

平成 30 年 6 月 14 日現在

機関番号：13901

研究種目：挑戦的萌芽研究

研究期間：2016～2017

課題番号：16K12432

研究課題名(和文) 精度保証のための高性能計算技術の創成

研究課題名(英文) Development of Technologies of High Performance Computing for Accuracy Assurance

研究代表者

片桐 孝洋 (KATAGIRI, TAKAHIRO)

名古屋大学・情報基盤センター・教授

研究者番号：40345434

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：高精度行列-行列積(HP_GEMM)の新実装方式の開発と性能評価を行なった。HP_GEMMを実現するスレッド並列化において疎行列化を行い演算量を削減する高性能実装方式を開発した。既存のスーパーコンピュータで複数の方式の実装評価を行った。これらの方式は、疎行列格納方式CRS、ELL、密行列積DGEMM方式、および、疎行列-ベクトル積の実装を問題特性を利用しスレッド実行効率を高める。性能評価により各実装方式の性能を明らかにした。

HP_GEMMの精度面について理論検討・実装・評価を行った。特に、行列積のエラーフリー変換を用いて隣接浮動小数点丸めを達成するアルゴリズムを開発した。

研究成果の概要(英文)：We establish new implementation methods and performance evaluation for high-precision matrix-matrix multiplication (HP_GEMM). A high-performance thread parallelization method for HP_GEMM is developed by sparselization to reduce computational complexities. In addition, we evaluate several implementation methods with an existing supercomputer. The implementations are utilizing with sparse storage formats, such as CRS and ELL, and dense matrix-matrix multiplication DGEMM, and implementations of sparse matrix-vector multiplications to increase efficiency of threads execution with problem-level parallelisms. Performance of thread execution of the implementations is clarified by the performance evaluation.

We discuss theory, implementation, and evaluation for accuracy of HP_GEMM. In particular, we develop an algorithm to establish rounding for the immediate floating-point neighbors in error-free transformation of the matrix-matrix computations.

研究分野：高性能計算

キーワード：高精度行列-行列積 分散並列化 反復改良 ベンチマーク 精度保証 スレッド並列化 疎行列化 PB LAS

1. 研究開始当初の背景

(1) 行列-行列積に代表される基本線形計算を集約したライブラリ、たとえば、BLAS (Basic Linear Algebra Subprograms) は、多くの線形計算で必須の処理である。

ところが、従来の数値計算では演算速度は考慮しているが、計算結果の正確性の考慮がなく、解の精度保証が重要な課題となっている。

一方で、BLAS を用いた汎用的な数値計算ライブラリの LAPACK などにおいて、精度保証をする研究や実装は少ない。

(2) この状況の中、BLAS を精度保証する研究が早稲田大学の 大石教授のグループにより進められており、実用的な成果が出てきている。

この一方で、大規模数値計算が行われているスーパーコンピュータで高性能実装を行う HPC 分野においては、BLAS の精度保証アルゴリズムを実装して評価する研究はまれであり、理論と実用の研究が乖離している。

(3) HPC 分野では現在、2020 年ごろ実現可能なエクサ級スパコン (次世代スパコン) に向けた計算機開発が行われている。次世代スパコン環境では問題の大規模化により、丸め誤差が蓄積し計算結果の信頼性が著しく下がることが懸念されている。

大規模計算の結果が不正確であれば、資源/時間/費用の無駄となり、大きな損害をうける。

そのため、精度保証アルゴリズムの高性能実装を行うことが緊急の課題となっている。

2. 研究の目的

(1) 本研究は早稲田大学大石グループで開発された高精度行列-行列演算 (尾崎の方法) の高精度行列-行列積 (HP_GEMM) アルゴリズムを基本とし、その並列化を中心する以下の原理を明らかにすることを目的とする。

I. HP_GEMM の高性能実装の原理: 「疎行列 - 密行列積」や「疎行列 - 疎行列」の実装方式

II. ハイブリッド MPI/OpenMP による高性能化の原理: 効率の良いプロセス/スレッド並列化方式

III. MPI 通信の最適化の原理: HP_GEMM 並列化時の効率の良い MPI 実装方式の仕組み

3. 研究の方法

(1) 平成 28 年度では主に、HP_GEMM 専用アルゴリズム開発と OpenMP 実装などの最適化方式の開発を行った。

片桐 (代表者) は、HP_GEMM における OpenMP 実装方式である逐次 DGEMM を用いた問題レベルの並列化、および、疎行列化による SpMV の実装の性能評価を行い、HP_GEMM 専用演算

化 (複数右辺の SpMV 演算) における高性能実装アルゴリズムを開発する。

(2) 尾崎 (分担者) は、尾崎の方法の開発者であるため、HP_GEMM アルゴリズムに熟知している。そのため、片桐と連携することで、高性能かつ独創性の高い実装への展開を行う。

(3) 荻田 (分担者) は、精度保証された内積演算で著名であり、精度保証アルゴリズムとベンチマーク化の観点で貢献を行う。

4. 研究成果

(1) 尾崎の方法は、入力行列に対して以下の式 (i) の無誤差変換を行う。

I) 行列 A と行列 B を下記のように分解する (インデックスが若いほうが高いビットを持つようにする)

$$\begin{aligned} A &= A(1)+A(2)+A(3)+\dots+A(p) \\ B &= B(1)+B(2)+B(3)+\dots+B(q) \quad \dots(i) \end{aligned}$$

II) 行列積 AB を以下のように計算する (p × q 個の行列積となる)

$$\begin{aligned} AB &= \\ &(A(1)+A(2)+\dots+A(p))(B(1)+B(2)+\dots+B(q)) \\ &= (A(1)+A(2)+\dots+A(p))(B(1)+B(2)+\dots+B(q)) \\ &= A(1) B(1)+ A(1) B(2)+ A(2) B(1)+\dots+A(p) B(q) \quad \dots(ii) \end{aligned}$$

ここで、式(i)の分解された行列積どうしの加算には、高精度な和を行う演算で加算される。

(2) 処理 I) の分解の仕方を工夫することで、処理 II) における行列積を無誤差の演算にすることができる。行列サイズが大きくなるに従い、ほとんどの演算時間は行列積部分の時間となる。

また、分解数である p、q を限定すれば、部分的な多倍長化演算と同様の効果を得ることができる。したがって、演算時間を考慮して高精度化を図ることが可能である。

(3) 本研究における尾崎の方法における OpenMP を用いたスレッド実装方式の開発により、従来方式に対して高速となる実装方式を開発した。

(4) 本研究では、以下の 11 種類の実装方式の開発に貢献した。

- i. dgemm による実装 (高精度ではない演算)
- ii. 尾崎の方法で dgemm を用いる実装 (従来の実装方式)

iii. 尾崎の方法において疎度 90%以上で CRS 形式による疎行列化を行う方式（並列化方式はループ内部）。

iv. 尾崎の方法において疎度 90%以上で CRS 形式による疎行列化を行う方式（並列化方式はループ外部）

v. 尾崎の方法において疎度 90%以上で CRS 形式による疎行列化を行う方式（並列化方式は、ループ内部で問題レベルの並列性を使う方法）

vi. 尾崎の方法において疎度 90%以上で CRS 形式による疎行列化を行う方式（並列化方式は、ループ内部で問題レベルの並列性を使う方法で、ベクトル数 100 の並列性を使う方法）。

vii. 尾崎の方法において疎度 90%以上で ELL 形式による疎行列化を行う方式（並列化方式はループ内部）。

viii. 尾崎の方法において疎度 90%以上で ELL 形式による疎行列化を行う方式（並列化方式はループ外部）

ix. 尾崎の方法において疎度 90%以上で ELL 形式による疎行列化を行う方式（並列化方式は、ループ内部で問題レベルの並列性を使う方法）

x. 尾崎の方法において疎度 90%以上で ELL 形式による疎行列化を行う方式（並列化方式は、ループ内部で問題レベルの並列性を使う方法で、ベクトル数 100 の並列性を使う方法）。

xi. 尾崎の方法を使わず、内積演算による高精度和の方式を利用して演算する方式。

(5) テスト行列として、行列 A を単位行列とある疎度分の要素に対して 0~1 の範囲で生成する。B は A の逆行列を用いる。

(6) 行列 A と行列 B に関する無誤差変換の分解数（それぞれ、 A_k および B_k ）について、そのうち疎行列化された数（ S_{pm} ）は以下の通りであった。

- $N=1000$: $A_k=2$, $B_k=3$, $S_{pm} = 2$

(7) この評価では以下の計算機を利用した。

- Fujitsu PRIMEHPC FX100 (FX100)
- 名古屋大学情報基盤センター設置
- CPU: SPARC64 XIfx, 2.2 GHz 32(+2) コア
- 記憶容量: 32 G、理論ピーク性能（ノード）: 1.1264 TFLOPS (倍精度), 2.2528 TFLOPS (単精度)
- キャッシュ構成
 - L1: 64KB (命令/データ分離, コア毎), L2: 24MB (共有)
 - 4 ウェイ
 - 1 ソケット当たり 16 コア, ノードあたり 2 ソケットの NUMA 構成
 - 富士通 MPI
 - コンパイラ: Fujitsu C/C++ Compiler Driver Version 2.0.0 P-id: T01776-01 (Jun 22 2016 14:52:00)

- コンパイラオプション:
 - ◇ 疎行列カーネル部分: -Kfast -Kopenmp
 - ◇ それ以外: -O0 -Kopenmp
- メモリアクセス性能 (node あたり): 240 GB/秒 (入力/出力ごと)

(8) 尾崎の方法は C のコードで記述されたものを利用し、FX100 に任意精度である MPFR ライブラリを導入してこれを真値として相対誤差を調べた。評価は 1 ノード 32 スレッドに固定した。富士通ライブラリの FX100 向け BLAS (スレッド並列版、および逐次版の双方) を BLAS 演算部分には利用している。

(9) 表 1 に、性能評価の結果を示す。

表 1 入力行列に対して従来実装 2 を 1 とした場合の速度向上率 (カーネル時間のみ)

実装方式	速度向上率
ii	1.00
iii	4.84
iv	22.70
v	14.47
vi	8.27
vii	3.78
viii	6.74
ix	3.76

表 1 の結果は、このテスト行列の場合、従来での尾崎の方法の実装方式 2 に対し、実装方式 4 が、カーネル時間の観点で約 22 倍も高速化されている。

このように、速度高速化の観点での有効性を示すことができた。

(10) 平成 28 年度には、精度保証付き数値計算の観点に基づくベンチマークを設計するため、線形問題について、高性能計算環境に適用可能な高精度計算法を開発した。

この方法は、反復改良法をベースとしており、問題の難しさに応じて適応的に計算精度を変化させ、最終的な結果精度を達成するためのものである。

また、計算精度を変化させて線形計算を実行するためには、行列ベクトル積、ベクトルの内積の計算について高精度化が必要であり、これを高性能計算環境で達成するためのアルゴリズムを内積計算ベースで試作し、次年度に向けての準備を整えることができた。

(11) 平成 29 年度は、高精度な行列・行列積について、目標となる精度を設定し、それを厳密に達成するアルゴリズムについて検討を行った。

提案する手法は PBLAS、特に PDGEMM を大きく依存するために、高速・高並列化効率が期待される。また、倍精度よりも少々高い精

度に特化された高精度計算法のプログラムを分散メモリ環境用に試作した。

(12)平成 29 年度は、高精度行列積の精度面についての理論・実装・評価を行った。特に、行列積のエラーフリー変換を用いて隣接浮動小数点丸めを達成するアルゴリズムの開発を行った。

隣接浮動小数点丸めとは、真値に隣り合う浮動小数点数のどちらかを返す丸めであり、非常に高精度なことが知られている。

この精度保証法および性能評価を行った。問題が良条件な場合、近似計算の 7 倍程度の時間で隣接浮動小数点丸めが達成されることを数値実験により示した。

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 Shuntaro Ichimura, Takahiro Katagiri, Katsuhisa Ozaki, Takeshi Ogita, Toru Nagai, Threaded Accurate Matrix-Matrix Multiplications with Sparse Matrix-Vector Multiplications, Proceedings of IEEE IPDPSW2018, 2018, 1-10
- ② 荻田 武史, 尾崎 克久, 柏木 雅英, 片桐 孝洋, HPC 分野における精度保証付き数値計算学の展開、研究報告ハイパフォーマンスコンピューティング (HPC)、2017-HPC-160、2017、1-5
- ③ 市村 駿太郎, 片桐 孝洋, 尾崎 克久, 荻田 武史, 永井 亨, 荻野 正雄, マルチコア計算機による高精度行列 - 行列積アルゴリズムの性能評価、研究報告ハイパフォーマンスコンピューティング (HPC)、2017-HPC-160、2017、1-8

[学会発表] (計 10 件)

- ① Takahiro Katagiri, Shuntaro Ichimura and Kenya Yamada, High Precision Computing of Matrix-Matrix Multiplications and a New Approach of Auto-Tuning to Numerical Libraries by Deep Learning, SIAM Conference on Parallel Processing for Scientific Computing 2018 (SIAM PP18), 2018
- ② 尾崎 克久, 行列積に対する隣接浮動小数点丸めの結果を返すアルゴリズム、日本応用数学会部会連合発表会、2018
- ③ Takeshi Ogita, Development of Verified Numerical Computations in High-Performance Computing Environments, 2018 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing (ATAT in HPSC 2018) (招待講演), 2018
- ④ Takeshi Ogita, Accurate and Verified

Numerical Computations with HPC, SIAM Conference on Parallel Processing for Scientific Computing 2018 (SIAM PP18), 2018

- ⑤ Takahiro Katagiri, Satoshi Ohshima, Masaharu Matsumoto, Algorithmic Revolution and Auto-tuning for Matrix Computations in Post Moore's Era, SIAM Conference on Computational Science and Engineering (CSE17), 2017
- ⑥ Takeshi Ogita, Iterative refinement of eigenvectors of symmetric matrices with clustered eigenvalues, SIAM Conference on Computational Science and Engineering (CSE17), 2017
- ⑦ Katsuhisa Ozaki, Takeshi Ogita, Faithful Rounding for Matrix Multiplication, SIAM Conference on Computational Science and Engineering (CSE17), 2017
- ⑧ 荻田 武史, 線形問題における HPC 向きの精度保証法と反復改良法、第 9 回自動チューニング技術の現状と応用に関するシンポジウム(ATTA2017) (招待講演)、2017
- ⑨ 荻田 武史, 数値線形代数における反復改良法、第 17 回 AT 研究会オーペンアカデミックセッション(ATOS17), (招待講演)、2017
- ⑩ 荻田 武史, 実対称行列の固有値分解に対する反復改良法、応用解析研究会、2016

[図書] (計 0 件)

なし

[産業財産権]

○出願状況 (計 0 件)

なし

○取得状況 (計 0 件)

なし

[その他]

ホームページ等

<http://www.abc-lib.org/MyHTML/getpaper.html>

6. 研究組織

(1) 研究代表者

片桐 孝洋 (KATAGIRI, Takahiro)
名古屋大学・情報基盤センター・教授
研究者番号: 40345434

(2) 研究分担者

荻田 武史 (OGITA, Takeshi)
東京女子大学・現代教養学部・准教授
研究者番号: 00339615

尾崎 克久 (OZAKI, Katsuhisa)
芝浦工業大学・システム理工学部・准教

授

研究者番号：90434282

(3)連携研究者

なし

(4)研究協力者

市村 駿太郎 (ICHIMURA, Syuntaro)