

大規模線形行列方程式に対する数値解法的高速化に関する研究

佐竹 祐樹（北海道大学）

概要

線形行列方程式は、制御理論や画像処理、物理シミュレーションなど幅広い分野に現れるため、その求解は重要である。特に、科学技術の発展に伴い、より大規模な問題の求解が求められており、大規模線形行列方程式に対する実用的な高速数値解法の構築は重要な研究課題である。本研究では、より大規模な線形行列方程式の求解を可能にすることを目標として、数値および HPC の両面からのアプローチから実用的数値解法の構築に取り組む。2024 年度には、行列方程式に対する数値解法に関連して、低ランク型 Krylov 部分空間法、シフト線形行列方程式に対する数値解法、縦長行列に対する QR 分解に関する実装・研究開発をそれぞれ行った。これらを踏まえて、2025 年度も継続課題として、引き続き行列方程式に対する実用的数値解法の構築に向けて研究を進めていく予定である。

1 共同研究に関する情報

1.1 共同研究を実施した拠点名

- 北海道大学 情報基盤センター
- 東京大学 情報基盤センター
- 京都大学 学術情報メディアセンター

1.2 課題分野

- 大規模計算科学課題分野

1.3 参加研究者の役割分担

- 佐竹祐樹: 課題代表、研究統括、行列方程式に対する数値解法の実装・研究開発
- 深谷猛: 課題副代表、統括補佐、QR 分解、並列計算に関する助言
- 岩下武史: Krylov 部分空間法、前処理、並列計算に関する助言
- 曾我部知広: シフト線形方程式、Krylov 部分空間法、前処理に関する助言
- 萬本遼太郎: シフト線形方程式に対する数値解法の並列化

2 研究の目的と意義

線形行列方程式は、制御理論や振動解析などの工学分野や、画像処理などの情報学分野に現れる。また、偏微分方程式の離散化においても現れることか

ら、幅広い分野においてその求解が重要である。特に、科学技術の発展とともに、各応用分野において、よりサイズの大きい方程式の求解が必要となっており、大規模な線形行列方程式に対する高速な数値解法の構築は重要な研究課題といえる。

上記のように幅広い応用があるにも関わらず、大規模線形行列方程式に対する数値解法に関する研究は、数値線形代数分野における他の研究課題に比べてあまり注目されてこなかった。その一因として、線形行列方程式はある種の連立 1 次方程式とみなせる点が挙げられる。大規模連立 1 次方程式に対する数値解法に関する研究は、古くから現在に至るまで多くの研究者によって行われてきた。したがって、線形行列方程式に特化した数値解法でなくても、連立 1 次方程式に対する多くの数値解法を利用することができる。

その一方で、線形行列方程式のまま求解を行うことで、連立 1 次方程式とみなして求解を行う場合に比べて所要メモリの大幅な削減と計算の高速化が期待できることから、20 年ほど前から行列方程式の数値解法に関する研究が盛んに行われるようになってきた。現在も数値解法的高速化に向けた研究が行われているが、数理的な観点からの研究がほとんどであり、実際の計算機上での実装（並列計算等）を考慮した研究はあまり行われていない。今後応用分野

において、さらに大規模な問題の求解が必要になってくると考えられ、また、「ムーアの法則」が限界に近づきつつある現状を踏まえると、数理的視点に加えて HPC の観点からのアプローチによる数値解法的高速化が必要である。本研究では、数理及び HPC の両面の視点から大規模線形行列方程式に対する数値解法的高速化を行うことが目的である。

本研究では、連立 1 次方程式とみなすのではなく、行列方程式のまま求解することで所要メモリを大きく削減できることから、より大規模な問題への対応が可能となる。また、数理及び HPC の両面から計算的高速化を実現することで、真に実用的な数値解法の確立につながる。線形行列方程式は、上述の応用分野に加え、非線形行列方程式の求解にも用いられるため、これらの理由から非常に幅広い分野への波及効果が期待できる。

3 当拠点公募型研究として実施した意義

上述のように、実用性の観点から、実際の計算機上での実装を考慮した数値解法の構築が重要である。特に、現実的な応用分野に現れる大規模問題を考える場合、豊富な計算資源を有する計算機システムを用いる必要がある。したがって、本研究の目的を達成する上で、JHPCN の各拠点の計算機環境での数値解法の開発を行うことは理に適っている。

また、JHPCN のシンポジウム等で多くの応用分野の研究者との交流が可能であり、本研究に対して応用分野の研究者からの知見が得られるとともに、応用分野の研究者に本研究の手法を認知・利用してもらうことで、将来的に幅広い分野への貢献が期待される。

4 前年度までに得られた研究成果の概要

該当なし

5 今年度の研究成果の詳細

本年度では、大規模線形行列方程式に対する数値解法に関連して下記の 3 つの内容について実装・研究開発を行った。

(1) 低ランク型 Krylov 部分空間法の実装

行列方程式の一種である Sylvester 方程式を対象として、低ランク型 Krylov 部分空間法（低ランク BiCGSTAB 法）の実装を行った。低ランク型

Krylov 部分空間法は Krylov 部分空間法（連立 1 次方程式に対する主要解法）をベースにした反復法で、近似解を低ランク行列で近似することでメモリを節約して計算を行う。ただし、近似解のランクは反復を繰り返すことで増加するため、近似解の小さい特異値を何らかの基準にしたがって打ち切ることで低ランク性を維持する。小さい特異値を打ち切る基準 ϵ を変えたときの収束履歴および近似解のランク変化を図 1 に示す。図 1 から、打ち切り基準 ϵ が最終的に到達する残差の精度に影響することが確認できる。打ち切り基準 ϵ を小さく設定する（厳しくする）ほど精度が上がり、逆に大きく設定するとほとんど相対残差ノルムは下がらなくなる。したがって必要な精度に応じて、打ち切り基準を設定する必要がある。ただし、打ち切り基準を小さくするほど、近似解のランクが増加しやすく、所要メモリ・計算時間が増加してしまう。これらを踏まえて、打ち切り基準を動的に変化させることで、所要メモリを抑えながら所望の精度に到達させることができるか今後検討する必要がある。

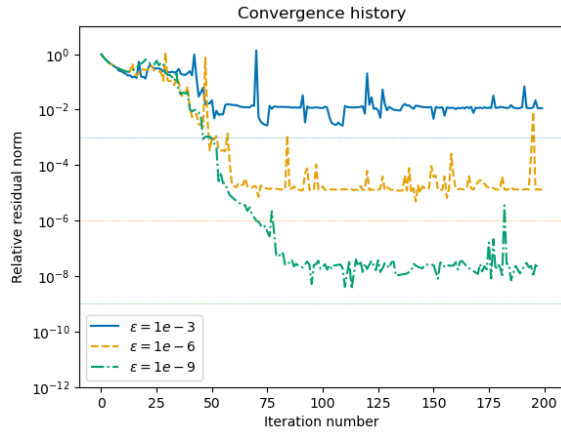
(2) シフト線形方程式に対する LOPBiCG 法のハイブリッド並列化

シフト線形方程式

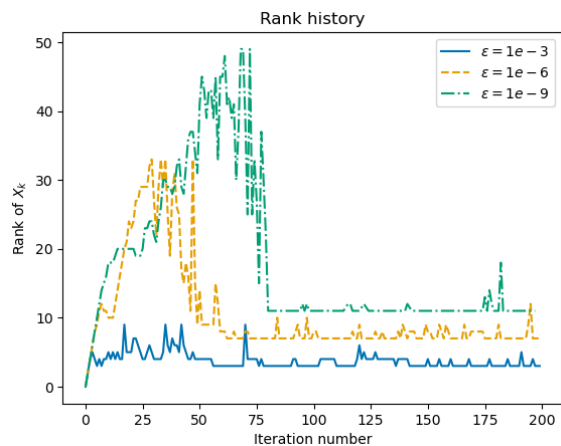
$$(A + \sigma_i I) \mathbf{x}^{(i)} = \mathbf{b}, \quad i = 1, \dots, m \quad (1)$$

は、行列方程式の一種である Sylvester 方程式の特殊な場合に相当し、格子量子色力学や大規模電子構造計算などの分野に現れる。シフト線形方程式に対する数値解法として、シフト付き Krylov 部分空間法と呼ばれる解法群が知られており、近年、シフト付き Krylov 部分空間法の一つとして Shifted LOPBiCG 法が提案された。この手法は、スイッチング技術（収束するまでの計算を効率的に行うための手法）を非対称な係数行列をもつシフト線形方程式に対して適用するために提案されている。本研究では、最新のシフト付き Krylov 部分空間法の一つである Shifted LOPBiCG 法の分散並列環境における MPI と OpenMP を用いたハイブリッド並列実装を行った。

係数行列（疎行列）および計算に必要な各ベクトル全てに対して行方向の 1 次元ブロックデータ分散を行い、スカラーパラメータは各プロセスが冗長に保持するように計算を行った。Shifted LOPBiCG 法における主な計算は、



(a) 相対残差履歴



(b) 近似解のランク変化

図 1: Sylvester 方程式（行列サイズ： $n = 25^3$ ）に対する低ランク BiCGSTAB 法の相対残差履歴と近似解のランク変化

- 行列ベクトル積
- ベクトルの内積
- ベクトルの演算（スカラー倍、ベクトル同士の和）

であり、上述のデータ分散により、ベクトルの演算は各プロセスで独立に処理できる。行列ベクトル積、ベクトルの内積では集団通信が必要となるが、それぞれ Allgather、Allreduce を用いることで処理できる。また、各プロセス内の処理に関して OpenMP を用いてスレッド並列化を行った。

MPI を用いたプロセス並列化においては、プロセス間の集団通信が並列化の速度向上を阻害する原因となり得る。そこで、Shifted LOPBiCG 法のアルゴリズムを利用して、通信と計算のオーバーラップにより通信時間の隠蔽を行った。Shifted LOPBiCG 法では反復毎にシード方程式とシフト方程式の 2 種類の方程式に対する計算を行っており、通信が必要

となるのはシード方程式に関する計算のみである。また、シフト方程式に関する計算では、次の反復の通信に必要な演算は行わない。これらの事実から、シード方程式の計算における行列ベクトル積（通信）と 1 反復前のシフト方程式の計算をオーバーラップすることで、通信時間の隠蔽を行った。

上述の Shifted LOPBiCG 法のハイブリッド並列化について、東京大学情報基盤センターの Wisteria / BDEC-01 (Odyssey) を用いて性能評価を行った。問題には SuiteSparse Matrix Collection から取得した行列を用いた。

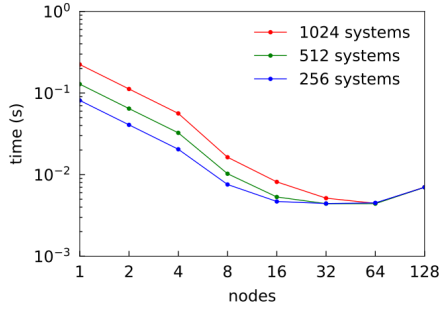
図 2 は、ノード数を変化させたときの 1 反復あたりの計算時間の変化を示している。各図における赤、緑、青の線はそれぞれ方程式の数 $m = 1024$ 、512、256 を表している。いずれの行列も 16～32 ノードあたりまで台数効果を確認することができた。32、64、128 ノードでは、方程式の数 m の違いによる反復当たりの実行時間の違いはほとんど見られなかった。

次に、シフト方程式の計算を用いたオーバーラップの有無による実行時間の比較を図 3 に示す。time ratio は、オーバーラップありの実行時間をオーバーラップなしの実行時間で割った値である。したがって、この値が小さいほど高速化できていることを表す。図 3 から、方程式の数が多いほど本手法が優位な傾向にあることを確認できた。

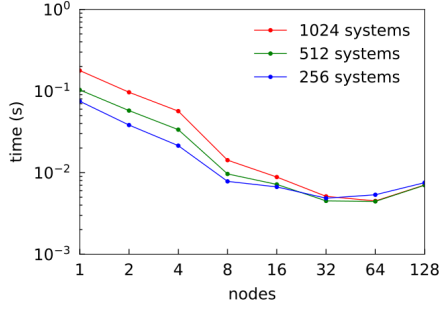
最後に、行列 ss 、 $m = 1024$ の場合に対して計算を行った際の、反復毎の計算時間の内訳を図 4 に示す。各図において、水色の部分はシード方程式の計算、緑色の部分は通信（Allgather）とシフト方程式の計算、斜線部分は通信（Allgather）のみの実行時間を表す。また、未収束の方程式の数を実線で示しており、反復回数が増えるにつれて未収束の方程式の数が減少し、最終的に全て収束している。16 ノードおよび 32 ノードでは、最初はシフト方程式の計算時間で通信時間を隠蔽できているが、徐々に未収束の方程式の数が減少することでシフト方程式の計算も減り、途中からは通信時間が支配的となった。64 ノードおよび 128 ノードでは最初から通信時間が支配的であった。

(3) 縦長行列に対する QR 分解

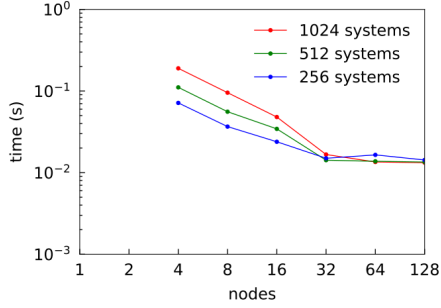
大規模行列方程式に対する数値解法においては、アルゴリズム中で縦長行列に対する QR 分解を計算する必要がある。今後、行列方程式に対する数値解法



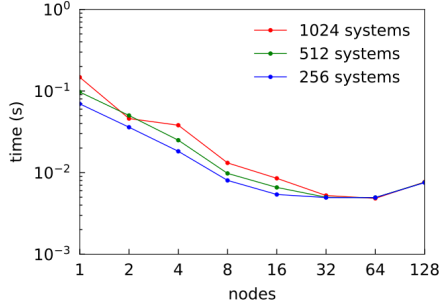
(a) atmosmodl



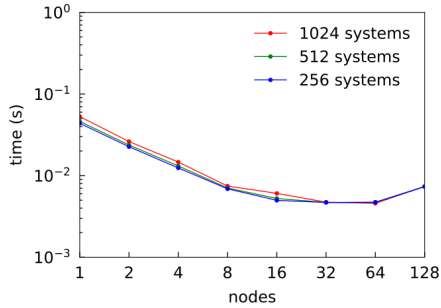
(b) cage14



(c) cage15

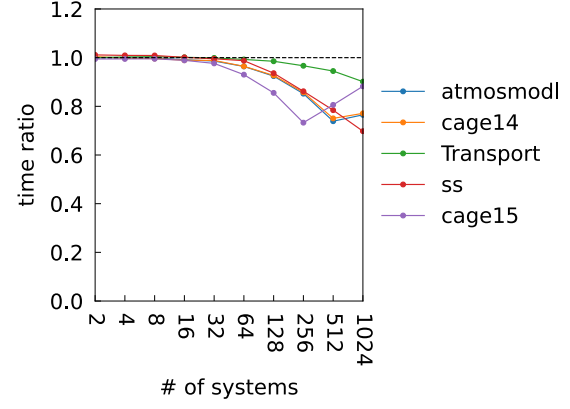


(d) ss

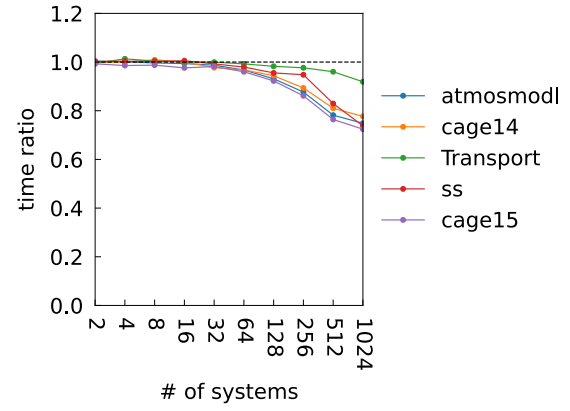


(e) transport

図 2: ノード数を変化させたときの 1 反復当たりの実行時間



(a) 16 ノード



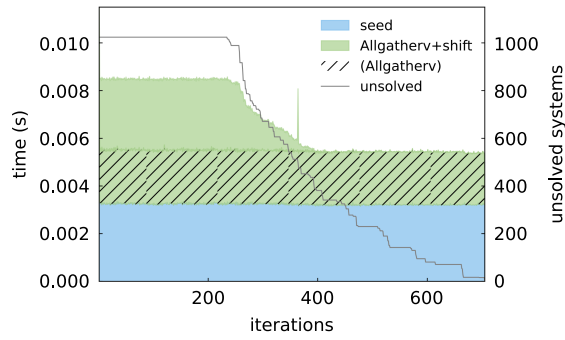
(b) 32 ノード

図 3: 方程式の数 m の違いによるオーバーラップの効果の違い。time ratio は (オーバーラップありの場合の実行時間) / (オーバーラップなしの実行時間) を表す。

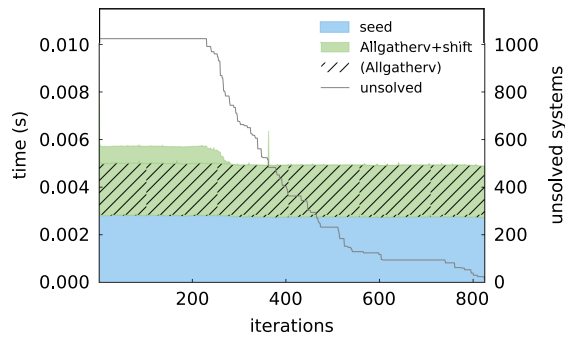
を分散並列実装することを踏まえて、分散並列環境における QR 分解アルゴリズム (TSQR、Householder QR、CGS2、Shifted CholeskyQR3 の 4 種類) の性能評価を行った。

北海道大学情報基盤センター Grand Chariot で性能評価した結果を図 5 に示す。テスト行列は乱数で生成し、行列のサイズは行数 $m = 2^{24}$ 、列数 $n = 16$ 、64、256 で実験を行った。図 5 から分かるように、分散並列環境において、深谷らが提案した通信回避型のアルゴリズムである Shifted CholeskyQR3 が有用であることを確認した。

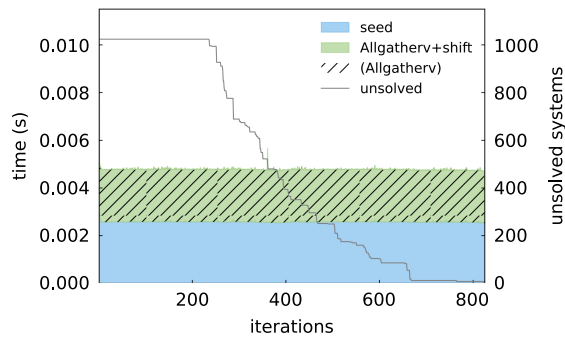
低ランク型 Krylov 部分空間法のアルゴリズム内で行われる QR 分解の対象となる縦長行列は、通常ランク落ちしている。一方で、縦長行列向けに有効であるとされる CholeskyQR 型の QR 分解アルゴリズムは、そのままランク落ちした行列に適用すると計算が破綻してしまうという問題点がある。そこ



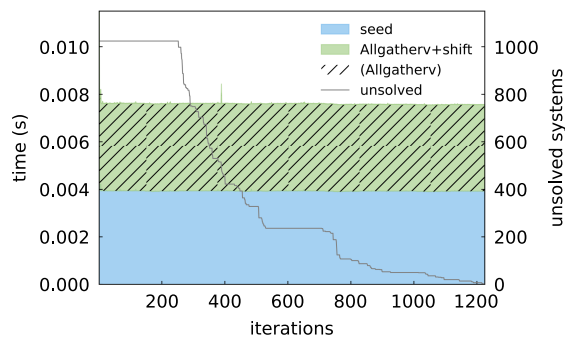
(a) 16 ノード



(b) 32 ノード

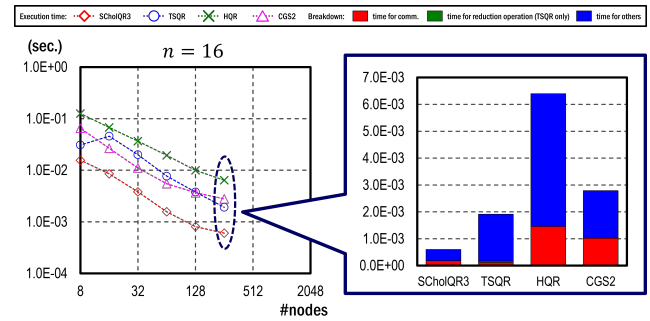


(c) 64 ノード

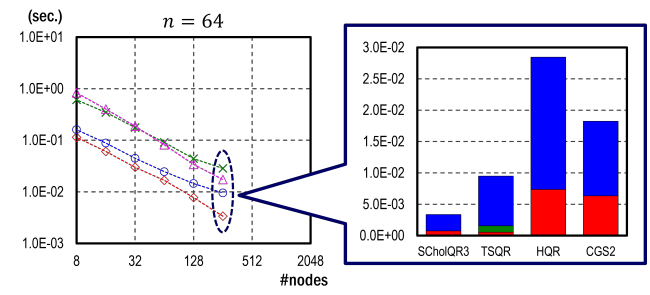


(d) 128 ノード

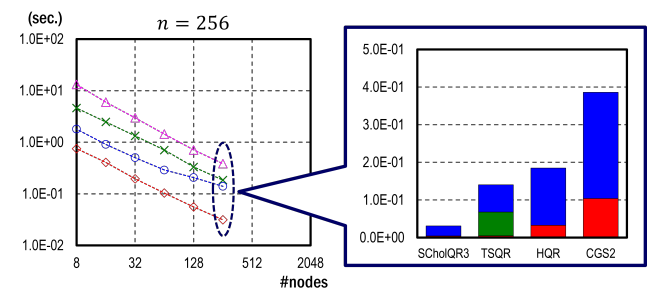
図 4: ノード数による反復毎の実行時間の内訳の変化 (行列名: ss、 $m = 1024$)。



(a) $n = 16$



(b) $n = 64$



(c) $n = 256$

図 5: ノード数による反復毎の実行時間の内訳の変化 (行列サイズ: $m = 2^{24}$)。

で、CholeskyQR 型アルゴリズムのランク落ちしている場合への拡張 (列ピボット付き QR 分解に対する CholeskyQR 型アルゴリズム) に関して、実装・性能評価を行い、その有効性を確認した。

6 今年度の進捗状況と今後の展望

申請時に設定した 2 種類の解法 (低ランク型 Krylov 部分空間法と射影法) のうち、低ランク型 Krylov 部分空間法を中心に実装を行った。射影法については、2025 年度に引き続き実装を行っていく予定である。

また、申請時には想定していなかったシフト線形方程式に対する数値解法に関して一定の成果を得ることができた。シフト線形方程式に対する数値解法 (シフト付き Krylov 部分空間法) は、分散並列環境

における実装での性能評価に関する既存研究は報告されていなかったため、提案手法の有用性を確認できたことは非常に有意義である。シフト線形方程式はある種の行列方程式と等価であることから、本研究は行列方程式に対しても有意義な結果である。等価な行列方程式は限定的であるが、本研究の手法をベースに、より広範囲の行列方程式に対する数値解法の開発にも今後取り組む予定である。

2024 年度には低ランク型 Krylov 部分空間法の分散並列実装も進めていたが、計算過程における QR 分解のアルゴリズムにおいてランク落ちの影響を考慮する必要が出てきたため、途中でアルゴリズムの修正が必要となった。そのため、並列化に関して遅れが生じたが、この点については上述の深谷らを中心に研究が行われている CholeskyQR 型の列ピボット付き QR 分解アルゴリズムを低ランク型 Krylov 部分空間法のアルゴリズムに組み込むことで対応する予定である。

申請時、初年度においてどこまで進められるか予測できていなかった点と課題を進める上で方針にやや修正が必要になったことから、当初の計画に比べて全体的に遅れが生じた。一方で、新たな方向性の成果も得られた点は評価できている。2025 年度は、今年度の反省も踏まえて計画的に進めていきたい。