様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科研費 KAKENHI

科学研究費助成事業　　研究成果報告書

平成　３０　年　　６　月　２０　日現在

機関番号：１２１０２
研究種目：基盤研究(C)（一般）
研究期間：2013～2017
課題番号：２５３３０１４１

研究課題名（和文）高精度演算の併用によるクリロフ部分空間法アルゴリズムの収束改善と高速化

研究課題名（英文）Improvement of Convergence and Performance for the Krylov Subspace Methods using High-Precision Arithmetic

研究代表者

長谷川　秀彦（HASEGAWA, Hidehiko）

筑波大学・図書館情報メディア系・教授

研究者番号：２０１６４８２４

交付決定額（研究期間全体）：（直接経費）　　3,800,000円

研究成果の概要（和文）：　倍精度演算の組み合わせで４倍精度を実現する Double-double演算と８倍精度を実現する Quad-double演算を用いてクリロフ部分空間法の収束改善を試みた。
　クリロフ部分空間法ではDouble-double演算の使用が効果的だったが、ランチョス法を用いた３重対角化では有効でなく、アルゴリズムによって高精度演算の効果が異なることが判明した。演算コストの高い高精度演算の使用をおさえた異なる精度によるリスタートを用いた混合精度演算では、問題によって有効性が大きく異なり、切り替えポイントの決定を含め、どれか一つの方法でよいという結論にはならなかった。

研究成果の概要（英文）：The Krylov Subspace Methods have some problems such as slow convergence rate or divergence, difficulty of parallelization of preconditioners, and rounding errors. The use of High-Precision arithmetic has a possibility to eliminate these problems, however it's costly.

We choose Double-double and Quad-double arithmetics as High-Precision arithmetic. The use of High-Precision arithmetic improved iterative process in Krylov Subspace methods, but did not improve a tridiagonalization process in Lanczos tridiagonalization method. From these results, the use of High-Precision arithmetic is effective for many algorithms, but not for all. To reduce computation cost, we combined three different arithmetic precisions such as double, Double-double, and Quad-double. There are many kind of combinations, and some of them were practically effective, but there was no method to fit for all test problems. As a software library we should resolve a problem how to relate combination methods and test problems.

研究分野：コンピュータサイエンス

キーワード：高精度演算　混合精度演算　Double-double演算　Quad-double演算　反復法　リスタート　疎行列　連立一次方程式

１．研究開始当初の背景

コンピュータを活用した数値シミュレーションの核は「離散化された連立一次方程式の解法」であり、そこでは大規模な疎行列を係数とする連立一次方程式に対して、クリロフ部分空間法と総称される反復法が用いられている。記述の精緻化に伴う問題の大規模化、あるいは問題の複雑化により、一般の倍精度演算を用いた反復法では収束させることが難しかったり、得られた解の誤差が大きかったりという問題に遭遇することが増えてきた。これまでは、前処理を工夫し、問題の条件を改善することで収束性の向上と誤差の軽減に対応してきたが、効果的な前処理は演算順序に依存性が強く、大規模な分散並列計算には向かない。いっぽうコンピュータの進歩は、演算については各段の性能向上が達成されているが、メモリアクセスに対してはさほど大きな性能向上はない。

そこで、コンピュータの演算性能の向上を高精度演算に活用することと、分散並列計算環境に適した演算依存性の少ない前処理を併用することで、丸め誤差の少ない反復プロセスを実現し、クリロフ部分空間法の収束性を向上させることが最終的な狙いである。高品質な解が、大規模な分散並列計算環境上で少ない反復回数によって得られるなら、数値シミュレーション全般が高速化できる。

２．研究の目的

演算精度を向上させれば、相対的に丸め誤差の影響が少なくなり、問題は良条件な方向に向かい、求解は容易になる。また、良条件の問題であれば、依存性の強い前処理を使う必要性も薄れ、並列化の容易な軽い前処理で済むようになる。最良のケースでは前処理が不要となり、前処理なしのクリロフ部分空間法で済む。前処理なしのクリロフ部分空間法では、内積演算と疎行列ベクトル積のみが並列化阻害要因となるが、アルゴリズムは汎用的で、良い前処理を探索して実装するのに比べれば微々たるコストである。

大きな研究の目的は以下のとおりである：

(1) まずは、各種のクリロフ部分空間法に対して高精度演算を適用し、収束までの反復回数と収束した解の品質にどのような影響があるかを調べる

(2) クリロフ部分空間法のアルゴリズムによって、また対象とする問題（係数行列と右辺）によって影響は異なるので、それらの関係性を明らかにする

(3) 高精度演算の演算コストが高いことは明らかであり、高精度演算の比率を下げ、一般の Double 演算を活用することが計算時間の観点では重要である。高精度演算の比率を減少させることによる収束性の悪化と１反復に要する計算時間の向上のトレードオフを検討する

使用する高精度演算は、特殊な付加ハードウェアを必要とせず、一般的なコンピュータで高速に実行でき、コンピュータの進歩が容易に享受できることが必要である。そこで、一般的な倍精度演算の組合せで４倍精度演算を実現する Double-double や８倍精度演算を実現する Quad-double を採用した。これらの四則演算には数十倍から数百倍の倍精度演算が必要になるが、メモリアクセスは２倍（４倍精度）か４倍（８倍精度）である。現代のコンピュータの演算性能とメモリアクセス性能の向上度を考えれば、メモリアクセスに対して演算が多いことは問題にならない。

研究の中心は、高精度化による問題の良条件化と、それに伴う収束性の改善であり、収束に要する反復回数の削減と収束した解の品質を重視する。言い換えれば、数値シミュレーションの要求精度が Double 演算の精度 10^{-10} であるところに、10^{-30} の精度を持つ Double-double 演算や 10^{-60} の精度を持つ Quad-double 演算を投入し、クリロフ部分空間法を用いてより少ない反復回数で質の良い解を得ることである。Double-double 演算や Quad-double 演算の実行時間はマシン環境や実装に強く依存するため、Double-double 演算や Quad-double 演算の計算時間は参考程度にとどめ、Double-double 演算や Quad-double 演算の並列化・高速化は別の研究テーマとして取り扱う。

詳細な評価をするまでもなく、Double-double 演算や Quad-double 演算が高コストであることは明らかである。そこで、反復回数は少ないが１反復あたりの計算時間がかかる高精度演算と、反復回数は多いが１反復あたりの計算時間が少ない一般の Double 演算のトレードオフを考える必要がある。特定の変数を高精度化し１反復内に Double 演算と高精度演算を混在させるのがよいのか、あるいは Double 演算の反復と高精度演算の反復を組み合わせるのがよいのかも検討する必要がある。反復計算の質を保ちながら、いかに演算コストを低減させるかも実用的な観点では重要である。

反復法では問題によって収束特性が異なるため、なんらかの方法で問題の性質を検出し、それに応じた対応をアルゴリズム中に組み込む必要がある。このような適応的なアルゴリズムの適切性を評価するには、テストデータセットを用いた大規模な数値実験が必要となる。このような実験は、どのような問題に対してどのアルゴリズムが有効かという分類問題にも発展する。

基本的スタンスは、倍精度演算と Double-double 演算や Quad-double 演算を組み合わせて、問題依存性の少ない現実的かつコストパフォーマンスのよいクリロフ部分空間法アルゴリズムの実現法を探索することにある。

３．研究の方法

まずは､実験に使える Double-double 演算､Quad-double 演算、標準的な Double 演算を混在させて使用できる実験環境が必要である。C などのプログラミング言語用ライブラリでは、コンパイルによって高速性が得られるが、データの型に強く依存したプログラムを書かなければならないことと、データ型の混在が難しいため、実行速度の面では不利だがインタプリタ型の計算環境を準備する。

また、数値シミュレーションで使われる連立一次方程式の多くは、係数行列が疎行列であり、より大きな問題を扱うために、疎構造も準備する必要がある。いっぽう、クリロフ部分空間法では、反復プロセス中で係数行列が更新されることはないため、係数行列を Double 精度で持つ。このことで、係数行列部分のメモリ容量の増大を押さえるとともに、高速化の際の FLOPS/BYTE を改善する。

実際、誤差・残差の観点から、どのような問題には、どれくらいの精度が必要かを数値実験により明らかにする：

1) 異なった演算精度で実行することにより、それぞれの問題の特性を明らかにする。同時に、問題のグループ分け、アルゴリズムとの相性などを探索する
2) 部分的なアルゴリズムの高精度化（混合精度演算の活用）によって、精度向上が達成できないかを調べる
3) 異なる演算精度による restart によって、精度向上が達成できないかを調べる。自動的な restart を行うための工夫とその効果を調べる
4) データ構造の改善、演算の並列化などによって、高速な実行を図る（収束への影響はないが、実行環境に強く依存する）。精度が不十分であっても、その演算が高速なら、多少の精度劣化を許容して性能向上が達成できる可能性がある
5) 同じような原理にもとづいた反復計算、たとえば固有値計算に対しても高精度化の効果があるのかを調査する

最終的に、問題のグループ分けとそれにふさわしい（混合精度演算を組み込んだ）クリロフ部分空間法アルゴリズムが提供できるはずである。コストパフォーマンスの観点から、単一の演算精度のプログラムでこのような目的を達成することは不可能であり、結果的に混合精度演算活用が必要になる。問題に合わせた自動的な演算精度コントロールがどのくらい可能かは、トライアンドエラーで調べるしかない。

研究成果は、１：BiCG 法に限定した混合精度演算と精度切り替え方法の評価、２:固有値計算における高精度化、３と４：テスト環境の構築と高速化である。まだ、全体像はまとめきれていない。

４．研究成果

１）安定性に優れている BiCG (Bi-Conjugate Gradient) 法を用いて Double 演算と Double-double 演算の組み合わせを評価した (Preconditionig 2017)。

All Double, 使用する変数の部分的な Double-double 化、反復の途中で Double から Double-double への切替え(DQ-SWITCH)、Full Double-double を ASIC_100ks, epb3, memplus, TSOPF_RS_b39_c7 などに（いずれも The University Florida Matrix Collection）適用した。部分的な Double-double の使用の改善効果はわずかで、問題によっては効果のない場合も観察された。Full Double-double は、きわめて安定な収束だったが、やはりコストが大きかった。異なる演算精度による restart 法である DQ-SWITCH は、Full Double-double より１割ほど反復回数が増加するが、Full Double-double と同様に安定な収束を示した。係数行列を Double 精度とし、Double-double を上位の変数配列と下位の変数配列に保持し、AVX2 を用いて高速化した環境では、Full Double-double の約半分の計算時間となった。今回のテスト問題では、ふつうの Double 演算よりも速く安定に解が得られている。

収束の状況をみて自動的に restart を決定する方法では、最適なタイミングが得られなかった。DQ-SWITCH はきわめて有効かつシンプルなアルゴリズムであるが、現時点では自動的に切り替えるタイミングは見つけられていない。

２）固有値問題への有効性を確認するため、Double-double と Quad-double を対称行列の３重対角化アルゴリズムに適用して有効性を評価した（EPASA2015：International Workshop on Eigenvalue Problems: Algorithms, Software and Applications in Petascale Computing)。

３重対角化アルゴリズムとしてはハウスホルダー３重対角化、Lanczos 法、Arnoldi 法、３重対角行列の固有値計算には QR 法を用いた。固有ベクトルは計算していない。３重対角化と固有値計算を Double, Double-double, Quad-double を用いて実行し、厳密計算で得られた約 200 桁の固有値との精度比較を行った。QR 法を用いた計算では、Double でも十分な精度が得られ､対称行列に対する固有値問題では、精度劣化が発生するのは直交変換を用いた３重対角化であり、得られた３重対角行列に対する固有値計算は誤差の影響を受けにくいことが明らかになった。固有値計算が誤差の影響を受けにくいのであれば、３重対角行列からの固有ベクトル計算も何らかの方法で精度よく実行できるはずである（つまり、前段の３重対角化とは独立に考えられる）。いっぽう、３重対角化部分のふるまいはアルゴリズムと演算精度によって大きく異なり、最終的な固有値の

品質を決定していた。丸め誤差に敏感と言われている Lanczos 法は Double-double, Quad-double 程度の高精度演算では精度が改善せず、Arnoldi 法は易しい問題にのみ Double-double, Quad-double が有効だった。
　連立一次方程式に対するクリロフ部分空間法は高精度化が有効であり、また３重対角化アルゴリズムと比べて安定したアルゴリズムであることが明らかになった。固有値計算も重要なアプリケーションで、アルゴリズムの改良は必要だが、アルゴリズム全体の単純な高精度化ではうまくいかず、高精度化が有効に働く別アプローチが必要である。

３）MATLAB 上に MuPAT (Multiple Precision Arithmetic Toolbox) を実装し、並列処理を用いた高速化を進めている（第 16 回情報科学技術フォーラム講演）。
　これまで Scilab 上に実装していた MuPAT を MATLAB 上に実装し、基本的な機能は MATLAB からも使えるようになった。これにより、ツールボックスと高精度演算を用いたクリロフ部分空間法アルゴリズムを渡せば、だれもが手元のマシンで効果をテストできる。より多くの問題に対してテストを実行してもらうことにより、より汎用的でロバストなコードの開発が可能になる。現在、シングルプロセッサを対象にした高速化 (FMA, SIMD, OpenMP) が進行中であり、高速版はより現実的なスピードにできるはずである。
　なお、ツールボックスは公開準備中、自動切り替え版の DQ-SWITCH のコードは未完成である。

４）数値シミュレーションで用いられる行列は大規模な疎行列であり、メモリ容量と演算数の節約のためには疎データ構造の活用が必須である。Double 演算、Double-double 演算、Quad-double 演算の混合使用を前提に、Scilab 上の MuPAT (Multiple Precision Arithmetic Toolbox) に疎データ構造を実装し、その有効性を評価した (PPAM2013:Tenth International Conference on Parallel Processing and Applied Mathematics)。
　３種類の数値型と演算精度に対して、それぞれ密構造と疎構造の併存を許したツールボックスを開発した。システム作成上、多くの組み合わせに対応したコードを準備する必要があるという問題点はあるが、ユーザからは演算精度とデータ構造を意識せずに MuPAT が使用できる。たとえば、四則演算 +-*/ はすべてのデータ型とデータ構造で共通に使えるので、演算精度やデータ構造の詳細を意識せずにアルゴリズム開発が行えるようになっている。
　特別な高速化は適用していないが、メモリ容量と演算回数の節約により、疎行列の非ゼロ要素数に対応した高速化が達成されている。人的エラーの削減、システム資源の節約の観点で、有効なアルゴリズムの開発環境である。

全体を通じて、以下のことが達成できた：

・Double 精度、 Double-double 精度、Quad-double 精度を混在して使える計算環境を構築し、そこでの高速演算を可能にした

・混合精度アルゴリズムの導入が有効な問題はみつけた

・BiCG 法に有効な混合精度演算の導入方法をみつけた。異なる演算精度による restart アルゴリズムとして一般化できるはずである

・演算精度の自動選択が可能になるほど、問題と精度の関係は明らかになっていない。演算精度やアルゴリズム選択に対する Automatic Tuning は今後の課題である

・固有値問題、対称行列の３重対角化に対して、興味深い結果が得られた

　コンピュータの高性能化により、複数の演算精度を同時かつ使い分けて利用できるようになり、アルゴリズム開発に対して演算精度選択という自由度が付与された。今後、計算パワーを、計算の質の向上に活用すべきだと考えている。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　３件）
① Kohei Asami, Ryoya Ino, <u>Emiko Ishiwata</u>, and <u>Hidehiko Hasegawa</u>. Comparison of Tridiagonalization Methods using High Precision Arithmetic with MuPAT on Scilab, Eigenvalue Problems: Algorithms, Software and and Applications in Petascale Computing, Lecture Notes in Computational Science and Engineering 117, pp. 125-141, Springer, 2017 at EPASA 2015: International Workshop on Eigenvalue Problems: Algorithms, Software and Applications in Petascale Computing, September 2015, Tsukuba, Japan, DOI: 10.1007/978-3-319-62426-6_9.（査読あり）
② T. Saito, S. Kikkawa, <u>E. Ishiwata</u>, and <u>H. Hasegawa</u>, Effectiveness of sparse data structure for double-double and quad-double arithmetics, Lecture Notes in Computer Science 8384, pp. 643-651, Springer, 2014 at the Tenth International Conference on Parallel Processing and Applied Mathematics (PPAM 2013), Part 1, Sep. 8-11, 2013, Warsaw, Poland, DOI:10.1007/978-3-642-55224-3_60.（査読

あり）
③吉川 慧子，斉藤 翼， 石渡 恵美子， 長谷川 秀彦． Scilab における高精度演算環境 MuPAT の実装， 図書館情報メディア研究, Vol. 11, No. 1, pp. 23-46 (2013)（査読あり）

〔学会発表〕（計 5件）
① 長谷川 秀彦， 椎葉 健， 石渡 恵美子． MATLAB 上での高精度演算の実装について， 第 16 回情報科学技術フォーラム講演論文集 第１分冊， 東京， 2017-9，東京大学，2017, B-007
② H. Hasegawa and T. Hishinuma. Robust and Fast BiCG Method using SIMD-Accelerated DD Arithmetic, The International Conference on Preconditioning Techniques for Scientific and Industrial Applications (Preconditioning 2017), Poster Presentation, Vancouver, BC, Canada, July 31 - August 2, 2017.
③ Kohei Asami, Ryoya Ino, Emiko Ishiwata, and Hidehiko Hasegawa. Eigenvalue computation using High Precision Arithmetic Toolbox MuPAT on Scilab, International Workshop on Eigenvalue Problems: Algorithms; Software and Applications, in Petascale Computing, Poster Presentation, Tsukuba, Japan, September 14-16, 2015
④ 長谷川 秀彦． 高精度演算を用いた混合精度反復法， 日本応用数理学会 三部会連携「応用数理セミナー」,東京， 2013.12.27.，東京大学本郷キャンパス，2013
⑤ T. Saito, S. Kikkawa, E. Ishiwata, and H. Hasegawa, Effectiveness of sparse data structure for double-double and quad-double arithmetics, Tenth International Conference on Parallel Processing and Applied Mathematics (PPAM 2013), Warsaw, Poland, Sep. 8-11, 2013

〔図書〕（計 0件）

〔産業財産権〕

○出願状況（計 0件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計 0件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

６．研究組織
(1)研究代表者
長谷川 秀彦（HASEGAWA, Hidehiko）
筑波大学・図書館情報メディア系・教授
研究者番号：２０１６４８２４

(2)研究分担者
石渡 恵美子（ISHIWATA, Emiko）
東京理科大学・理学部第一部応用数学科・教授
研究者番号：３０２８７９５８