graph partitioning, defines the interface variables that is solved iteratively using a Schur complement approach, while the variables associated with the interior sub-graphs are handled by a sparse direct solver. Although the Schur complement system is usually more tractable than the original problem by an iterative technique, preconditioning treatment is still required. For that purpose, we further study and compare the parallel preconditioners proposed by the teams, as well as design new parallel techniques. Linear systems with a few tens of millions unknowns have been solved on a few thousand of processors using the designed software prototypes. However, with the increasing demand for high resolution simulations, the size of the linear systems reaches hundreds of millions. In this joint research effort, we propose to collaboratively develop novel algorithms that will greatly improve our solvers scalability, especially in exploiting the emerging advances of petascale many-core computers.