

Title	メモリ空間の再利用に注目した再帰Strassenアルゴリズム
Author(s)	李, 睿智
Citation	
Issue Date	2023-06
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18460
Rights	
Description	Supervisor:井口 寧, 先端科学技術研究科, 修士 (情報科学)

概要

行列の乗算は基本的な線形代数学の一部であり、機械学習、画像処理など様々な科学技術分野の数値計算に重要な役割を果たしている。ディープラーニング、エッジコンピューティングなど最近流行している技術は特に膨大なデータがあり、高精度が必要、大規模の行列乗算を必要となる機会が増えた。したがって大規模行列乗算を高速化する Strassen アルゴリズムが注目されている。

Strassen アルゴリズムは大きいサイズの行列の場合、標準的な行列乗算法と比べてより高速に計算できるアルゴリズムである。 M が偶数である $M \times M$ の正方行列 $C = A \times B$ の場合、まず C 、 A 、 B はそれぞれ行列数が $\frac{M}{2}$ 、4つのサブ行列に分け、Winograd で改良した Strassen アルゴリズムで7回の乗算と15回の加減算を行うことにより結果を求める。 M が大きくなると Strassen アルゴリズムの演算回数が標準乗算より少なくなり、高速になる。 A の行数と B の列数が対応する長方形行列は同じように考えることができ、padding 法で M が奇数の場合も対応できる。

Strassen アルゴリズムを GPU に応用するにはメモリ使用量の問題がある。これまでの研究には Malloc で確保した計算途中の結果を一時的に保存するための Temporary 行列を使用した、グローバルメモリ空間が有限な GPU にとって Strassen アルゴリズムの利点である大規模の行列乗算を応用しづらい問題がある。本研究では、NVIDIA GPU Tesla A100 を中心に、V100 と P100 を対比対象でメモリ空間の節約研究を行った。

メモリ空間を節約するため、再帰的な Strassen アルゴリズムにおいて、Malloc によるメモリ割り当てを避け、上位レベルの未使用行列を下位レベルに提供し、サポート行列として中間結果を保存する空間を提供することで、Malloc で確保した Temporary 行列全体を削除する方法を提案した。再利用できるメモリ空間がない問題に対して、メモリ空間使用分析アルゴリズムで下位 Level に提供できるメモリ空間を検索する上で計算順番の調整を行った。実験した結果、メモリ空間を重視した先行研究と比べて、Level-4 の計算に最大 18.11% のグローバルメモリが節約できた。結果により、最適化 Level の問題と更に高速化の問題が発見した。

実行速度が一番速い再帰の深度の問題、最適化 Level の問題に対して、GPU の乗算、加減算の性能関数の測定により判断する式を提案した。実験した結果、誤差が 2-7% あるが、各 Level の境界の関係が理論と合っている。

この上、更に高速化の探索について、先行研究を基にして、乗算だけではなく加減算の並列を含むことを中心に、部分同期を用いた並列処理手法を提案した。実験した結果、A100 で 40,960 サイズに対して、提案手法が最大 CUBLAS11.7 の標準行列乗算より 1.451 倍の SpeedUp が達成できた。