GPU における 4 倍精度浮動小数点演算を用いたクリロフ部分空間法

Krylov subspace methods using quadruple precision floating-point arithmetic on GPUs

椋木大地[1]、高橋大介[2]

Daichi Mukunoki[1] and Daisuke Takahashi[2]

1) 筑波大学大学院システム情報工学研究科 2) 筑波大学システム情報系

1) Graduate School of Systems and Information Engineering, University of Tsukuba,
2) Faculty of Engineering, Information and Systems, University of Tsukuba,
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573

疎行列の連立一次方程式に対する反復解法として広く用いられているクリロフ部分空間法の収束性は浮動小数点演算の丸め誤差に影響されやすく，倍精度演算の代わりに 4 倍精度演算を用いて丸め誤差を減らすことにより，求解に必要な反復回数を削減できる場合がある[1]．一方で，GPU においては密行列ベクトル積などの演算に対して，ソフトウェア処理による 4 倍精度浮動小数点演算を行っても演算性能がメモリ律速となり，その実行時間は倍精度の場合の 2 倍程度となる場合がある[2]．クリロフ部分空間法は疎行列ベクトル積などのメモリインテンシブな演算で構成されているため，1 反復あたりの実行時間は，4 倍精度演算を行っても倍精度の高々2 倍程度で済む可能性がある．したがって，クリロフ部分空間法において 4 倍精度演算を用いることで反復回数が倍精度演算の半分以下となる場合には，倍精度演算の代わりに 4 倍精度演算を用いることで求解の高速化が期待できる．

本研究ではクロリフ部分空間法の一種である Conjugate Gradient（CG）法および Bi-Conjugate Gradient Stabilized（BiCGStab）法について，4 倍精度演算を用いた実装を GPU 上に行った．そして The University of Florida Sparse Matrix Collection から収集した疎行列を用いて，NVIDIA Tesla K20X GPU において倍精度版と性能を比較し，4 倍精度演算を用いることで求解が高速化できるケースを探した．また，NVIDIA cuSPARSE5.0 の倍精度 ILU(0)前処理を適用した場合についても検討を行った．

その結果，前処理なしの CG 法・BiCGStab 法においては，倍精度演算で求解可能な問題においても，4 倍精度演算を行うことで求解速度が高速化できるケースが存在した．これらのケースにおいて，4 倍精度版の 1 反復あたりの実行時間は倍精度版の約 1.4-1.6 倍であった．また，ILU(0)前処理を適用した場合においても，4 倍精度演算を用いることにより反復回数が削減され，求解時間を短縮できるケースが存在した．しかし前処理によって反復回数が削減される一方で，前処理に多くの実行時間を要したため，前処理なしの場合と比べて計算時間が増大する結果となった．本研究では反復回数を削減し求解速度を高速化する手段として，4 倍精度演算を用いる有効性について議論を行う．

[1] Hasegawa, H.: Utilizing the quadruple-precision floating-point arithmetic operation for the Krylov Subspace Methods, Proc. SIAM Conference on Applied Linear Algebra (LA03), 2003.
[2] 椋木大地，高橋大介：GPU における 3 倍・4 倍精度浮動小数点演算の実現と性能評価，情報処理学会論文誌 コンピューティングシステム（ACS41），Vol. 6，No. 1，pp. 66-77，2013.