Out-of-core algorithms for dense matrix factorization on GPGPU

Numerical Linear Algebra Techniques for Scientific Computing

Ed D'Azevedo
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6367

Abstract

This work adapts out-of-core algorithms for dense matrix factorization for solving large problems on General Purpose Graphics Processing Unit (GPGPU). Dense matrix computations arises in diverse applications such as in modeling the response and heating of fusion plasma to radio frequency (RF) waves, modeling radiation heat transfer, boundary element method, and large scale linear least squares problems. The MAGMA library (`http://icl.cs.utk.edu/magma/`) achieves very high performance on GPGPU for dense matrix computations. However, the largest problem size is limited to the amount of local device memory on GPGPU. This work adapts out-of-core algorithms for solving large problems on GPGPU so that a matrix of size say 10 GBytes can still be factored on GPGPU with only 1 GBytes of device memory. The Cholesky factorization using a column-panel oriented left-looking algorithm with variable column panel width to take full advantage of available device memory. It achieves over 95% performance of MAGMA with the overhead for data movement accounting for about 5% of time.

	MAGMA 1.0	Out-of-core algorithm
N=25,000	266 Gflops/s	246 Gflops/s
N=35,000	Out of memory	263 Gflops/s

Table 1: Comparison of MAGMA 1.0 to out-of-core Cholesky factorization (DPOTRF) using only 1 GBytes out of 5 GBytes of Nvidia M2070.