T2R2 東京科学大学 リサーチリポジトリ Science Tokyo Research Repository

論文 / 著書情報 Article / Book Information

論題(和文)	FMM と H^2(HSS) 行列のトレードオフについて				
Title(English)	Tradeoff Between FMM and H^2(HSS) matrices				
著者(和文)	横田理央				
Authors(English)	Rio Yokota				
出典(和文)	計算工学, Vol. 21, No. 4, pp. 34983501				
Citation(English)	Journal of the Japan Society for Computational Engineering and Science, Vol. 21, No. 4, pp. 34983501				
 発行日 / Pub. date	2016, 10				

FMM と \mathcal{H}^2 (HSS) 行列のトレードオフについて

横田 理央

1 諸論

FMM は粒子同士の相互作用を計算する N 体問題の 高速化手法として用いられてきたが、近年その代数学 的な拡張である \mathcal{H}^2 行列 [1] や HSS 行列 [2] などの研 究が盛んに行なわれている。従来の FMM では多重極 展開や局所展開などの解析的な低ランク近似を用いる ため、Green 関数解の存在する Poisson、Helmholtz、 Maxwell、Stokes などの限定的な偏微分方程式にしか 用いることができなかった。また、演算の種類も密行 列・ベクトル積に限定されていた。これとは対照的に、 H² 行列や HSS 行列などは幅広い問題に適用するこ とができ、行列分解や固有値解析にも用いることがで きる。しかし、kernel independent FMM [3] を用い ることで Gauss 関数や Matérn 関数など、Green 関 数以外の計算にも適用が可能になり、inverse FMM [4] を用いることで逆行列も計算できるようになるた め、従来の FMM の短所は改善されつつある。つまり、 FMM と \mathcal{H}^2 行列は同様の問題に適用することがで き、低ランク近似を多重極展開を用いて解析的に行う \boldsymbol{b} , adaptive cross approximation [5] \boldsymbol{b} interpolative decomposition[11] を用いて代数学的に行って行列を 保存するかの違いになる。したがって、問題に応じて 演算とメモリのトレードオフを考慮して適切な手法を 選択することが重要である。

2 階層的低ランク近似の多様性

FMM と \mathcal{H}^2 行列は階層的低ランク近似法の中でも 両極端に位置づけられるが、図 1 に示すようにそれら の間には様々な手法が存在する。ここでは、解析的な FMM と代数学的な \mathcal{H}^2 行列の中間に位置する手法を 取り上げ、演算とメモリのトレードオフに関して検討 する。

筆者紹介



横田 理央

東京工業大学、学術国際情報センター、准 教授。専門は高速多重極法の並列化、最 適化、GPU や Xeon Phi などへの実装。 2009 年 ACM Gordon Bell 賞 (価格性能 部門) 受賞。

		KIFN	IM			
ラン	ダム	black-bo	ox FMM	変換行	列の対角化	
	サンプリン	グ	並進・回転	云対称性		
H²行列	変換行	列の圧縮		変換行列	の保存	FMM
代数学的					解析的(約	後何学的)
メモリ [Byt	es]				演算[Flops]

図1 階層的低ランク近似法における演算とメモリの トレードオフ

2.1 解析的低ランク近似

図1に示す階層的低ランク近似法における演算と メモリのトレードオフの一端には、純粋に解析的な FMM などの手法がある。これらの手法はマトリック スフリーの \mathcal{H}^2 行列と見ることもでき、高い演算密度 を有するため GPU や Xeon Phi などでも高い演算性 能を発揮することができる。演算密度が高い手法は多 く場合、演算量も多い。例えば、密行列積や密行列分 解などは高い演算密度を有するが、演算量は $\mathcal{O}(N^3)$ で ある。ここで述べる手法は同じ密行列の計算を $\mathcal{O}(N)$ もしくは $\mathcal{O}(Nlog^pN)$ で行うことができるにもかかわ らず、高い演算密度を有する手法である。

これらの手法が次世代計算機に適していることを示 す一例として、Treecode の大規模 GPU 並列実装であ る Bonsai コードが挙げられる [6]。Bonsai コードは ORNL の Titan において 18,600 GPU を用いた計算 で混合精度で 24.77 PFlops を達成した。ただし、この 高い演算性能はチャージが全て正であることを利用し た二重極の消滅と、四重極までで十分な精度が得られ るというアプリケーションの性質を利用した GPU レ ジスタの最適化が行われた結果であり、より高い精度 が必要とされる一般的な科学技術計算にはそのまま適 用できないことに注意されたい。

2.2 FMM の変換行列

図1で変換行列と呼んでいるのはFMM における多 重極展開を局所展開に変換する行列である。FMM の 計算時間の大半はこの多重極展開から局所展開への変 換に費やされるため、球面調和関数の回転 [7]、平面 波 [8] を用いた方法など様々なものが提案されている。 HSS 行列の表記法 [2] を用いると低ランク近似により 密行列は UBV の3つの行列に分解できる。ここで、 V が多重極展開、BV が局所展開、B が変換行列に相 当する。つまり、*B* 行列を保存するのが代数学的手法 (\mathcal{H}^2 ,HSS) であり、行列を保存せず毎回計算し直すの が解析的手法 (FMM) である。図1に示す「変換行列 の保存」は FMM でありながら変換行列 *B* を保存する 手法である。

ただし、FMM のように空間を等方的に分割するよ うな八分木を用いる場合、相対的な位置関係が同じ多 重極展開と局所展開の変換行列 B は同じになるという 性質があり、図1の「並進・回転対称性」はこれを利用 して B を重複して保存することを避ける手法を表して いる。相対的な位置関係の組み合わせは問題サイズ N に依存せず一定であり、八分木の階層間でも定数倍し て再利用できるため、FMM においては B を保存する ために必要なメモリは $\mathcal{O}(1)$ である。このため、FMM では B を保存したとしてもメモリ消費量は問題になら ない。 \mathcal{H}^2 行列や HSS 行列も階層化を行う前に密行列 を生成する関数に立ち返って幾何学情報を抽出し、計 算点の空間分割に規則性を持たせることができれば、 非対角成分に現れる B の多くが同じ値になることが予 想される。この性質をうまく利用することができれば \mathcal{H}^2 行列や HSS 行列のメモリ消費量を $\mathcal{O}(1)$ に低減で きる可能性がある。

図1の「変換行列の対角化」は多重極展開・局所 展開の基底に平面波[8]などを用いてBを対角化す ることでBVを求める際の演算量を低減する手法を 表している。上記の方法以外にもFMMの変換行列 の計算を高速化する方法として、局所的に多重極展 開や局所展開の次数を変化させるものがある。Very fast multipole method (VFMM)[9], error controlled FMM[10]などがこれにあたる。このような手法が有 効な理由は、FMMの変換行列を適用する多重極展開 と局所展開のペアの距離にばらつきがあり、遠いもの は多少次数を落として計算しても精度に影響がないこ とと、粒子の持つチャージの大きさに空間的なばらつ きがあり、値が小さいところでは次数を落としても全 体の精度に影響がないことである。

2.3 準解析的低ランク近似

初期の FMM は多重極展開から局所展開への変換行 列 B を解析的に求めるための式が存在するような問題 に用途が限定されていた。しかし、多重極展開や局所 展開を用いず、それと等価な作用を作り出す擬似粒子 を用いる「KIFMM」[3] や、それと等価な Chebyshev ノードを用いる「black-box FMM」[12] などでは、変 換行列 B を求める式は元の密行列を求めるための式 と同じになるため幅広い問題に適用することができ る。これは、解析的な FMM から代数学的な方向に また一歩近いづいた手法であると言える。このよう な擬似粒子の概念をより抽象化しものが interpolative decomposition[11] などのサンプリングを用いた手法 であり図1には「サンプリング」と表されている。サ ンプリングは「ランダム」性を利用した低ランク近似法 [14] と組み合わせることで高速かつ堅牢な手法となる。

図 1 の「変換行列の圧縮」は一度解析的に求めた変換 行列 *B* を SVD を用いて再圧縮する方法 [13] である。 これは SVD が多重極展開よりも低いランクまで圧縮 できることから、FMM 自体の高速化のために用いら れてきた手法であるが、*B* を求めるために代数学的な 処理を行っていることを考えると FMM よりは \mathcal{H}^2 行 列に近い手法である。

2.4 代数学的低ランク近似

代数学的低ランク近似は線形代数と数値解析の2分 野、アメリカとヨーロッパでそれぞれ独立に発展した きた経緯があり、類似した概念に異なる名称が付いて いるケースが散見される。表1に代数学的低ランク近 似法の代表的なものとそれぞれを3つの特性をもとに 分類したものを示す。BLR 以外の手法は全て階層的で あり、非対角成分を階層的にブロック化する。許容条 件 (admissibility) に「strong」とあるものは非対角成 分の分割をより細かく行い、「weak」の場合は非対角 成分の分割が粗い。FMM の用語を用いると、strong admissibility は隣り合うセルとは低ランク近似しな い、weak admissibility は隣り合うセルとは低ランク近 似することを意味する。表1にある「基底のネスト」 とは、階層的に表された行列の上層の大きいブロック の基底 U,V をより下層の小さいブロックの基底から 求めることをいう。FMM との対応関係でいうと、基 底をネストしない場合は $\mathcal{O}(NlogN)$ の Treecode に なる。

表 1 を上から見ていくと、 \mathcal{H} -matrix[15] は階層 的で許容条件は strong、基底はネストされていな い。この基底をネストしたものが \mathcal{H}^2 -matrix[1] で ある。HODLR[16] と HSS[2] は \mathcal{H} -matrix と \mathcal{H}^2 matrix の許容条件をそれぞれ weak にしたものであ る。BLR[17] は階層的でなく行列を単にプロック化 し、それぞれに低ランク近似を行う。

表1 代数学的低ランク近似の分類

手法	階層的	許容条件	基底のネスト
\mathcal{H} -matrix [15]	yes	strong	no
\mathcal{H}^2 -matrix [1]	yes	strong	yes
HODLR $[16]$	yes	weak	no
HSS[2]	yes	weak	yes
BLR [17]	no	weak	no



図 2 2 次元 Laplace 方程式の Green 関数から得ら れる密行列の行列・ベクトル積にかかる時間

表1に示した分類は行列のブロック化の構造に関す るもので、低ランク近似の手法はこれとは独立に選定 することができる。もっとも単純な方法としては特異 値分解 (SVD) があるが、より高速に低ランク近似を行 う手法は多数提案されている。例えば、QR 分解やLU 分解において適切なピボティングを行うことで低ラ ンク近似を行うことができる。Rank-revealing QR は Chan ら [18] によって提案され、適切なピボティングは Guら [19] によって示された。Rank-releaving LU は Chan ら [20] によって提案され、適切なピボティングは Miranian ら [21] によって示された。pseudo-skeletal 法 [22] や ACA[5] などは最小ランク近似を保証しな い代わりに高速な手法提案されている。ただし、これ らの手法は適切なピボティングが行われず低ランク近 似に失敗することもある [23]。ACA の改良版である HCA[24] は最小ランク近似を保証しながらも ACA と 同じ計算コストを有する。これ以外にも interpolative decomposition (ID)[11] などのサンプリングを用いつ つ条件数の良い基底行列を作る方法も提案されており、 これらの手法をランダマイズ手法 [14] と組み合わせる ことで高速で安定な低ランク近似法を実現することが できる。

3 FMM と HSS の直接比較

前述のように、FMM と \mathcal{H}^2 ・HSS 行列は共通の問 題に適用できる手法であるにも関わらず、これまで直 接的な計算時間の比較は少ない。特に階層的低ランク 近似の分野では前節で述べたように類似した手法が 多数提案されているが、手法間の優位性が明確に示さ れていない。ハイパフォーマンスコンピューティング の観点からこれらの手法の高性能実装同士の直接比較 は重要な課題であるといえる。ここでは、FMM の高 性能実装である exaFMM と HSS の高性能実装である STRUMPACK を同等の問題に関して同等の計算機上で 行う。

問題としては 2 次元の Lapalce 方程式の Green 関 数から得られる密行列の行列・ベクトル積を選定した。 計算環境としては 12 コアの Ivy Bridge (E5-2695v2) を用いて、まずは1コアのみを使用した状態で比較を 行った。図 2 にこのときの FMM と HSS の実行時間 を示す。横軸は密行列のサイズ $N \times N$ における N を とり、縦軸には計算時間を秒単位で示す。FMM には 一定のオーバーヘッドがあるため、N が小さい時には 実行時間が理論的な $\mathcal{O}(N)$ に従わない挙動を示してい るが、Nが一定以上大きくなるとFMM, HSS ともに $\mathcal{O}(N)$ の挙動を示している。ただし、FMM は同じサ イズの問題に関しては HSS よりも 1000 倍近く高速で あることが分かる。この大きな違いは、HSS の代数学 的な低ランク近似に要する時間が FMM の解析的な低 ランク近似に要する時間に比べてはるかに長いためで ある。

4 結言

階層的低ランク近似法を解析的な FMM から代数 学的な H²・HSS 行列まで統一的な視点から分析・比 較を行った。得られた知見は以下の通り要約できる。 FMM で多用される並進・回転対称性を HSS 行列に 応用することでメモリ消費量を大幅に低減できる可 能性がある。FMM も今では逆行列を求めたり Green 関数以外の関数にも適用できるため、対象となるアプ リケーションは HSS 行列に近づきつつある。また、 FMM を HSS 行列の低ランク近似法として用いるこ ともできる。FMM と HSS 行列の直接比較を行った場 合、FMM の方が 1000 倍高速であるという結果が得ら れた。

謝辞

本研究は JSPS 科研費 16H05859,16H02827 の助成を 受けたものです。

参考文献

- W. Hackbusch, B. Khoromskij, and S. A. Sauter. O(N) H²-Matrices. In H. Bungartz, R. Hoppe, and C. Zenger, editors, Lectures on Applied Mathematics. Springer-Verlag (2000)
- [2] S. Chandrasekaran, P. Dewilde, M. Gu, W. Lyons, and T. Pals, A Fast Solver for HSS Representations via Sparse Matrices. SIAM Jour-

nal on Matrix Analysis and Applications, Vol. 29, No. 1, pp. 67–81 (2006)

- [3] L. Ying, G. Biros, and D. Zorin, A kernelindependent Adaptive Fast Multipole Algorithm in Two and Three Dimensions. Journal of Computational Physics, Vol. 196, No. 2, pp. 591–626 (2004)
- [4] S. Ambikasaran and E. Darve, The Inverse Fast Multipole Method, arXiv:1407.1572v1 (2014)
- [5] M. Bebendorf, Approximation of Boundary Element Matrices. Numerische Mathematik, Vol. 86, pp. 565–589 (2000)
- [6] J. Bédorf, E. Gaburov, M. S. Fujii, K. Nitadori, T. Ishiyama, and S. Portegies Zwart, 24.77 Pflops on a Gravitational Tree-code to Simulate the Milky Way Galaxy with 18600 GPUs. In Proceedings of the 2014 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (2014)
- [7] C. A. White and M. Head-Gordon, Rotating Around the Quartic Angular Momentum Barrier in Fast Multipole Method Calculations, The Journal of Chemical Physics, Vol. 105, No. 12, pp. 5061–5067 (1996)
- [8] L. Greengard and V. Rokhlin. A New Version of the Fast Multipole Method for the Laplace Equation in Three Dimensions, Acta Numerica, Vol. 6, pp. 229–269 (1997)
- [9] H. G. Petersen, D. Soelvason, J. W. Perram, and E. R. Smith, The Very Fast Multipole Method, The Journal of Chemical Physics, Vol. 101, No. 10, pp. 8870–8876 (1994)
- [10] H. Dachsel, Corrected Article: "An Errorcontrolled Fast Multipole Method ", The Journal of Chemical Physics, Vol. 132, pp. 119901 (2010)
- [11] H. Cheng, Z. Gimbutas, P. G. Martinsson, and V. Rokhlin, On the Compression of Low Rank Matrices, SIAM Journal on Scientific Computing, Vol. 26, No. 4, pp. 1389–1404 (2005)
- [12] W. Fong and E. Darve. The Black-Box Fast Multipole Method. Journal of Computational Physics, Vol. 228, pp. 8712–8725 (2009)
- [13] Z. Gimbutas and V. Rokhlin, A Generalized Fast Multipole Method for Nonoscillatory Kernels, SIAM Journal on Scientific Computing,

Vol. 24, No. 3, pp. 796-817 (2002)

- [14] E. Liberty, F. Woolfe, P. G. Martinsson, V. Rokhlin, and M. Tygert, Randomized Algorithms for the Low-Rank Approximation of Matrices. PNAS, Vol. 104, No. 51, pp. 20167– 20172 (2007)
- [15] W. Hackbusch, A Sparse Matrix Arithmetic Based on *H*-Matrices, Part I: Introduction to *H*-Matrices, Computing, Vol. 62, pp. 89–108 (1999)
- [16] S. Ambikasaran and E. Darve, An $\mathcal{O}(N \log N)$ Fast Direct Solver for Partial Hierarchically Semi-separable Matrices, Journal of Scientific Computing, Vol. 57, pp. 477–501 (2013)
- [17] P. Amestoy, C. Ashcraft, O. Boiteau, A. Buttari, J.-Y. L ' Excellent, and C. Weisbecker, Improving Multifrontal Methods by Means of Block Low-Rank Representations, SIAM Journal on Scientific Computing, Vol. 37, No. 3, pp. A1451–A1474 (2015)
- [18] T. F. Chan, Rank Revealing QR Factorizations, Linear Algebra and its Applications, Vol. 88/89, pp. 67–82 (1987)
- [19] M. Gu and S. C. Eisenstat, Efficient Algorithms for Computing a Strong Rank-revealing QR Factorization, SIAM Journal on Scientific Computing, Vol. 17, No. 4, pp. 848–869 (1996)
- [20] T. F. Chan, On the Existence and Computation of LU-factorizations with Small Pivots, Mathematics of Computation, Vol. 42, No. 166, pp. 535–547 (1984)
- [21] L. Miranian and M. Gu, Strong Rank Revealing LU Factorizations, Linear Algebra and its Applications, Vol. 367, pp. 1–16 (2003)
- [22] S. A. Goreinov, E. E. Tyrtyshnikov, and N. L. Zamarashkin. A theory of pseudoskeleton approximations. Linear Algebra and its Applications, Vol. 261, No. 1-3, pp. 1–21 (1997)
- [23] S. Börm, L. Grasedyck, and W. Hackbusch, Introduction to Hierarchical Matrices with Applications, Engineering Analysis with Boundary Elements, Vol. 27, No. 405–422 (2003)
- [24] S. Börm and L. Grasedyck, Hybrid Cross Approximation of Integral Operators, Numerische Mathematik, Vol. 101, pp. 221–249 (2005)