

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Linear Time Dense Direct Solver Without Trailing Sub-matrix Dependencies
著者(和文)	MAQianxiang
Author(English)	Qianxiang Ma
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12880号, 授与年月日:2024年9月20日, 学位の種別:課程博士, 審査員:横田 理央,吉瀬 謙二,宮崎 純,DEFAGO XAVIER,小野 峻佑
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12880号, Conferred date:2024/9/20, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

論文審査の要旨及び審査員

報告番号	甲第 号		学位申請者氏名	Qianxiang Ma		
		氏名	職名		氏名	職名
論文審査 審査員	主査	横田 理央	教授	審査員	小野 峻佑	准教授
	審査員	吉瀬 謙二	教授			
		宮崎 純	教授			
		Xavier Defago	教授			

論文審査の要旨 (2000 字程度)

本論文は「Linear Time Dense Direct Solver Without Trailing Sub-Matrix Dependencies」と題し、 $N \times N$ の密行列の LU 分解の計算量を $O(N^3)$ から $O(N)$ に低減できる階層的低ランク近似法において処理の依存関係を解消するアルゴリズムを提案し、その分散並列実装と GPU 実装の性能測定を行ったもので、英文で書かれており全 6 章からなる。

1 章「Introduction」では、高性能計算分野において中心的な役割を果たしてきた線形代数ライブラリの発展の歴史を振り返りながら近年注目されている階層的低ランク近似法の利点と問題点について述べている。具体的には、階層的低ランク近似法が多くのアプリケーションにおいて密行列分解の計算量を $O(N^3)$ から $O(N)$ に低減できる利点について触れ、分散並列化を行う際に生じるデータの依存関係が性能低下をもたらしていることを問題点として挙げている。その上で、本論文の目的として LU 分解におけるデータの依存関係を解消し、分散並列実装および GPU 実装における並列化性能の向上を挙げている。

2 章「Background on Hierarchical Low-rank Approximation」では、本論文の前提となる密行列の低ランク近似について解説している。また、密行列を低ランク行列に圧縮するアルゴリズムとして現存する特異値分解(SVD), randomized SVD, interpolative decomposition, adaptive cross approximation などの手法を挙げ、それぞれの長所と短所を述べている。さらに、密行列を再帰的に分割し低ランク近似を行う上で有効なデータ構造について触れ、LU 分解を行う際のデータの依存関係にそれらがどのように関係しているかについて解説している。また、当該分野の研究が主に応用数学の研究者を中心に行われてきたため、高性能計算分野で頻繁に用いられている高速化・並列化の技法を適用した例が少ないことを問題として投げかけている。

3 章「The Use of Shared Basis and the Existing Best Practices」では、低ランク近似の基底を行ブロックや列ブロック間で共有することで可能となる HSS-ULV 法と H^2 -ULV 法を用いた LU 分解について述べている。HSS-ULV 法ではデータの依存関係を解消できる一方、弱許容条件の HSS 行列形式を用いるため適用できるアプリケーションが限定的であることを指摘している。逆に、 H^2 -ULV 法では強許容条件の H^2 行列を用いるため広範なアプリケーションに適用可能であるが、共有基底の再圧縮が必要になることからデータの依存関係が再び生じてしまうというジレンマについて概念図や数式を交えながら解説している。

4 章「Data-parallel ULV Factorization for H^2 -matrix on Many CPUs」では、3 章で述べた HSS-ULV 法と H^2 -ULV 法の長所を兼ね備えた提案手法とその CPU 上での分散並列実装について述べている。 H^2 -ULV 法においてデータの依存関係が生じる原因となっている共有基底の再圧縮を避けるために、密ブロックと低ランクブロックの処理を分け、密ブロックのみの LU 分解によって生じた fill-in ブロックを低ランクブロックと併せて共有基底を構築することが可能であることを解説している。これにより、 H^2 -ULV 法のように広範なアプリケーションに適用でき、HSS-ULV 法のようにデータの依存関係のない手法を提案している。また、単一の CPU 上でのスレッド並列化効率と複数の CPU 上での分散並列化効率を計測し、関連手法 LORAPO との比較において最大で 4,700 倍の高速化を実現できたことを示している。

5 章「Data-parallel ULV Factorization for H^2 -matrix on Many GPUs」では、4 章で述べた提案手法の GPU 実装と前進後退代入の並列化について述べている。提案手法を GPU 向けに実装する上で処理をバッチ行列積の形に変換することが重要であることを明らかにし、それを高速に実装するための時空間局在性を向上させる様々な工夫について述べている。また、CPU 上では支配的であった LU 分解の計算時間が GPU 上で高速化されることで前進後退代入の計算時間が支配的になることを明らかにした上で、これまで並列化が不可能だと思われていた前進後退代入の部分のデータの依存関

係を解消することで並列に計算できることを示している。これにより、既存手法の LORAPO に対して最大で 13,300 倍の高速化を実現している。

6 章「Conclusion」では、得られた知見と結論についてまとめている。主要な結論としては、密行列の LU 分解が階層的低ランク近似法を用いることで $O(N^3)$ から $O(N)$ の計算量に低減できるだけでなく、データの依存関係を解消することで並列化効率の高いアルゴリズムに変換できること、この提案手法を利用することでマルチスレッド実装、分散並列実装、GPU 実装を行う際に大幅な性能向上が得られること、それにより既存手法に対して大幅な高速化が得られることを挙げている。

以上より、本論文で提案している手法を用いることで高性能計算分野において中心的な役割を果たしてきた線形代数ライブラリの演算量を $O(N^3)$ から $O(N)$ に低減しながらも、分散並列化、GPU 実装の効果を最大限に引き出す並列化効率を実現し、それにより既存手法に対して大幅な高速化を実証できている。これは高性能計算分野において工学的に重要な貢献であるため博士（工学）の学位に相応しいと判断できる。

注意：「論文審査の要旨及び審査員」は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。