

令和 3 年 5 月 19 日現在

機関番号：10101

研究種目：若手研究

研究期間：2018～2020

課題番号：18K18058

研究課題名（和文）HPCの視点に基づくテンソル分解アルゴリズムの高性能化

研究課題名（英文）Study on improving algorithms for tensor decomposition based on the HPC viewpoint

研究代表者

深谷 猛（Fukaya, Takeshi）

北海道大学・情報基盤センター・助教

研究者番号：30633846

交付決定額（研究期間全体）：（直接経費） 3,100,000円

研究成果の概要（和文）：本研究では、データサイエンス等の基盤技術の一つであるテンソル分解の計算手法の高性能化を目的として、従来の数理的な視点に加えて、高性能計算（HPC）の視点から最新の計算機の特徴を考慮することで、より効率的な計算手法の構築を試みた。具体的な成果としては、代表的なテンソル分解計算の主要部に対して、条件に応じて適切な実装方法を選択することで計算を高速化することが可能であることを示した。また、テンソル分解と関連の深い行列計算に対して、既存手法よりも効率的な計算手法を新たに提案し、その有効性を確認した。

研究成果の学術的意義や社会的意義

本研究では、様々な分野での活用が期待されているデータサイエンスの基盤技術の一つであるテンソル分解の高性能化を目的とした研究を行った。また、計算機の複雑化・多様化により、数理とHPCの両方の視点が効率的な計算の実現に不可欠であり、本研究を通して、その重要性を示すことができた。国内のHPC分野におけるテンソル分解の研究事例が乏しかったので、本研究をきっかけとして、今後、様々な研究が展開されることが期待される。

研究成果の概要（英文）：In this research, we aimed for improving algorithms of tensor decomposition, which is one of building blocks in data science applications. In addition to the traditional mathematical viewpoint, we investigated efficient algorithms based on the HPC viewpoint, in which the characteristics of recent computers are considered. Through this research, we found that a dominant computation in a typical tensor decomposition can be accelerated by appropriately selecting kernels depending on the conditions such as the size of a tensor. We also presented a new efficient algorithm for a matrix computation deeply related to tensor decomposition.

研究分野：高性能計算、数値線形代数、応用数学

キーワード：テンソル分解 線形計算 並列計算

## 様式 C-19、F-19-1、Z-19（共通）

### 1. 研究開始当初の背景

様々な分野でデータサイエンスの活用が期待され、大量のデータを効率的に扱う技術の必要性が高まっている。例えば、データを行列として扱うことが多いため、主成分分析やデータ圧縮に利用される特異値分解をはじめとした行列計算手法の高性能化は、データサイエンス分野に大きなインパクトを与える。

一方、近年、データの大規模化とともに、複雑なデータをより適切に扱うための手法として、テンソル分解が注目を集めており、脳科学分野での活用例などが報告されている。今後、より多くの分野にデータサイエンスが普及し、また、IT 技術の更なる進歩にともない、より大量かつ複雑なデータの取得が可能となる状況において、今以上にテンソル分解を用いたデータ処理が活発になることが予想される。したがって、テンソル分解の計算手法を高性能化することは非常に重要な課題の一つである。

このように、テンソル分解に関する研究が活発化しているが、最近の研究状況を概観すると、数理的視点から数多くの研究が行われている一方で、高性能計算（HPC）視点での研究に限られていることが確認できる。テンソル分解の計算を実際に行う計算機のアーキテクチャは年々複雑かつ多様になっており、その性能を十分に引き出すことは容易ではない。そのため、高性能なアルゴリズムを開発するためには、数理と HPC の両面からアルゴリズムの研究を行う必要がある。実際、行列計算分野において、近年有効性が注目されている通信回避型アルゴリズムなどは、数理と HPC の両方の視点でアルゴリズムの再構成を行った成果であり、テンソル分解計算に対しても、同様の試みを実施する価値が十分にあると言える。

### 2. 研究の目的

前項に記載した背景を踏まえて、本研究では、「HPC の視点から、テンソル分解の計算手法の高性能化に貢献する」という点を大局的な目的として掲げる。具体的には、「計算時間の意味で、より高性能なテンソル分解の計算手法を確立する」ことを最終目的として、数理と HPC の両面からテンソル分解の計算手法を研究する。

現在、数理的な視点から提案されたテンソル分解のアルゴリズムは、演算回数や収束性の点において優れているが、一方で、現在の計算機においては「演算回数 $\div$ 実行時間」ではないので、計算時間の面でベストとは限らない。本研究では、数理的視点に、メモリアクセスコストや並列化効率などの HPC 的視点を加えた上で、総合的にアルゴリズムを研究することで、上述の目的の達成を目指す。

また、テンソル分解に対する HPC 視点での研究が乏しい現状に対して、本研究を行うことで、新しい問題点を明らかにすることができる。これにより、過去の行列計算の場合と同様に、自動チューニング技術やライブラリ・フレームワークの研究・開発といった、テンソル分解に関するより高度かつ実用上重要な HPC 技術の研究への展開につながることを期待される。

### 3. 研究の方法

本研究では、主要なテンソル分解とその代表的な計算手法（例：ALS 法を用いた CP 分解、HOSVD を用いた Tucker 分解）を研究対象とする。また、テンソル分解の過程で出現する典型的な行列計算（例：縦長行列の QR 分解）についても、研究対象に含める。これらの計算に関して、マルチコア CPU やメニーコア CPU 環境を想定し、シングルノード上での高性能化を考える。

研究手順としては、まず、一般的な実装の性能を分析し、高性能化に向けた課題を特定する。次に、実装方法の改善を検討することで、性能向上を図る。その後、アルゴリズム自体の改良を検討し、更なる性能向上を目指す。

### 4. 研究成果

主な研究成果は以下の通りである。

#### (1) MTTKRP カーネルの高性能化に関する研究

テンソルの CP (CANDECOMP/PARAFAC) 分解は、行列の低ランク近似に類似した、主要なテンソル分解の一つである。また、その代表的な計算手法として、ALS (Alternating Least Square) 法が知られている。ALS 法は反復型のアルゴリズムであり、各反復において、テンソルを行列化した結果と Khatori-Rao 積 (KRP) との積の計算 (MTTKRP: Matricize Tensor Times KRP) を行うが、この処理が計算時間の大部分を占めることが分かっている。そこで、本研究では、MTTKRP の高速化を試みる。

ALS 法における MTTKRP の計算では、異なるモードでテンソルの行列化を行うため、数学的には同じ計算でも、計算機上でのメモリアウトが異なる。そのため、この点を踏まえて、適切な MTTKRP の実装を選択することが必要であると考えた。そこで、本研究では、複数の実装を用意し、条件（テンソルの階数、大きさ、モード）ごとの各実装の性能を分析した。

図1は得られた結果の一例である。階数と次元の異なるテンソルにおける、各モードのMTTKRPについて、実装方法によって性能（計算時間）が大きく異なっていることが確認できる。具体的には、 $K=3$  ( $N=960$ ) におけるモード1や3の場合には、既存（Intel MKL）のBLASライブラリで提供されているスレッド並列版DGEMMルーチンが最も高速であるが、それ以外の場合には、逐次版DGEMMをOpenMPで並列実行する実装が高速であった。また、素朴な実装（単純な多重ループ）の性能は著しく低いことも確認できた。今後の展望としては、計算対象の行列やアルゴリズム中のモードに応じて、適切なMTTKRP実装を自動的に切り替える機能を組み込んだ計算手法の開発等が期待される。

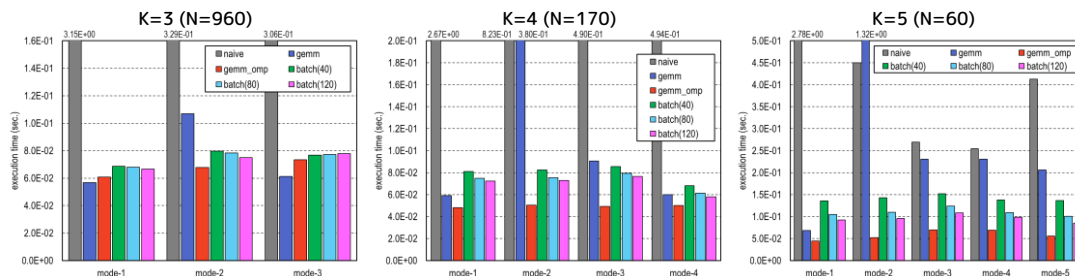


図1：各モードにおけるMTTKRPに対する異なる実装方法の性能比較結果  
(テンソル： $K$ 階・ $N$ 次、環境：北大 Grand Chariot1 ノード・40 スレッド)

## (2) 縦長行列のQR分解に関する研究

テンソル分解のアルゴリズムでは、テンソルを行列化して計算を行うことが多々ある。テンソルを行列化する場合、一般的な行列計算で現れる行列とは異なった形状の行列となる場合が多く、縦長行列はその一例である。そのため、縦長行列のような特徴的な形状の行列を対象とした基本的な行列計算の高速化は、テンソル分解の高性能化において重要な課題の一つとなる。そこで、本研究では、縦長行列に対する効率的なQR分解の計算手法について研究を行った。

これまでに、我々はコレスキーQR型のQR分解アルゴリズムの研究を行っている。今回の研究では、これまで対応していなかった難しい（条件数が大きい）行列への対応方法を研究し、シフトを用いたShifted Cholesky QR法を前処理とする新しい計算手法を提案した。

図2は、縦長行列のQR分解に対する各アルゴリズムの性能を、最新のマルチコアCPU上で比較した結果である。条件数の大きい行列に対して十分な精度で計算をすることが可能なアルゴリズム同士を比較しており、計算時間の点で提案手法（ScholQR3）が優位であることが確認できた。加えて、分散並列環境上でも性能評価を実施しており、Oakforest-PACSの4096ノードまでを用いた実験において、提案手法の有効性も確認されている。

また、研究期間の後半では、数値的にランク落ちした行列に対応するため、列ピボットを用いたコレスキーQR型のアルゴリズムの研究も開始した。現状、プロトタイプ版のアルゴリズムが完成した段階ではあるが、期待通りの計算精度と性能が得られることを確認しており、今後は理論的な解析や詳細な性能評価を実施する予定である。

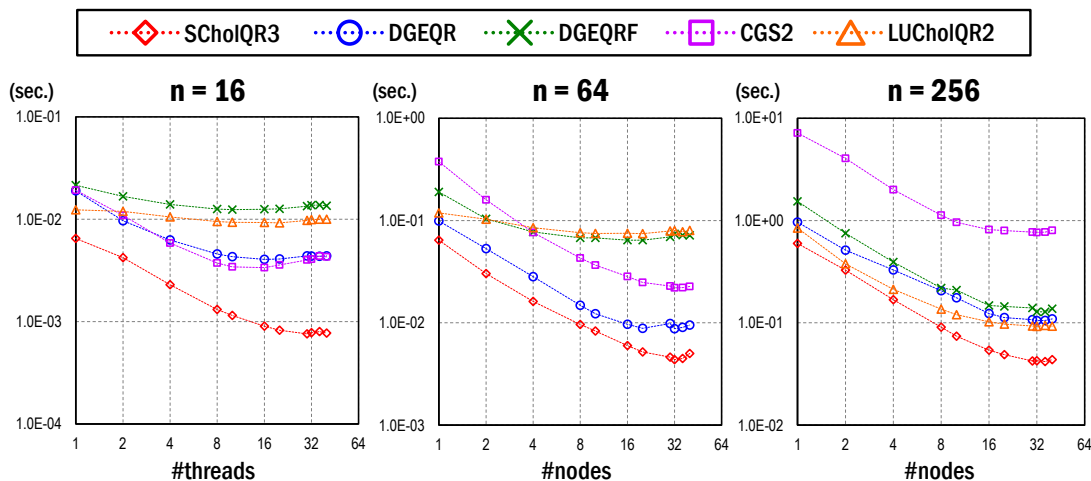


図2：縦長行列のQR分解に対する各アルゴリズムの性能比較結果  
(行列：10万× $n$ 、環境：北大 Grand Chariot1 ノード・最大40スレッド)

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 1件/うちオープンアクセス 1件）

1. 著者名 Takeshi Fukaya, Ramaseshan Kannan, Yuji Nakatsukasa, Yusaku Yamamoto, Yuka Yanagisawa	4. 巻 42
2. 論文標題 Shifted Cholesky QR for Computing the QR Factorization of Ill-Conditioned Matrices	5. 発行年 2020年
3. 雑誌名 SIAM Journal on Scientific Computing	6. 最初と最後の頁 A477 ~ A503
掲載論文のDOI（デジタルオブジェクト識別子） 10.1137/18M1218212	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1. 著者名 深谷猛	4. 巻 2018-HPC-167
2. 論文標題 マルチコアCPU環境における密テンソル向けALS法の性能評価	5. 発行年 2018年
3. 雑誌名 情報処理学会研究報告：ハイパフォーマンスコンピューティング	6. 最初と最後の頁 1-9
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計10件（うち招待講演 0件/うち国際学会 6件）

1. 発表者名 深谷 猛
2. 発表標題 テンソル分解におけるMTTKRPのスレッド並列化に関する考察
3. 学会等名 第24回計算工学講演会
4. 発表年 2019年

1. 発表者名 深谷 猛
2. 発表標題 ALS法を用いた密テンソルのCP分解におけるMTTKRPの性能評価
3. 学会等名 The 3rd cross-disciplinary Workshop on Computing Systems, Infrastructures, and Programming (xSIG2019)
4. 発表年 2019年

1. 発表者名 深谷 猛
2. 発表標題 HPC視点に基づくテンソル分解アルゴリズムの高性能化
3. 学会等名 第11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(ATTA2019)
4. 発表年 2019年

1. 発表者名 Takeshi Fukaya
2. 発表標題 Recent progress of the Cholesky QR factorization
3. 学会等名 2019 Mini-Workshop on Computational Science (MWCS2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Takeshi Fukaya, Ramaseshan Kannan, Yuji Nakatsukasa, Yusaku Yamamoto, Yuka Yanagisawa
2. 発表標題 Shifted CholeskyQR3 for High Performance Tall-Skinny QR Factorization
3. 学会等名 2020 SIAM Conference on Parallel Processing for Scientific Computing (SIAM PP20) (国際学会)
4. 発表年 2020年

1. 発表者名 深谷猛
2. 発表標題 密テンソルに対するALS法の実装方法に関する考察
3. 学会等名 日本応用数理学会2018年度年会
4. 発表年 2018年

1 . 発表者名 Takeshi Fukaya, Yusaku Yamamoto
2 . 発表標題 An overview of various algorithms for computing tall-skinny QR factorization
3 . 学会等名 he 37th JSST Annual International Conference on Simulation Technology (JSST2018) ( 国際学会 )
4 . 発表年 2018年

1 . 発表者名 Takeshi Fukaya, Ramaseshan Kannan, Yuji Nakatsukasa, Yusaku Yamamoto, Yuka Yanagisawa
2 . 発表標題 Performance Evaluation of the Shifted Cholesky QR Algorithm for Ill-Conditioned Matrices
3 . 学会等名 SC ' 18: The International Conference for High Performance Computing, Networking, Storage, and Analysis ( 国際学会 )
4 . 発表年 2018年

1 . 発表者名 Takeshi Fukaya, Ramaseshan Kannan, Yuji Nakatsukasa, Yusaku Yamamoto, Yuka Yanagisawa
2 . 発表標題 Shifted Cholesky QR algorithm for computing the QR factorization of ill-conditioned matrices
3 . 学会等名 2019 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing (ATAT2019) ( 国際学会 )
4 . 発表年 2019年

1 . 発表者名 Takeshi Fukaya, Ramaseshan Kannan, Yuji Nakatsukasa, Yusaku Yamamoto, Yuka Yanagisawa
2 . 発表標題 High performance QR factorization of ill-conditioned matrices based on the Cholesky QR algorithm
3 . 学会等名 2019 SIAM Conference on Computational Science and Engineering (CSE19) ( 国際学会 )
4 . 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	山本 有作 (Yamamoto Yusaku)		
研究協力者	中務 佑治 (Nakatsukasa Yuji)		
研究協力者	柳澤 優香 (Yanagisawa Yuka)		
研究協力者	Kannan Ramaseshan (Kannan Ramaseshan)		

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関		
英国	University of Oxford		