



HOKKAIDO UNIVERSITY

Title	Sparse Linear Solvers for High-Performance Computing Devices [an abstract of dissertation and a summary of dissertation review]
Author(s)	鈴木, 謙吾
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第16172号
Issue Date	2024-09-25
Doc URL	https://hdl.handle.net/2115/93432
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	doctoral thesis
File Information	Kengo_Suzuki_abstract.pdf, 論文内容の要旨



学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称 博士（情報科学） 氏名 鈴木 謙吾

学 位 論 文 題 名

Sparse Linear Solvers for High-Performance Computing Devices (高性能計算デバイスのための疎行列線形ソルバ)

数値解析は、現代の多くの学問分野を支える基盤技術の一つである。例えば、理工学分野における物理シミュレーションに関し、多くの応用研究が実施されている。それらの数値解析のアプリケーションにおいて、多くの現象は、大規模かつ疎な係数行列からなる線形方程式で記述され、その求解手法である、疎行列線形ソルバの実行が全体の処理を律速する。そのため、疎行列線形ソルバの高性能化は、多様な学問分野の高度化に資する、重要な研究課題である。

近年、消費電力の問題などから、計算デバイスの動作周波数の向上が停滞している。そのため、現在の高性能計算デバイスの多くは、動作周波数の向上に依存せず、高い演算性能を実現すべく、従来と異なるアーキテクチャを採用している。例えば、本研究で対象とする、Single Instruction Multiple Data (SIMD) 命令に対応した CPU や GPU は、ベクトル処理に適したアーキテクチャを有している。ベクトル処理では、一つの命令により複数の演算が実行されるため、低い動作周波数においても、高い演算性能が実現される。したがって、それらの高性能計算デバイスの特徴を最大限に活用することが、疎行列線形ソルバの高性能化に繋がると期待される。しかしながら、その特徴的なアーキテクチャにより、従来のアルゴリズムでは、高性能計算デバイスの性能を十分に引き出すことが困難である。そこで本研究では、代表的な疎行列線形ソルバである、前処理付きクリロフ部分空間法を基に、現在の高性能計算デバイスを十分に活用可能な、新たなアルゴリズムを検討した。具体的には、SIMD 演算に対応する Intel 社 CPU 及び、NVIDIA 社 GPU の二種類の計算デバイスを対象とし、それぞれのデバイスにおいて、新たなアルゴリズムを開発、評価した。加えて、整数演算を利用することで、求解精度を維持しながら、データ量削減を実現する手法についても検討を行った。これは、一度のベクトル演算で処理可能な要素数の増加、及び、データ転送量の削減によるメモリ帯域幅の有効利用を目的とした。さらに本手法は、整数演算の回路構造の簡単さから、再構成可能デバイス等の回路資源が限定された環境においても優位に働くことが期待できる。

本論文は、以下の内容を含む全 6 章から構成される。

第 1 章において、本研究の背景と位置付け、貢献について記述した後、第 2 章において、前処理付きクリロフ法について説明を与えた。また、本研究で対象とする、前処理付きクリロフ部分空間法における、SIMD 演算、GPU、及び低ビット演算の利用に関し、先行研究を示しながら、課題を明らかにした。

第 3 章では、クリロフ部分空間法を CPU で実行する際の著名な前処理手法である不完全 LU(ILU) 法において、SIMD 演算を効率的に利用する手法について検討を行った。SIMD 演算を利用した行列ベクトル積演算の数値実験結果を示しながら、近年の CPU において効果的に SIMD 演算を活用する一つの方向性として、ILU 分解において発生するフィルインに対し、SIMD 長に対応するブロック構造の制約を課し、得られる三角行列にブロック構造を創出する手法を提案した。この制約を課した場合においても、十分な収束性が実現されること、また、それらの三角行列を用いた代入計算におい

て, SIMD 演算を効果的に活用することができ, 大幅な求解時間の削減が達成されることを示した.

第 4 章では, 前処理付きクリロフ部分空間法において, GPU を効率的に活用する手法について検討を行った. 前処理付きクリロフ部分空間法の GPU 実装では, 特に前処理部分の並列性が重要となることを考慮し, 二つの方向性から, 新たなアルゴリズムを検討した. 一つ目の方向性では, グラフ彩色に基づく係数行列の並べ替えに着目し, 収束性を維持したまま, ILU 前処理の並列性を高める手法を検討した. CPU 上で有効性が確認された既存手法の GPU 上での性能評価, 及び, 新たな並べ替え手法の提案, 評価を行った. ILU 前処理の GPU 実装では, 収束性と並列性のトレードオフが課題であった. グラフ彩色に基づく係数行列の並べ替えにより, そのトレードオフが解消可能なことを示した. また, 同時に, 係数行列の非零要素数が多い問題に対し, 高い並列性が得られず, 良い性能が実現されない場合があることも確認した. そこで, 二つ目の方向性として, 係数行列の逆行列を直接近似する手法に注目し, 非零要素数が多い問題に対する並列性の向上を図った. 高速に逆行列の近似を構成可能なアルゴリズムを提案し, 有効性を確認した. GPU 上において, 比較的に非零要素数が多い問題を対象とする場合においても, 高性能な疎行列線形ソルバを実現できることを示した.

第 5 章では, ベクトル演算の更なる有効活用, 及び, データ転送量の削減を目的に, 低ビットデータ型に基づくクリロフ部分空間法を検討した. 特に, 低ビット整数 (固定小数点) 型に注目し, その有用性を検証した. 近年, 低ビット浮動小数点型の利用に関し, 多くの研究が実施されており, 一定の有効性が確認されている. しかし同時に, 求解精度の低下が問題となる例も報告されており, 求解精度を維持した低ビット化が課題となっている. 固定小数点型は指数部を必要としないため, 浮動小数点型と比較し, 有効桁の観点において優位性が期待できる. 反復改良の枠組みを用い, クリロフ部分空間法を整数演算により実現する, 新しいアルゴリズムを開発した. 数値実験により, 同ビット数の浮動小数点型を利用する場合と比較し, 求解精度, 実行時間共に高性能となるケースを確認し, 整数演算が浮動小数点演算の代替となり得ることを明らかにした.

第 6 章では, 本研究により得られた成果を総括し, 今後の展望について述べた.

以上のように, 本研究は, 高性能計算デバイスの活用による, 前処理付きクリロフ法に基づく疎行列線形ソルバの高性能化を目的とする. 高性能計算デバイスの機構に適した構造を有するアルゴリズムの開発, 評価, 及び, 低ビット整数演算の利用を可能にする手法, 実装の枠組みの検討を通じ, 高性能な線形ソルバの実現に加え, 現在, 研究開発が進められている, 新たな高性能計算デバイスを活用する一つの方向性を示した. これらの成果は, 将来にわたる, 疎行列線形ソルバの継続的な高性能化に貢献すると期待できる.