The 8th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2022
5-9 June 2022, Oslo, Norway

A FACTORIZATION ALGORITHM FOR SPARSE MATRIX WITH MIXED PRECISION ARITHMETIC

Atsushi Suzuki[1]

[1] RIKEN Center for Computational Science, Minatojima-minami-machi, Chuo-ku, Kobe, 650-0047, Japan, Atsushi.Suzuki.aj@riken.jp

Keywords: *sparse direct solver, mixed precision arithmetic, high condition number*

There are two kinds of demand in finding a solution of linear system with large sparse matrix obtained from discretization of partial differential equations, by using mixed precision arithmetic. One is for solving the system with very high condition number in numerical simulation of complex physical model, e.g., the elasticity equation for mechanical problem for composite material, and the drift-diffusion equation for semi-conductor problem, where the variation of the coefficient of the matrix is very wide due to exponential weight to deal with the drift operator. In this case, it is required to use higher precision arithmetic than "double" precision, like "double-double" 128bit data and the linear system will be solved by a monolithic direct factorization solver. Since there is no dedicated hardware to perform quadruple precision arithmetic, using data type "double-double" is an efficient way to utilize the existing hardware units.

The other is for solving rather moderate problem mainly by lower precision arithmetic like single floating point. Factorization solver will be efficient as a local solver in preconditioner on modern architecture with combination of single and double floating point hardware units. Performing factorization of sparse matrix in single precision, whose coefficients are given in double precision, is recognized as perturbation to the finite element discretization of the bilinear form like approximate numerical quadrature. Therefore, a solution of the linear system by single precision could be a good initial guess for an iterative refinement procedure. However, we need to pay attention for the case a matrix is not invertible, which may happen for a stiffness matrix of a floating sub-problem of the elasticity problem. The zero energy mode consists of rigid body movements, which may be perturbed due to numerical truncation during conversion from "double" to "single" precision.

For both purposes, it is essential to distinguish either part of the algorithm that is allowed to be calculated by lower precision or one should be done by higher precision. Our strategy for factorization of the sparse matrix is to follow the elimination tree that is generated by nested-dissection ordering with modification to separate moderate part from difficult one by threshold pivoting [1]. In new implementation, the Schur complement is generated in higher precision by iterative refinement for linear systems with multiple right-hand side whose dimension is equal to the size of the Schur complement. This hybrid computation works better than global iterative refinement process with solutions by lower-precision factorization of the whole matrix.

REFERENCES

[1] Suzuki A. and Roux F.-X. A dissection solver with kernel detection for symmetric finite element matrices on shared memory computers, *Int. J. Num. Meth. Engng.*, Vol. **100**, pp. 136–164, 2014.