An N-Body Solver for Square Root Iteration

Matt Challacombe,* Terry Haut,† and Nicolas Bock‡

Theoretical Division, Los Alamos National Laboratory

We develop the Sparse Approximate Matrix Multiply n-body solver for first order Newton Schulz iteration of the matrix square root and inverse square root. The solver performs an n-body occlusion-cull, yeilding a bounded relative error in the matrix-matrix product and reduced complexity for problems with structured metric decay. This complexity reduction cooresponds to the hierarchical resolution of algebraic structures within recursive volume of the product, a consequence of metric locality. For square root iteration, strongly localized sub-volumes are culled about plane-diagonals and along the cube-diagonal, cooresponding to resolution of the identity.

The main contributions of this paper are bounds on the `SpAMM` product and demonstration of a new algebraic locality that develops in these sub-volumes with strongly contractive identity iteration. This contraction cooresponds to the deflation of sub-volumes onto plane diagonals of the resolvent, and to a stronger bound on the `SpAMM` product.

Also, we carry out a first order Frëchet analyses for single and dual channel instances of the square root iteration, and look at bifurcations due to ill-conditioning and a too-agresive `SpAMM` approximation. Then, we show that extreme `SpAMM` approximations and strongly contractive identity iteration can be achieved through iterated regularization, and demonstrate the potential for orders of magnitude acceleration with product representation of the inverse factor.

I. INTRODUCTION

In many areas of application, long range, high value correlations lead to matrix equations with decay properties. By decay, we mean an approximate inverse relationship between matrix elements and an associated distance; this may be a simple inverse relationship between matrix elements and the Cartesian distance between corresponding support functions, or it may involve a non-Euclidean distance, *e.g.* a generalized measure between character strings in a training library [].

Matrix equations with decay have history and recent development in the statistics and statistical physics litterature [1–5]. Also recently, methods for meshfree interpolation are demonstrating remarkable predictive power through delocalized correlations and cooresponding ill-conditioned matrix equations with extreme slow decay [], a problem equivalent to the LCAO Gaussian basis problem in quantum chemistry []. Generally, local support functions are correlated through Lowdin's symmetric orthogonalization based on the matrix inverse square root [6, 7], yeilding representation independent matrix equations. In electronic structure, important long-range correlations manifest in slow decay properties of the gap shifted matrix sign function, as projector of the effective Hamiltonian (Fig. I). Both of these matrix problem with decay, the matrix sign function and the matrix inverse square root, are related by Higham's identity:

$$\mathrm{sign}\left(\begin{bmatrix} 0 & \boldsymbol{s} \\ \boldsymbol{I} & 0 \end{bmatrix}\right) = \begin{bmatrix} 0 & \boldsymbol{s}^{1/2} \\ \boldsymbol{s}^{-1/2} & 0 \end{bmatrix}. \quad (1)$$

The theory and computation of these matrix functions is given in Higham's reference [8].

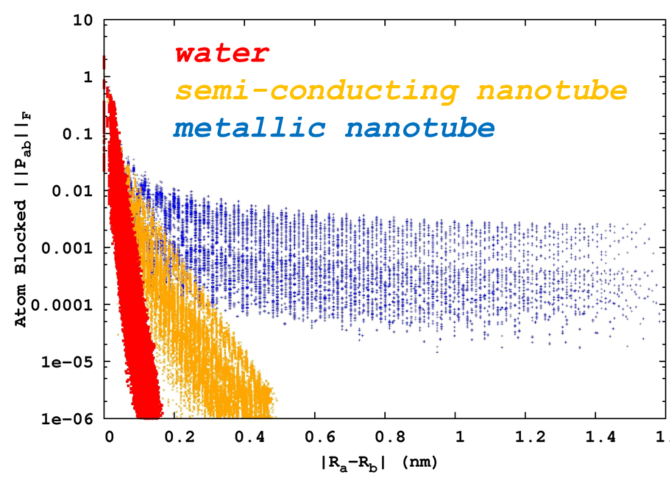

FIG. 1: Examples from electronic structure of decay for the spectral projector (gap shifted sign function) with respect to local (atomic) support. Shown is decay for systems with correlations that are short (insulating water), medium (semi-conducting 4,3 nanotube), and long (metalic 3,3 nanotube) ranged, from exponential (insulating) to algebraic (metallic).

A well conditioned matrix $\boldsymbol{s}$ may often correspond to matrix sign and inverse square root functions with rapid exponential decay, and be amenable to *ad hoc* matrix truncation or "sparsification", $\bar{\boldsymbol{s}} = \boldsymbol{s} + \boldsymbol{\epsilon}^{\boldsymbol{s}}_\tau$, where $\boldsymbol{\epsilon}^{\boldsymbol{s}}_\tau$ is the error introduced according to some criterion τ. The criterion τ might be a drop-tolerence, $\epsilon^{\boldsymbol{s}}_\tau = \{-s_{ij} * \hat{\boldsymbol{e}}_i \,|\, |s_{ij}| <$

*Electronic address: `matt.challacombe@freeon.org`; URL: `http://www.freeon.org`
†Electronic address: `haut@lanl.gov`
‡Electronic address: `nicolasbock@freeon.org`; URL: `http://www.freeon.org`

$\tau\}$, a radial cutoff, $\epsilon_\tau^{\boldsymbol{s}} = \{-s_{ij} * \hat{\boldsymbol{e}}_i \,|\, \|\boldsymbol{r}_i - \boldsymbol{r}_j\| > \tau\}$, or some other approach to truncation, perhaps involving a sparsity pattern chosen *a priori* for computational expedience. Then, the sparse general matrix-matrix multiply (`SpGEMM`) [9–12] may be employed, yielding fast solutions for multiplication rich iterations with fill-in modulated by truncation. Comprehensive surveys of these methods in the numerical linear algebra are given by Benzi [13, 14], and by Bowler [15] and Benzi [16] for electronic structure.

Often however, matrix truncation is ineffective for ill-conditioned problems, because of slow decay, and because of increased numerical sensitivities to poorly controled (absolute) truncation errors, *e.g.* in the matrix-product:

$$\overline{\boldsymbol{a}\cdot\boldsymbol{b}} \;=\; \boldsymbol{a}\cdot\boldsymbol{b} \;+\; \boldsymbol{\epsilon}_\tau^{\boldsymbol{a}}\cdot\boldsymbol{b} \;+\; \boldsymbol{a}\cdot\boldsymbol{\epsilon}_\tau^{\boldsymbol{b}} \;+\; \mathcal{O}(\tau^2)\,. \qquad (2)$$

An alterative approach is to find a reduced rank approximation closed under the opperations of interest [**?**]. However, compression to a reduced rank may be expensive if the rank is not much much smaller than the dimension. Both of these methods, truncation and rank reduction, are focused on matrix data as the target for compresion. In this contribution, our target for compression is instead the matrix product itself. For problems with decay, we show that an underlying metric locality, together with a new form of algebraic locality, can lead to complexity reduction under contractive iteration and the n-body occlusion-cull. The organization of this paper follows.

Contents

I. Introduction 1

II. Sparse Approximate Matrix Multiplication (`SpAMM`) 2
- A. Occlusion-Cull 2
- B. Bound 3
- C. Related Research 3

III. First Order Newton-Shulz Iteration 4
- A. Sign Iteration 4
- B. Square Root Iteration 4
- C. Mapping 5

IV. Error Flows in Square Root Iteration 5
- A. Stability 5
- B. Frëchet Derivatives 5
- C. Displacements 5
- D. Most Approximate but Still Stable 6

V. Regularization 7
- A. A Product Representation 7
- B. Effective by one order 7

VI. Locality 8
- A. Spatial, metric and temporal locality 9
- B. Algebraic locality 9
- C. Complexity reduction 9

VII. Summary 9

VIII. Conclusions 11

A. Implementation 11

B. Data 12

C. TEMPERARY:NOTES FOR REGULARIZATION SECTION 12

References 12

II. SPARSE APPROXIMATE MATRIX MULTIPLICATION (`SpAMM`)

In this contribution, we consider an N-body approach to the approximation of matrix functions with decay, based on the quadtree data structure [**? ?**]

$$\boldsymbol{a}^i = \begin{bmatrix} \boldsymbol{a}_{00}^{i+1} & \boldsymbol{a}_{01}^{i+1} \\ \boldsymbol{a}_{10}^{i+1} & \boldsymbol{a}_{11}^{i+1} \end{bmatrix}, \qquad (3)$$

and orderings that are locality preserving []. Orderings that preserve data locality are well developed and ***generic*** in the database theory [], providing fast spatial and metric queries. Locality enabled, fast data access is central to b-body approximations [], and an important problem for enterprise [] and runtime systems [], moreso with memory hierarchies becoming increasingly asynchronous and decentralized [**?**]. For matrices with decay, orderings that preserve locality lead to blocked-by-magnitude matrix structures with well segregated neighborhoods, inhabited by matrix elements of like size, and efficiently resolved by the quadtree data structure [].

A. Occlusion-Cull

The Sparse Approximate Matrix Multiply (`SpAMM`) carries out occlusion-culling to find only the most important sub-volumes in an approximate matrix product. `SpAMM` has evolved from a row-coloumn oriented skipout mechanism within the BCSR and DBCSR structures [], to hierachical approaches based on the quadtree and related to the occlusion-culling found in advanced mechanics and graphics methodologies [], with occlusion the avoidence of unessesary tree-work and culling the collection of significant tasks. Here, we ammend the `SpAMM` occlusion-cull with the recursion:

$$\boldsymbol{a}^i \otimes_\tau \boldsymbol{b}^i = \begin{cases} \emptyset \quad \texttt{if} \quad \|\boldsymbol{a}^{\mathtt{i}}\| \|\boldsymbol{b}^{\mathtt{i}}\| < \tau \|\mathtt{a}\| \|\mathtt{b}\| \\ \boldsymbol{a}^i \cdot \boldsymbol{b}^i \quad \texttt{if}(\mathtt{i} = \mathtt{leaf}) \\ \begin{bmatrix} \boldsymbol{a}^{i+1}_{00} \otimes_\tau \boldsymbol{b}^{i+1}_{00} + \boldsymbol{a}^{i+1}_{01} \otimes_\tau \boldsymbol{b}^{i+1}_{10} \,, & \boldsymbol{a}^{i+1}_{00} \otimes_\tau \boldsymbol{b}^{i+1}_{01} + \boldsymbol{a}^{i+1}_{01} \otimes_\tau \boldsymbol{b}^{i+1}_{11} \\ \boldsymbol{a}^{i+1}_{00} \otimes_\tau \boldsymbol{b}^{i+1}_{01} + \boldsymbol{a}^{i+1}_{01} \otimes_\tau \boldsymbol{b}^{i+1}_{11} \,, & \boldsymbol{a}^{i+1}_{00} \otimes_\tau \boldsymbol{b}^{i+1}_{01} + \boldsymbol{a}^{i+1}_{01} \otimes_\tau \boldsymbol{b}^{i+1}_{11} \end{bmatrix} \quad \texttt{else} \end{cases} , \tag{4}$$

which bounds the relative occlusion error

$$\frac{\|\boldsymbol{\Delta}^{a\cdot b}_\tau\|}{n^2} \;\leq\; \tau \, \|\boldsymbol{a}\| \, \|\boldsymbol{b}\| \,, \tag{5}$$

that occurs in the approximate product

$$\widetilde{\boldsymbol{a} \cdot \boldsymbol{b}} \;\equiv\; \boldsymbol{a} \otimes_\tau \boldsymbol{b} \;=\; \boldsymbol{a} \cdot \boldsymbol{b} + \boldsymbol{\Delta}^{a\cdot b}_\tau \,, \tag{6}$$

where $\|\cdot\| \equiv \|\cdot\|_F$ is a sub-multiplicative norm [].

B. Bound

We now prove (5).

Proposition 1. *Let $\tau_{A,B} = \tau \|A\| \|B\|$. Then for each i, j,*

$$\left| (A \otimes_\tau B)_{ij} - (A \cdot B)_{ij} \right| \leq n \tau_{A,B},$$

and

$$\|A \otimes_\tau B - A \cdot B\|_F \leq n^2 \tau_{A,B}.$$

Proof. We first show the following technical result: it is possible to choose $\alpha_{lij} \in \{0, 1\}$ such that

$$(A \otimes_\tau B)_{ij} = \sum_{l=1}^{n} A_{il} B_{lj} \alpha_{lij}, \tag{7}$$

In addition, if $\alpha_{lij} = 0$, then $|A_{il}| \, |B_{lj}| < \tau_{A,B}$. . To show this, we use induction on the number $k_{\max}$ of levels.

First, if $k_{\max} = 0$,

$$A \otimes_\tau B = \begin{cases} 0 & \text{if } \|A\|_F \|B\|_F < \tau_{A,B}, \\ A \cdot B & \text{else.} \end{cases}$$

Therefore, $A \otimes_\tau B$ is of the form (7) with either all $\alpha_{lij} = 0$ or all $\alpha_{lij} = 1$. Moreover, if $\alpha_{lij} = 0$, then $|A_{il}| \, |B_{lj}| \leq \|A\|_F \|B\|_F < \tau_{A,B}$.

Now assume that the claim holds for $k_{\max} - 1$. We show that it holds for $k_{\max}$. Indeed, if $\|A\|_F \|B\|_F < \tau_{A,B}$, we have that $A \otimes_\tau B = 0$, which is of the form (7) with all $\alpha_{lij} = 0$. Also, if $\alpha_{lij} = 0$, then $|A_{il}| \, |B_{lj}| < \|A\|_F \|B\|_F < \tau_{A,B}$.

Now assume that $\|A\|_F \|B\|_F \geq \tau_{A,B}$. Then

$$A \otimes_\tau B = \begin{pmatrix} A_{11} \otimes_\tau B_{11} + A_{12} \otimes_\tau B_{21} & A_{11} \otimes_\tau B_{12} + A_{12} \otimes_\tau B_{22} \\ A_{21} \otimes_\tau B_{11} + A_{22} \otimes_\tau B_{21} & A_{21} \otimes_\tau B_{21} + A_{22} \otimes_\tau B_{22} \end{pmatrix}.$$

We need to consider four cases: $i \leq n/2$ and $j \leq n/2$, $i > n/2$ and $j > n/2$, $i > n/2$ and $j \leq n/2$, and, finally, $i > n/2$ and $j > n/2$. Since the analysis is similar for all four cases, we only consider $i \leq n/2$ and $j \leq n/2$. We have that

$$\begin{aligned} (A \otimes_\tau B)_{ij} &= (A_{11} \otimes_\tau B_{11} + A_{12} \otimes_\tau B_{21})_{ij} \\ &= \sum_{l=1}^{n/2} (A_{11})_{il} (B_{11})_{lj} \, \alpha^{(1)}_{lij} + \\ &\qquad \sum_{l=1}^{n/2} (A_{12})_{il} (B_{21})_{lj} \, \alpha^{(2)}_{lij} \\ &= \sum_{l=1}^{n} A_{il} B_{lj} \alpha_{lij}, \end{aligned}$$

where we used the induction hypothesis in the second equality.

Now suppose that $\alpha_{lij} = 0$ for some l. Then $\tilde{\alpha}^{(1)}_{lij} = 0$ if $l \leq n/2$ or $\tilde{\alpha}^{(2)}_{l-n/2,ij} = 0$ $l > n/2$. If, e.g., $\tilde{\alpha}^{(2)}_{l-n/2,ij} = 0$, then $|A_{il}| \, |B_{lj}| = \left| (A_{12})_{i,l-n/2} \right| \left| (B_{21})_{l-n/2,j} \right| < \tau_{A,B}$, where we used the induction hypothesis in the final inequality. The analysis for $l \leq n/2$ is similar, and the claim follows.

We can now finish the proof of Proposition 1. Indeed, by (7),

$$\begin{aligned} \left| (A \otimes_\tau B)_{ij} - (A \cdot B)_{ij} \right| &\leq \sum_{l=1}^{n} |A_{il} B_{lj}| \, |\alpha_{lij} - 1| \\ &= \sum_{\alpha_{lij}=0} |A_{il} B_{lj}| \,. \end{aligned}$$

In addition, if $\alpha_{lij} = 0$, then $|A_{il} B_{lj}| < \tau_{A,B}$ and the lemma follows.

□

C. Related Research

`SpAMM` is perhaps most closely related to the Strassen-like branch of fast matrix multiplication [17, 18]. In the Strassen-like approach, disjoint volumes in (abstract) tensor intermediates are omitted recursively []. In the `SpAMM` approach to fast multixplication, the numerically most significant volumes in näive (ijk) tensor intermediates are culled, with error bounded by Eq. (5). This bound makes $\otimes_\tau$ a *stable* form of fast multiplication, as explained by Demmel, Dumitriu and Holz (DDH; Ref. [**?**]).

`SpAMM` is a n-body method for fast matrix multiplication, related to the generalized methods popularized by Grey [19, 20]. In our development, generalization reflects the *genericity* [] of recursive data access [], enabling range querries, metric querries, higher dimensional querries and so on, with common frameworks, structures and runtimes. So far, we have prototyped n-body solvers for mainstay problems in modern electronic structure theory [**? ?**], involving Fock exchange [], semi-local exchange-correlation functionals [], the Hartree (Coulomb) interaction [], matrix sign function [] and the matrix inverse square root (this work). This contribution is cornerstone for the simplification and evolution of these solvers.

Top-down n-body recursion and breadth-first map-reduction may be viewed as two sides of the same problem [**?**]. Emergent data frameworks and functional programming languages that support generic recursion and map skelitinizations may enable early exploitation of commodity (decentralized) concurence by scientific n-body solvers, as well as software sustainability [**?**]. For centralized, distributed architectures, n-body methods offer well established protocols for turning spatial and metric locality into data and temporal locality []. Recently, Driscoll *et. al* showed perfect strong scaling and communication optimality for pairwise n-body methods [**?**]. Bridging the gap between n-body solver and fast matrix multiplication, we recently demonstrated strong scaling for fast matrix multiplication (`SpAMM`) [].

This work offers a data local alternative to fast non-deterministic methods for sampling the product, which include sketching [21–26], joining [27–33], sensing [] and probing []. These methods involve a wheighted (probablistic) and on the fly sampling, with the potential for complexity reduction in the case of random distributions. `SpAMM` also employs on the fly wheighted sampling, but with compresion through locality, brought about by algebraic correlations (towards identity) and also in the metric structure, through strong Euclidean locality.

Finally, previous work on the scaled NS iteration has heavily influenced this work. Formost is Higham, Mackey, Mackey and T (HMMT; Ref. []) masterwork on convergence of NS iteration under all groups, wherein HMMT also develop Frëchet analyses for single square root iteration at the fixed point. Also, important inspiration comes from Chen and Chow's [] approach to scaled NS iteration for ill-conditioned problems [], and from the Helgaker groups work on NS iteration, whose notation we follow in part [].

III. FIRST ORDER NEWTON-SHULZ ITERATION

There are two common, first order NS iterations; the sign iteration and the square root iteration, related by the square, $\boldsymbol{I}\left(\cdot\right) = \mathrm{sign}^2\left(\cdot\right)$. These equivalent iterations converge linearly at first, then enter a basin of stability marked by super-linear convergence.

A. Sign Iteration

For the NS sign iteration, this basin is marked by a behavioral change in the difference $\delta \boldsymbol{X}_k = \widetilde{\boldsymbol{X}}_k - \boldsymbol{X}_k = \mathrm{sign}\left(\boldsymbol{X}_{k-1} + \delta\boldsymbol{X}_{k-1}\right) - \mathrm{sign}\left(\boldsymbol{X}_{k-1}\right)$, where $\delta\boldsymbol{X}_{k-1}$ is some previous error. The change in behavior is associated with the onset of idempotence and the bounded eigenvalues of $\mathrm{sign}'\left(\cdot\right)$, leading to stable iteration when $\mathrm{sign}'\left(\boldsymbol{X}_{k-1}\right)\delta\boldsymbol{X}_{k-1} < 1$. Global perturbative bounds on this iteration have been derived by Bai and Demmel [34], while Byers, He and Mehrmann [] developed asymptotic bounds. The automatic stability of sign iteration is a well developed theme in Ref.[8].

B. Square Root Iteration

In this work, we are concerned with resolution of the identity []

$$\boldsymbol{I}\left(\boldsymbol{s}\right) = \boldsymbol{s}^{1/2} \cdot \boldsymbol{s}^{-1/2}\,, \tag{8}$$

and its low-complexity computation with fast methods.

Starting with eigenvalues rescaled to the domain $(0,1]$ with the easily obtained largest eigenvalue, $\boldsymbol{s} \leftarrow \boldsymbol{s}/s_{N-1}$, and with $\boldsymbol{z}_0 = \boldsymbol{I}$ and $\boldsymbol{x}_0 = \boldsymbol{y}_0 = \boldsymbol{s}$, the cooresponding canonical, "dual" channel square root iteration is:

$$\begin{aligned} \boldsymbol{y}_k \;&\leftarrow\; h_\alpha\left[\boldsymbol{y}_{k-1}\cdot\boldsymbol{z}_{k-1}\right]\cdot\boldsymbol{y}_{k-1} \\ \boldsymbol{z}_k \;&\leftarrow\; \boldsymbol{z}_{k-1}\cdot h_\alpha\left[\boldsymbol{y}_{k-1}\cdot\boldsymbol{z}_{k-1}\right]\;, \end{aligned} \tag{9}$$

converging as $\boldsymbol{y}_k \to \boldsymbol{s}^{1/2}$, $\boldsymbol{z}_k \to \boldsymbol{s}^{-1/2}$ and $\boldsymbol{x}_k \to \boldsymbol{I}$, with eigenvalues aggregated towards 0 or 1 by the NS map $h_\alpha[\boldsymbol{x}] = \frac{\sqrt{\alpha}}{2}\left(3 - \alpha\boldsymbol{x}\right)$ []. As in the case of sign iteration, this cannonical iteration was shown by Higham, Mackey, Mackey and Tisseur [35] to remain strongly bounded in the superlinear regime, by idempotent Frechet derivatives about the fixed point $\left(\boldsymbol{s}^{1/2}, \boldsymbol{s}^{-1/2}\right)$, in the direction $\left(\delta\boldsymbol{y}_{k-1}, \delta\boldsymbol{z}_{k-1}\right)$:

$$\delta\boldsymbol{y}_k \;=\; \frac{1}{2}\delta\boldsymbol{y}_{k-1} - \frac{1}{2}\boldsymbol{s}^{1/2}\cdot\delta\boldsymbol{z}_{k-1}\cdot\boldsymbol{s}^{1/2} \tag{10}$$

$$\delta\boldsymbol{z}_k \;=\; \frac{1}{2}\delta\boldsymbol{z}_{k-1} - \frac{1}{2}\boldsymbol{s}^{-1/2}\cdot\delta\boldsymbol{y}_{k-1}\cdot\boldsymbol{s}^{-1/2}\,. \tag{11}$$

In addition to the dual channel instance, we also consider the "single" channel version of square root iteration,

$$\begin{aligned} \boldsymbol{z}_k \;&\leftarrow\; \boldsymbol{z}_{k-1}\cdot h_\alpha\left[\boldsymbol{x}_{k-1}\right]\;, \\ \boldsymbol{x}_k \;&\leftarrow\; \boldsymbol{z}_k^\dagger\cdot\boldsymbol{s}\cdot\boldsymbol{z}_k\;. \end{aligned} \tag{12}$$

C. Mapping

$\otimes_\tau$ is also a map. atm, stocastic unrpredictable, but really deterministic. excersize Control

In this contribution, we consider another aspect of convergence, namely the (hopefully) linear approach towards stability of the iteration

$$\widetilde{\boldsymbol{x}}_k \leftarrow \widetilde{\boldsymbol{y}}_k\left(\widetilde{\boldsymbol{x}}_{k-1}\right) \otimes_\tau \widetilde{\boldsymbol{z}}_k\left(\widetilde{\boldsymbol{x}}_{k-1}\right) , \tag{13}$$

made difficult by ill-conditioning and a sketchy $\otimes_\tau$.

Then, using

$$h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right] = h_\alpha\left[\boldsymbol{x}_{k-2}\right] + h'_\alpha \delta \boldsymbol{x}_{k-2} \tag{14}$$

Initially, h'_α at the smallest eigenvalue x_0 controls the rate of progress towards idempotence. As recently shown by Jie and Chen [36], for very ill-conditioned problems, a factor of two reduction in the number of NS steps can be achieved by choosing $\alpha \sim 2.85$, which is at the edge of stability. As argued by Pan and Schreiber [37], Jie and Chen [36], switching or damping the scaling factor towards $\alpha = 1$ at convergence is important, shifting emphasis away from the behavior of x_0 towards *e.g.* $x_i \in [0.01, 1]$, emphasizing overall convergence of the broad distribution [**?**]. In an approximate algebra like `SpAMM`, the potential for eigenvalues to fluctuate out of the domain of convergence must be considered. This is addressed in Section **??**.

IV. ERROR FLOWS IN SQUARE ROOT ITERATION

A. Stability

Stability in the square root iteration is determined by the differential

$$\delta \boldsymbol{x}_k = \boldsymbol{x}_{\delta \widehat{\boldsymbol{y}}_{k-1}} \times \delta y_{k-1} + \boldsymbol{x}_{\delta \widehat{\boldsymbol{z}}_{k-1}} \times \delta z_{k-1} + \mathcal{O}(\tau^2) \tag{15}$$

which must remain bounded below one to avoid divergence. The cooresponding Frëchet derivatives are

$$\boldsymbol{x}_{\delta \widehat{\boldsymbol{y}}_{k-1}} = \lim_{\tau \to 0} \frac{\boldsymbol{x}\left(\boldsymbol{y}_{k-1} + \tau \delta \widehat{\boldsymbol{y}}_{k-1}, \boldsymbol{z}_{k-1}\right) - \boldsymbol{x}_k}{\tau} \tag{16}$$

and

$$\boldsymbol{x}_{\delta \widehat{\boldsymbol{z}}_{k-1}} = \lim_{\tau \to 0} \frac{\boldsymbol{x}\left(\boldsymbol{y}_{k-1}, \boldsymbol{z}_{k-1} + \tau \delta \widehat{\boldsymbol{z}}_{k-1}\right) - \boldsymbol{x}_k}{\tau} , \tag{17}$$

along unit directions of the previous errors $\delta \widehat{\boldsymbol{y}}_{k-1}$ and $\delta \widehat{\boldsymbol{z}}_{k-1}$, by an ammount determined by the displacements $\delta y_{k-1} = \|\delta \boldsymbol{y}_{k-1}\|$ and $\delta z_{k-1} = \|\delta \boldsymbol{z}_{k-1}\|$. In the single instance, we have simply:

$$\delta \boldsymbol{x}_k = \boldsymbol{x}_{\delta \widehat{\boldsymbol{z}}_{k-1}} \times \delta z_{k-1} + \mathcal{O}(\tau^2) . \tag{18}$$

This formulation makes plain changes about the resolvent, seperating orientational effects of the directional derivatives, set mostly by the underlying exact linear algebra, from changes to error displacements, which involve both the action of derivatives on previous errors, as well as the `SpAMM` occlusion errors local to the product.

B. Frëchet Derivatives

In the dual instance, Frëchet derivatives occuring in Eq. (15) are:

$$\begin{aligned} \boldsymbol{x}_{\delta \widehat{\boldsymbol{z}}_{k-1}} = \boldsymbol{y}_{k-1} \cdot h'_\alpha \delta \widehat{\boldsymbol{z}}_{k-1} \cdot \boldsymbol{y}_{k-1} \cdot \boldsymbol{z}_k \\ + \boldsymbol{y}_k \cdot \delta \widehat{\boldsymbol{z}}_{k-1} \cdot h_\alpha\left[\boldsymbol{x}_{k-1}\right] \\ + \boldsymbol{y}_k \cdot \boldsymbol{z}_{k-1} \cdot \boldsymbol{y}_{k-1} \cdot h'_\alpha \delta \widehat{\boldsymbol{z}}_{k-1} , \end{aligned} \tag{19}$$

and

$$\begin{aligned} \boldsymbol{x}_{\delta \widehat{\boldsymbol{y}}_{k-1}} = h_\alpha\left[\boldsymbol{x}_{k-1}\right] \cdot \delta \widehat{\boldsymbol{y}}_{k-1} \cdot \boldsymbol{z}_k \\ + h'_\alpha \delta \widehat{\boldsymbol{y}}_{k-1} \cdot \boldsymbol{z}_{k-1} \cdot \boldsymbol{y}_{k-1} \cdot \boldsymbol{z}_k \\ + \boldsymbol{y}_k \cdot \boldsymbol{z}_{k-1} \cdot h'_\alpha \delta \widehat{\boldsymbol{y}}_{k-1} \cdot \boldsymbol{z}_{k-1} . \end{aligned} \tag{20}$$

Closer to the fixed point orbit, $\boldsymbol{y}_k \cdot \boldsymbol{z}_{k-1} \to \boldsymbol{I}$, $\boldsymbol{y}_{k-1} \cdot \boldsymbol{z}_k \to \boldsymbol{I}$, $h_\alpha\left[\boldsymbol{x}_k\right] \to \boldsymbol{I}$ and $h'_\alpha \to -\frac{1}{2}$ [**?**]. Then,

$$\boldsymbol{x}_{\delta \widehat{\boldsymbol{y}}_{k-1}} \to \delta \widehat{\boldsymbol{y}}_{k-1} \cdot \left(\boldsymbol{z}_k - \boldsymbol{z}_{k-1}\right) \tag{21}$$

and

$$\boldsymbol{x}_{\delta \widehat{\boldsymbol{z}}_{k-1}} \to \left(\boldsymbol{y}_k - \boldsymbol{y}_{k-1}\right) \cdot \delta \widehat{\boldsymbol{z}}_{k-1} . \tag{22}$$

Likewise, in the single channel instance:

$$\begin{aligned} \boldsymbol{x}_{\widehat{\boldsymbol{z}}_{k-1}} \to \left(\boldsymbol{z}_k - \boldsymbol{z}_{k-1}\right)^\dagger \cdot \boldsymbol{s} \cdot \delta \widehat{\boldsymbol{z}}_{k-1} \\ + \delta \widehat{\boldsymbol{z}}^\dagger_{k-1} \cdot \boldsymbol{s} \cdot \left(\boldsymbol{z}_k - \boldsymbol{z}_{k-1}\right) . \end{aligned} \tag{23}$$

About the fixed point then, error flow in the $\boldsymbol{y}_k$ and the $\boldsymbol{z}_k$ channels is tightly quenched, cooresponding to $\boldsymbol{x}_{\delta \widehat{\boldsymbol{x}}_{k-1}} \to \boldsymbol{I}$ and identity iteration [**?**].

C. Displacements

Countering oritentational convergence, determined almost entirely by the underlying exact itations, is the the compounding displacement error, determined by `SpAMM` occlusion in each of three products, at each step, and also involving previous errors. Here, we look at just the displacement $\delta \boldsymbol{z}_{k-1}$, which has the largest potential for divergence as we argue here and show numerically in the following section.

Including the `SpAMM` error in the $\widetilde{\boldsymbol{z}}_{k-1}$ update we have:

$$\begin{aligned} \widetilde{\boldsymbol{z}}_{k-1} &= \widetilde{\boldsymbol{z}}_{k-2} \otimes_\tau h_\alpha[\widetilde{\boldsymbol{x}}_{k-2}] \\ &= \Delta_\tau^{\widetilde{\boldsymbol{z}}_{k-2} \cdot h_\alpha[\widetilde{\boldsymbol{x}}_{k-2}]} + \widetilde{\boldsymbol{z}}_{k-2} \cdot h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right] . \end{aligned} \tag{24}$$

Then, with $h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right] = h_\alpha\left[\boldsymbol{x}_{k-2}\right] + h'_\alpha \delta \boldsymbol{x}_{k-2}$, and taking $\boldsymbol{z}_{k-1}$ from both sides, we find

$$\begin{aligned} \delta \boldsymbol{z}_{k-1} = \Delta_\tau^{\widetilde{\boldsymbol{z}}_{k-2} \cdot h_\alpha[\widetilde{\boldsymbol{x}}_{k-2}]} \\ + \delta \boldsymbol{z}_{k-2} \cdot h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right] + \boldsymbol{z}_{k-2} \cdot h'_\alpha \delta \boldsymbol{x}_{k-2} , \end{aligned} \tag{25}$$

which is bounded by

$$\delta z_{k-1} < \|\boldsymbol{z}_{k-2}\| \left(\tau\, \sigma_n \, \|h_\alpha \left[\widetilde{\boldsymbol{x}}_{k-2}\right]\| + h'_\alpha \delta y_{k-2} \|z_{k-2}\|\right) + \delta z_{k-2} \left(\|h_\alpha \left[\widetilde{\boldsymbol{x}}_{k-2}\right]\| + \|y_{k-2}\|\right) . \quad (26)$$

In Eq. (26), the term $h'_\alpha \delta y_{k-2} \|\boldsymbol{z}_{k-2}\|^2$ is volitile, tending towards $\delta y_{k-2}\, \kappa(\boldsymbol{s})/2$. Because of this sensitivity, and because the $\boldsymbol{y}_k$ product channel maintains fidelity of the starting eigen-basis, we single out this "sensitive" product for a higher level of precision; $\tau_s \ll \tau$. Still, we expect different behavior from the single instance $\widetilde{\boldsymbol{y}}_{k-1} = \widetilde{\boldsymbol{z}}^\dagger_{k-1} \otimes_{\tau_s} \boldsymbol{s}$, and the dual instance $\widetilde{\boldsymbol{y}}_{k-1} = h_\alpha[\widetilde{\boldsymbol{x}}_{k-1}] \otimes_{\tau_s} \widetilde{\boldsymbol{y}}_{k-1}$. This is because the spectral product is broader (resolving larger and smaller numbers) in the single instance and narrower in the dual instance.

D. Most Approximate but Still Stable

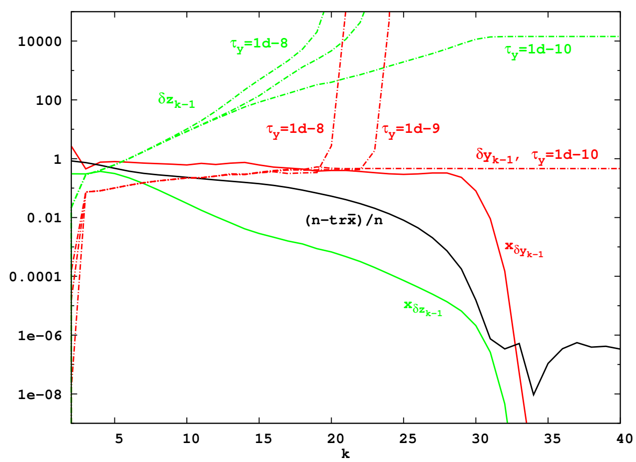

FIG. 2: Derivatives, displacements and trace error of the unscaled dual iteration. Derivatives are full lines, whilst the displacements cooresponding to $\tau_s = \{10^{-8}, 10^{-9}, 10^{-10}\}$ are the dashed lines. The trace error is shown as a full black line.

Experiments were carried out on the ill-conditioned ($\kappa(\boldsymbol{s}) = 10^{10}$) nanotube metric-matrices described in Appendix B. We picked $\tau = .001$ and block size $b = 64$. Then, we looked at stability with respect to the tighter τ_s threshold, the directional derivatives and the trace error.

In Fig. 2, unscaled results for the dual instance are shown. In Fig. 4, scaled results for the dual instance are given, recaiming approximatly 2/3 of the available 2× acceleration, with a 1/3 penalty due to the stabilization map described in Section III C. Then, in Fig. **??** we show results for the scaled single instance.

Not shown is complete fillin at convergence for even the most-approximate-yet-still-stable (MAYSS) value of τ_s. Also, with a less forgiving stability map, we find interesting left/right differences; namely, the right first

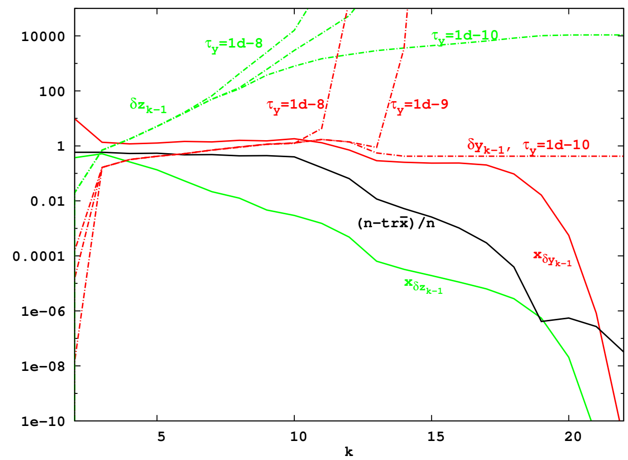

FIG. 3: Derivatives, displacements and the approximate trace of the scaled dual iteration. Derivatives are full lines, whilst the displacements cooresponding to $\tau_s = \{10^{-8}, 10^{-9}, 10^{-10}\}$ are the dashed lines. The trace error is shown as a full black line.

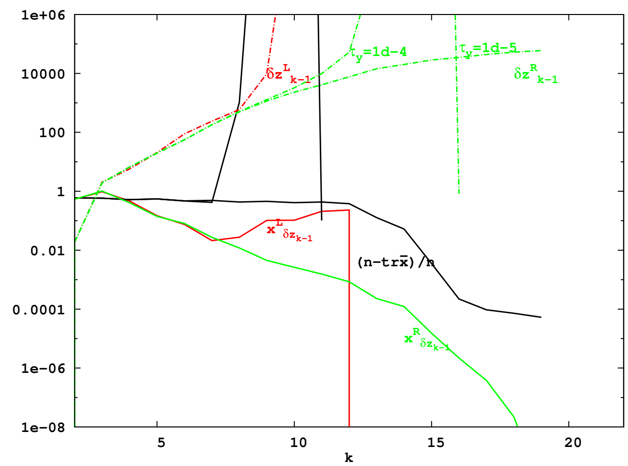

FIG. 4: Derivatives, displacements and the approximate trace of the scaled single iteration. Derivatives are full lines, whilst the displacements cooresponding to $\tau_s = \{10^{-3}, 10^{-4}, 10^{-5}\}$ are the dashed lines. The trace error is shown as a full black line.

product

$$\widetilde{\boldsymbol{x}}^R_k \leftarrow \widetilde{\boldsymbol{z}}^\dagger_k \otimes_\tau \left(\boldsymbol{s} \otimes_{\tau_s} \widetilde{\boldsymbol{z}}_{k-1}\right) , \quad (27)$$

is different from the left first product

$$\widetilde{\boldsymbol{x}}^L_k \leftarrow \left(\widetilde{\boldsymbol{z}}^\dagger_k \otimes_{\tau_s} \boldsymbol{s}\right) \otimes_\tau \widetilde{\boldsymbol{z}}_{k-1} , \quad (28)$$

which is unsuprising. The stability map parameters discussed Section III C are tuned away from such senstivities, and their local variation has a neglibible impact on stability. Also not shown, the intermediate volumes (before fill in) behave very differently with scaling; in all cases, the product volumes, proportional to the computatonal cost, are significantly increased by scaling. In the remainder of this work, we won't look at scaling again.

V. REGULARIZATION

Even for the most approximate but stable (MAYSS) approximations, our nanotube calculations become dense in both the data and task domains, even for very large 128× U.C. systems. For these ill-conditioned iterations, broad spectral resolutions that involving products of large and small numbers coorespond to delocalized products that are not tightly bound by the `SpAMM` approximation. However, as we show later in Fig. 12, similarly ill-conditioned problems may achieve substantial compression with just the MAYSS approximation.

A systematic way to reduce ill-conditioning is through Tikhonov regularization []. Regularization involes a small level shift of the eigenvalues, $\boldsymbol{s}_\mu \leftarrow \boldsymbol{s} + \mu \boldsymbol{I}$, altering the condition number of the shifted matrix to $\kappa(\boldsymbol{s}_\mu) = \frac{\sqrt{s_{N-1}^2+\mu^2}}{\sqrt{s_0^2+\mu^2}}$ [].

Achieving substantial acceleration with severe ill-conditioning may require a large level shift however, producing inverse factors of little practical use. One approach to recover a more accurate inverse factor is Riley's method based on Taylor's series [];

$$\boldsymbol{s}^{-1/2} = \boldsymbol{s}_\mu^{-1/2} \cdot \left(\boldsymbol{I} + \frac{\mu}{2} \boldsymbol{s}_\mu^{-1} + \frac{3\mu^2}{8} \boldsymbol{s}_\mu^{-2} + \ldots \right) \, . \quad (29)$$

For severely ill-conditioned problems and large level shifts, this expansion may converge very slowly. Also, adding powers of the full inverse may not be computationally effective.

A. A Product Representation

We introduce an alternative representation of the regularized inverse factor;

$$\boldsymbol{s}^{-1/2} \equiv \bigotimes_{\substack{\tau=\tau_0 \\ \mu=\mu_0}} |\, \tau\, \mu \, ; \boldsymbol{s}^{-1/2} \rangle \, , \quad (30)$$

which is a telescoping product of preconditioned "slices" starting with a most-approximate-yet-still-effective-by-one-order (MAYEBOO) preconditioner, $\boldsymbol{s}_{\tau_0 \mu_0}^{-1/2} \equiv |\, \tau_0\, \mu_0 \, ; \boldsymbol{s}^{-1/2} \rangle^1$. This sandwich of generic, thinly sliced `SpAMM` products allows to construct a nested scoping on precision, via τ, and in the condition number, controled by μ.

[1] Braket notation marks the potential for assymetries in the intermediate represenation.

This scheme rests on three factors: (**1**) the ability to find a most agressive but still effective first slice, corrective by one order in the condition, $\mu_0 = .1$, and by one order in the precision, $\tau_0 = .1$ (MAYEBOO), (**2**) an optimal $\tau\,\mu$ scoping in Eq. (35) and (**3**) optimized maps for generic thin slices. In the following, we demonstrate a first MAYEBOO preconditioner for the ill-conditioned nano-tube problem, and then we sketch a naive approach to building the full inverse factor.

B. Effective by one order

We look again at the $\kappa(\boldsymbol{s}) = 10^{10}$ nanotube series described in Appendix B, this time with extreme regularization, $\mu_0 = .1$, and at a finer granularity, $b = 8$. Culled $\boldsymbol{y}_k$ and $\boldsymbol{z}_k$ volumes (as percentage of the total work) for $36 - 128\times$ the (3,3) unit cell are shown for the MAYEBOO approximation in Fig. V B for the single instance, and in Fig. V B for the dual instance.

The behavior of these implementations is very different; in the single instance, a stable iteration could not be found at precision $\tau_0 = .1$. Also, disturbingly, the single iteration with inflating cull-volumes.

On the other hand, volume of the dual iteration is strongly contracted with resolution of the identity.

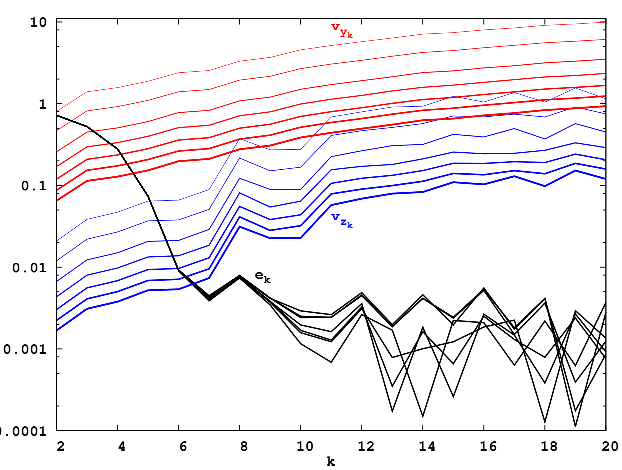

FIG. 5: Culled volumes in the thin slice, single instance approximation of $\boldsymbol{s}_{\tau_0 \mu_0}^{-1/2}$ for the (3,3) nanotube, $\kappa(\boldsymbol{s}) = 10^{10}$ matrix series described in Section B. In the "single" instance, it was not possible to achieve stability with $\tau_0 = .1$ In this "single" case, a thin slice cooresponds to $\mu_0 = .1, \tau_0 = 10^{-2}$ & $\tau_s = 10^{-4}$, and volumes are $\mathrm{v}_{\widetilde{\boldsymbol{z}_\mathrm{k}}} = \left(\mathrm{vol}_{\widetilde{\boldsymbol{z}}_{\mathrm{k}-1} \otimes_\tau \mathrm{h}[\widetilde{\boldsymbol{x}}_{\mathrm{k}-1}]}\right) \times 100\%/\mathrm{N}^3$ and $\mathrm{v}_{\widetilde{\boldsymbol{y}}_k} = \left(\mathrm{vol}_{\boldsymbol{s} \otimes_{\tau_s} \widetilde{\boldsymbol{z}}_k}\right) \times 100\%/N^3$. Line width increases with increasing system size. Also shown is the trace error, $\mathrm{e}_k = \left(N - \mathrm{tr}\, \boldsymbol{x}_k\right)/N$.

These results reflect very different cull-spaces. In the single instance, the spectral resolution of powers is not compresive, and $\widetilde{\boldsymbol{y}}_k^{\mathrm{single}} \rightarrow \boldsymbol{s}_{\tau_0 \mu_0}^{-1/2} \otimes_{\tau_0} \boldsymbol{s}_{\mu_0}$ is poorly bound by Eq. (5). In the dual case however, $\widetilde{\boldsymbol{y}}_k^{\mathrm{dual}} \rightarrow$

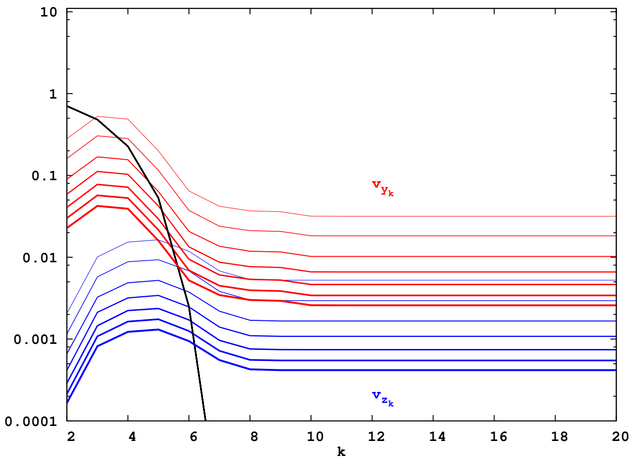

FIG. 6: Culled volumes in the thin slice, dual instance approximation of $\boldsymbol{s}^{-1/2}_{\tau_0\mu_0}$ for the (3,3) nanotube, $\kappa(\boldsymbol{s}) = 10^{10}$ matrix series described in Section B. The thin slice coresponds to $\mu_0 = .1, \tau_0 = .1$ & $\tau_s = .001$ with volumes $\mathrm{v}_{\widetilde{\boldsymbol{y}}_k} = \left(\mathrm{vol}_{h[\widetilde{\boldsymbol{x}}_{k-1}]\otimes_{\tau_s}\widetilde{\boldsymbol{y}}_k}\right)\times 100\%/N^3$ and $\mathrm{v}_{\widetilde{\boldsymbol{z}}_k} = \left(\mathrm{vol}_{\widetilde{\boldsymbol{z}}_{k-1}\otimes_\tau h[\widetilde{\boldsymbol{x}}_{k-1}]}\right)\times 100\%/N^3$. Line width increases with increasing system size. Also shown is the trace error, $\mathrm{e}_k = (N - \mathrm{tr}\,\boldsymbol{x}_k)/N$, which rapidly approaches 10^{-11} (not shown).

$\boldsymbol{I}_{\tau_0\mu_0}\otimes_{\tau_0}\boldsymbol{s}^{1/2}_{\tau_0\mu_0}$ and $\widetilde{\boldsymbol{z}}^{\mathrm{dual}}_k \rightarrow \boldsymbol{s}^{-1/2}_{\tau_0\mu_0}\otimes_{\tau_0}\boldsymbol{I}_{\tau_0\mu_0}$, with Eq. (5) tightening to

$$\Delta^{\boldsymbol{I}_{\tau_0\mu_0}\cdot\boldsymbol{s}^{1/2}_{\tau_0\mu_0}} < \tau n \|\boldsymbol{s}^{1/2}_{\tau_0\mu_0}\| \tag{31}$$

and

$$\Delta^{\boldsymbol{s}^{-1/2}_{\tau_0\mu_0}\cdot\boldsymbol{I}_{\tau_0\mu_0}} < \tau n \|\boldsymbol{s}^{-1/2}_{\tau_0\mu_0}\|\,, \tag{32}$$

as relative and absolute errors converge. This tightening is compressive, leading to complexities that are quadtree copy in place.

In the dual instance, the `SpAMM` approximation can be brought all the way to $\tau_0 = .1$ in the case of $\mu_0 = .1$. From this first slice $\boldsymbol{s}^{-1/2}_{\tau_0,\mu_0}$ then, a next level shifted preconditioner can be found, $\boldsymbol{s}^{-1/2}_{\tau_0\mu_1}$, based on the residual $\left(\boldsymbol{s}^{-1/2}_{\tau_0\mu_0}\right)^\dagger \otimes_{\tau_0} (\boldsymbol{s}+\mu_1\boldsymbol{I}) \otimes_{\tau_0} \boldsymbol{s}^{-1/2}_{\tau_0\mu_0}$, with *e.g.* $\mu_1 = .01$. It may then be possible to find the full (`SpAMM` most approximate) factor as the nested product of preconditioned thin slices;

$$\boldsymbol{s}^{-1/2}_{\tau_0} = \boldsymbol{s}^{-1/2}_{\tau_0\mu_n}\otimes_{\tau_0}\boldsymbol{s}^{-1/2}_{\tau_0\mu_{n-1}}\otimes_{\tau_0}\ldots\,\boldsymbol{s}^{-1/2}_{\tau_0\mu_0} \tag{33}$$

In this way, iterative regularization can be used to find a product representation of the inverse square root at a `SpAMM` resolution potentially far more permisive than otherwise possible. Likewise, it may be possible to obtain the full factor with increasing `SpAMM` resolution in the product representation:

$$\boldsymbol{s}^{-1/2} = \boldsymbol{s}^{-1/2}_{\tau_m}\otimes_{\tau_m}\boldsymbol{s}^{-1/2}_{\tau_{m-1}}\otimes_{\tau_{m-1}}\ldots\,\boldsymbol{s}^{-1/2}_{\tau_0} \tag{34}$$

taken over the sequence $1 > \tau_0 > \tau_1 > \cdots > \tau_n$. More generally,

$$\boldsymbol{s}^{-1/2} \equiv \bigotimes_{\substack{\tau=\tau_0\\ \mu=\mu_0}} |\,\tau\,\mu\,;\boldsymbol{s}^{-1/2}\rangle\,, \tag{35}$$

acknowledging the potential for a flexible path between precision and regularization. The braket notation marks the potential for assymetries in the intermediate represenation.

This most approximate (but effective) solution is (ideally) representative of one order in the precision, $\tau_0 \sim .1$, and corrective by one order in the condition, $\mu_0 \sim .1$, yeilding a thin, 0^{th} preconditioner, $\boldsymbol{s}^{-1/2}_{\tau_0\mu_0}$. This "thin" iteration may bring spectral resolution into alignment with norm magnitudes towards the resolvent $\boldsymbol{I}_{\tau_0\mu_0} \equiv \widetilde{\boldsymbol{I}}\,(\boldsymbol{s}_{\tau_0\mu_0})$, strengthening Eq. (5).

This thin product representation may have advantages: (**1**) Each thin solve involves a few generic and well behaved steps that may be narrowly optimized; (**2**) Each thin solve can be brought rapidly into compressive identity iteration; (**3**) The `SpAMM` bound is vastly strengthend, via Eqs. (31-32); (**4**) A new algebraic n-body form of locality is exploited; (**5**) The inverse factor can be applied incrementally; (**6**) Slice update and application is ammenable to continous temporal partitioning based on *e.g.* persistence data.

VI. LOCALITY

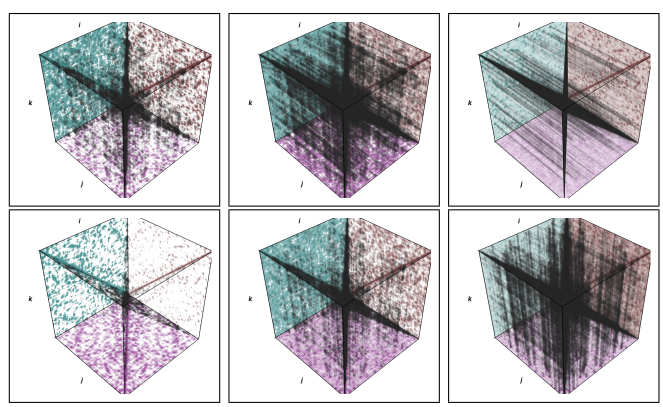

FIG. 7: The ijk task and data space for construction of the unregularized preconditioner $|\tau_0 = .001, \mu_0 = .0\,;\boldsymbol{s}^{-1/2}\rangle$, with dual instance square root iteration, and for the 6-311G** metric of 100 periodic water molecules at STP. At top its $\boldsymbol{y}_k = h_\alpha[\boldsymbol{x}_{k-1}]\otimes_{\tau_s}\boldsymbol{y}_{k-1}$ for $k = 0, 4, \&15$, while on the bottom we have $\boldsymbol{x}_k = \boldsymbol{y}_k\otimes_\tau\boldsymbol{z}_k$ for $k = 0, 4, \&15$. Maroon is $\boldsymbol{a}$, purple is $\boldsymbol{b}$, green is $\boldsymbol{c}$, and black is the volume $\mathrm{vol}_{a\otimes_\tau b}$ in the product $\boldsymbol{c} = \boldsymbol{a}\otimes_\tau\boldsymbol{b}$.

A. Spatial, metric and temporal locality

Astrophysical n-body algorithms employ range queries over spatial databases to hierarchically discover and compute approximations that commit only small errors. Often, these spatial databases are ordered with a space filling curve (SFC) [], which maps points that are close in space to an index where they are also close. The block-by-magnitude structures empowering the `SpAMM` approximation are *metric localities*; in quantum chemical examples they coorespond to an underlying SFC ordering of Carteisan coordinates.

Warren and Salmon showed how to parlay spatial locality into temporal locality, remapping and repartitioning the space filling curve to rebalance distributed n-body tasks, based on accumulated histories (persistence data). In a similar way, we showed how persistence can be used to achieve strong parallel scaling for `SpAMM` with commonly available runtimes []. Persistence data, providing temporal locality, may also be useful in mathematical approximation.

In Figure 7, we show $\otimes_\tau$ volumes for square root iteration, cooresponding to the metric of a small, periodic water box with the large, 6-311G** basis. For the 3-d periodic case, diminishing Cartesian seperations lead to long-skinny delocalizations (pillae) and much denser matrices, relative to *e.g.* a one-dimensional nano-tube. These delocalizations coorespond to weakness in Eq. (5), and to the tighter thresholds required to maintain a single iteration in the MAYSS approximation. This effect is even more pronounced in the single instance (not shown), where delocalizations are exagerated due to spectral resolutions that are broader. Eventually, Cartesian seperation will thin and trim the density of these delocalizations, leading to complexity reduction based on the effects of metric locality alone, as demonstrated in **??**.

B. Algebraic locality

In Figures 8 and 9, we show a new kind of locality that is uniquely exploited by n-body approximation of the square root iteration. This algebraic locality develops compressively towards convergence as the contractive identity iteration develops. We call this compression ***lensing***, involving collapse of the culled volume about plane diagonals of the resolvent. Lensing cooresponds to strengthening of Eq.(**??**) to yeild Eqs. (31) and (32), and strong convergence of the directional derivatives Eq.(**??**-**??**) to $\mathbf{0}$. This is an important, mitigating computational effect for the $\boldsymbol{y}_k$ channel that involves the tighter threshold, $\tau_s \sim 0.01 \times \tau$.

In addition, non-Euclidian measures are relevent for achieving metric locality in the `SpAMM` algebra, including information measures, space filling curve generalizations, as well as graph reorderings that envelope matrix elements about the diagonal [], a common approach in structural mechanics. In Figure 12 we show development of a first, unregularized (MAYSS) preconditioner for such an example; the structural matrix $\boldsymbol{s} =$ `bcsstk14` is a $\kappa(\boldsymbol{s}) = 10^{10}$ matrix cooresponding to the roof of the Omni Coliseum in Atlanta []. These results show remarkable gossamer sheeting and flattening along plane diagonals, at top for developmentent of $\boldsymbol{y}_k$, and hollow accumulation of $\mathrm{vol}_{\boldsymbol{y}_k \otimes_\tau \boldsymbol{z}_k}$ looking down at $\boldsymbol{y}_k$ (along bottom). Surprisingly, this example shows lensing for a relatively tight MAYSS approximation, while the equally ill-conditioned & lower dimensional $\kappa(\boldsymbol{s}) = 10^{10}$ nanotube MAYSS approximation remains full (dense) through U.C. $\times 128$.

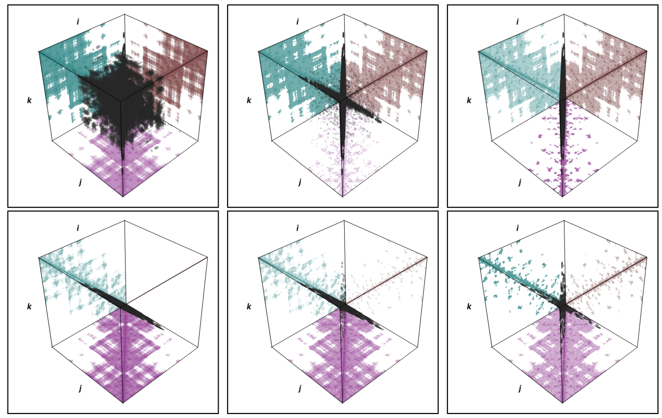

FIG. 8: The ijk task and data space for construction of the MAYEBOO preconditioner $|\tau_0 = .1, \mu_0 = .1\,; \boldsymbol{s}^{-1/2}\rangle$, with dual instance square root iteration, and for an $8\times$ U.C. (3,3) $\kappa(s) = 10^{11}$ nanotube. $\boldsymbol{y}_k$ appears wider than $\boldsymbol{z}_k$ because it is computed at a higher precision, $\tau_s = .001$, and because the first multiply involves $\boldsymbol{s}^2$. At top its $\boldsymbol{y}_k = h_\alpha[\boldsymbol{x}_{k-1}]\otimes_{\tau_s} \boldsymbol{y}_{k-1}$ for $k = 0, 4, \&16$, while on the bottom we have $\boldsymbol{x}_k = \boldsymbol{y}_k \otimes_\tau \boldsymbol{z}_k$ for $k = 0, 2, \&16$. Maroon is $\boldsymbol{a}$, purple is $\boldsymbol{b}$, green is $\boldsymbol{c}$, and black is the volume $\mathrm{vol}_{a\otimes_\tau b}$ in the product $\boldsymbol{c} = \boldsymbol{a}\otimes_\tau \boldsymbol{b}$.

C. Complexity reduction

Finally, we show complexity reduction at convergence of the MAYEBOO approximation relative to the MAYSS approximation, in Fig. 10 for periodic water boxes, and in Fig. 11 for the ill-conditioned nano-tube. The two-orders difference between $\boldsymbol{y}_k$ and $\boldsymbol{z}_k$ volumes cooresponds precicely to $\tau_s \sim \tau \times .01$, with $\boldsymbol{x}_k$ in between. Except for the slower trend in Fig. (10)'s $\boldsymbol{x}_k$ volume, we see the potential for continued strong acceleration with increasing system size. Understanding these subtlties is the subject of future work.

VII. SUMMARY

In this work, we developed the n-body solver `SpAMM` for square root iteration. Main contributions include a modified Cauchy-Schwarz criterion, Eq. **??**, and proof that the cooresponding relative product error is bound by

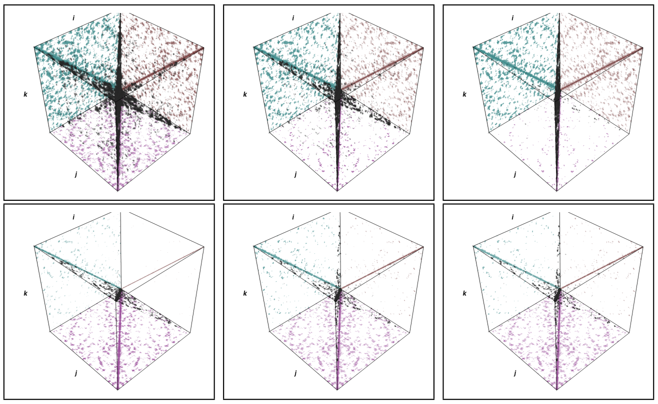

FIG. 9: The ijk task and data space for construction of the MAYEBOO preconditioner $|\tau_0 = .1, \mu_0 = .1\,; \boldsymbol{s}^{-1/2}\rangle$, with dual instance square root iteration and for 6-311G** metric of 100 periodic water molecules at STP. At top its $\boldsymbol{y}_k = h_\alpha[\boldsymbol{x}_{k-1}]\otimes_{\tau_s}\boldsymbol{y}_{k-1}$ for $k = 0, 4, \&15$, while on the bottom we have $\boldsymbol{x}_k = \boldsymbol{y}_k \otimes_\tau \boldsymbol{z}_k$ for $k = 0, 4, \&15$. Maroon is $\boldsymbol{a}$, purple is $\boldsymbol{b}$, green is $\boldsymbol{c}$, and black is the volume $\mathrm{vol}_{a\otimes_\tau b}$ in the product $\boldsymbol{c} = \boldsymbol{a}\otimes_\tau \boldsymbol{b}$.

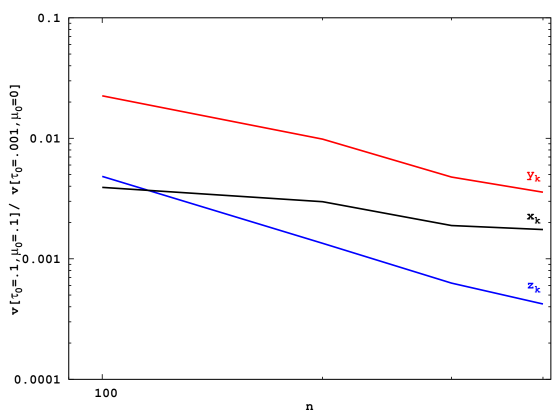

FIG. 10: Complexity reduction in metric square root iteration for periodic 6-311G** water. Shown is the ratio of lensed product volumes for the regularized most-approximate-yet-effective-by-one-order (MAYEBOO) approximation and the unregularized most-approximate-yet-still-stable (MAYSS) approximation.

Eq. (5). Also, we demonstrated a new kind of algebraic locality, lensing, that develops with strongly contractive identity iteration.

In Section **??**, we looked at stability leading to the basin of convergence and sensitivity of the three product channels $\boldsymbol{y}_k$, $\boldsymbol{z}_k$ and $\boldsymbol{x}_k$, for the `SpAMM` approximation in the cannonical "dual" instance, Eq. (**??**), and for the "single" instance, Eq. (**??**). Consistent with HMMT [], the $\boldsymbol{z}_k$ channel is sensitized by the full inverse, $\boldsymbol{s}^{-1}$, requiring a tighter threshold for that case, $\tau_s \ll \tau$. Later, we find that extra cost is strongly mitigated by lensing. Also in Section **??**, we looked at bifurcations of scaled and unscaled iterations for ill-conditioned systems, towards a most-approximate-yet-still-stable (MAYSS) preconditioner, and disected competing effects at the edge of stability, between compounding displacement magnitudes and strongly convergent directional derivatives.

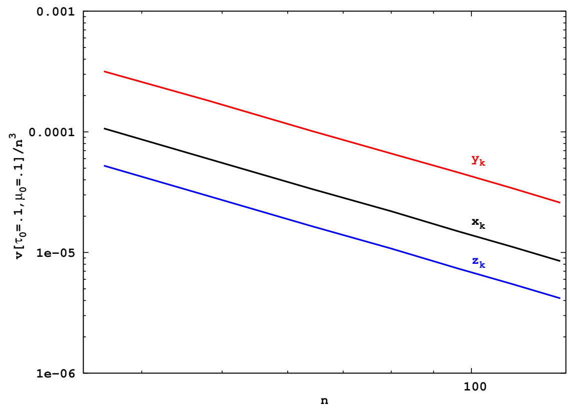

FIG. 11: Complexity reduction in metric square root iteration for periodic 6-311G** water. Shown is the ratio of lensed product volumes for the regularized most-approximate-yet-effective-by-one-order (MAYEBOO) approximation and the unregularized most-approximate-yet-still-stable (MAYSS) approximation.

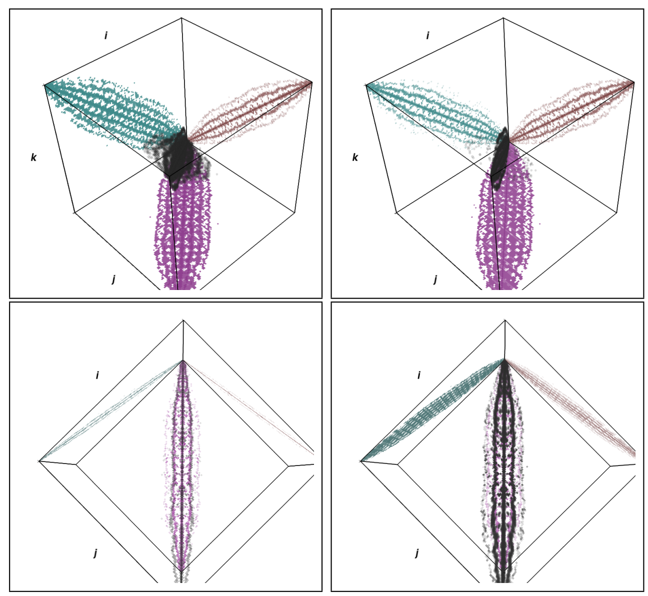

FIG. 12: The ijk task and data space for construction of the unregularized preconditioner $|\tau_0 = .001, \mu_0 = .0\,; \boldsymbol{s}^{-1/2}\rangle$, with the dual instance of square root iteration and for 6-311G** metric of 100 periodic water molecules at STP. At top its $\boldsymbol{y}_k = h_\alpha[\boldsymbol{x}_{k-1}]\otimes_{\tau_s}\boldsymbol{y}_{k-1}$ for $k = 0, 4, \&15$, while on the bottom we have $\boldsymbol{x}_k = \boldsymbol{y}_k \otimes_\tau \boldsymbol{z}_k$ for $k = 0, 4, \&15$. Maroon is $\boldsymbol{a}$, purple is $\boldsymbol{b}$, green is $\boldsymbol{c}$, and black is the volume $\mathrm{vol}_{a\otimes_\tau b}$ in the product $\boldsymbol{c} = \boldsymbol{a}\otimes_\tau \boldsymbol{b}$.

OK, WELL ITS HARD TO WRITE THIS WITH THAT PART BEING A BIT OF A MESS ... In Section **??**, we introduced iterated regularization for ill-conditioning, and developed a product representation of thin the potential for and found large differences between single and dual channel instances. For the regularized dual channel instance, we demonstrated iterations with volumes strongly contractive towards convergence, and we showed that this contraction cooresponds to Eq. (5) tightenging significantly. strong contractive iterations how the full inverse factor can e proved that at convergence, be achieved by products of generic, regularized, well conditioned and increasingly more accurate solutions; $s^{1/2} = \bigotimes_{\mu}^{\tau} |\mu, \tau\, ; s^{-1/2}\rangle$ cooresponding to a first most-approximate-yet-effective-by-one-order (MAYEBOO) preconditioner $|\mu = .1, \tau = .1\, ; s^{-1/2}\rangle$. We looked at the MAYEBOO approximation for both the single and dual instances, and found that even with the most permisive regularization, spectral resolution in the single instance is too broad to achieve strong lensing.

Finally, we looked at the MAYSS and MAYBOO approximations for periodic water systems, for ill-conditioned nanotubes, and for the ill-conditioned structural matrix `bcsstk14` for the Omni Coliseum in Atlanta []. For the problem of large basis periodic water systems, we find a MAYBOO/MAYSS volume compression of two to three orders. For the problem of an ill-conditioned nano-tube, we find a MAYBOO/MAYSS volume compression of ... Remarkably, the ill-conditioned `bcsstk14` was able to achieved a strongly lensed state in the MAYSS approximation, with remarkable gossamer sheeting and flattening along plane diagonals, and hollow, reticulate volumes of the resolvent.

VIII. CONCLUSIONS

This work is gauged against other methods for fast matrix multiplication discussed in Section **??**. Against `SpMM`, the n-body approach offers a bounded control over relative errors in the product and the ability to resolve complex algebraic structures, about plane-diagonals of the ijk-cube, along tall-skiny pillae and for volumetric contractions to lower dimensional objects via lensing. The n-body method uniquely and synnergistically exploits two distinct forms of locality, metric locality cooresponding to a Cartesian or non-Euclidean decay principle, and algebraic locality cooresponding to contractive identity iteration. Also, strong parallel scaling for the $\mathcal{O}(n)$ electronic structure problem has been demonstrated with the `SpAMM` kernel, a feature that remains elusive for methods based on `SpMM` [**?**].

Against methods for matrix compression [], as well as against Fast Matrix Multiplication of the Strassen type [], the quadtree data structure and the octree task space employed by `SpAMM` are entirely complimentary. In the case of sketch products [], persistence data can be used to identify and characterize pillae resulting from broad spectral resolutions; then, it may be possible to deploy streaming approaches for these tall-skinny delocalizations []. Thus, the concurent application of fast methods for matrix multiplication may be enabled by the database framework supporting n-body approximations.

Beyond the fast matrix multiply, n-body frameworks may enable additional, layered functionalities and economizations in complex solver ecosystems, with facile interopperability and mathematical agility, through generic recursion and skelitinization, and with common runtimes able to exploit temporal and data localities. For example, we recently generalized `SpAMM` recursion to the problem of Fock-exchange, with a recursive triple (hextree) metric querry on the Almlof-Alhrichs direct SCF criteria []. Also, mathematical equivalence with the matrix sign function, Eq. (**??**), and close structural relationships with the polar decomposition may enable to extend functionality of the n-body iterations developed here.

Despite these compeling features and related xxx, n-body square-root iteration must be gauged by its ability to compute a high quality inverse factor. Here, we have only looked at complexity and stability of the most approximate preconditioners; the most-approximate-yet-still-stable (MAYSS) approximation and the regularized most-approximate-yet-effective-by-one-order (MAYEBOO) approximations. However, these results are encouraging, showing the potential for several to many orders of magnitude reduction in complexity for the regularized approximation relative to the unregularized approximation, made possible by construction of a much lower precision preconditioner than would otherwize be possible, via Eq. (**??**), and also by opperations in the strongly contractive regime, under Eqs. (31-32). These and related preliminary results [] suggest that a `SpAMM` sandwich of thin, generic iterations may enable a competitive computational approach that avoids explicit computation of the ill-conditioned factor.

Appendix A: Implementation

FP, F08, OpenMP 4.0 In the current implementation, all persistence data (norms, flops, branches & *etc.*) are accumulated compactly in the backward recurrence. This persistence data that may be achieved by minimal locally essential trees [].

For these reasons, maintaining connection to the eigenvectors of s through a tighter first product is nessesary. In the single instance, and with a tighter "s" product, $\tau_s \ll \tau$, we find very interesting left/right differences; namely, the right first product

$$\widetilde{x}_k^R \leftarrow \widetilde{z}_k^\dagger \otimes_\tau \left(s \otimes_{\tau_s} \widetilde{z}_{k-1} \right) \ , \tag{A1}$$

is different from the left first product

$$\widetilde{x}_k^L \leftarrow \left(\widetilde{z}_k^\dagger \otimes_{\tau_s} s \right) \otimes_\tau \widetilde{z}_{k-1} \ . \tag{A2}$$

damping the inversion and the small value to be added c is called Marquardt-Levenberg coefficient

Map switching and etc based on TrX

Appendix B: Data

3,3 carbon nanotube with diffuse *sp*-function double exponential (Fig.)

Appendix C: TEMPERARY:NOTES FOR REGULARIZATION SECTION

The idea of preconditioning is that we can use a low tolerance τ_0 (e.g. $\tau_0 = 10^{-2}$) to cheaply obtain an approximation $R_0 \approx S^{-1/2}$. Then since $S_1 \equiv R_0 S R_0$ is close to the identity matrix I,

$$\|R_0 S R_0 - I\|_F \lesssim \tau_0,$$

we can use Newton Schulz on S_1 with a higher tolerance τ_1 to get an accurate approximation $R_1 = S_1^{-1/2}$ and using only a few iterations:

$$\|R_1 S_1 R_1 - I\|_F \lesssim \tau_1.$$

In particular, the matrix S_1, being close to the identity, is better conditioned than S and computing $S_1^{-1/2}$ requires much fewer Newton Schulz iterations. Moreover, since

$$\|(R_1 R_0)\, S\, (R_0 R_1) - I\|_F \lesssim \tau_1,$$

we see that $R_1 R_0$ is a τ_1 approximation to $S^{-1/2}$. Notice that, from the stability bound for SpAMM, we can replace all of the exact matrix multiplications with SpAMM multiplications.

To formalize this, let $S_0 = S$, and suppose that R_j is the approximation to $S_j^{-1/2}$ obtained via the Newton Schulz iteration with SpAMM tolerance τ_j, so that

$$\left\|R_j \otimes_{\tau_j} S_j \otimes_{\tau_j} R_j - I\right\|_F \lesssim \tau_j.$$

Then define $S_{j+1} = R_j \otimes_{\tau_K} S_j \otimes_{\tau_K} R_j$ and let R_{j+1} the approximation to $S_{j+1}^{-1/2}$ obtained via the Newton Schulz iteration with SpAMM tolerance τ_{j+1}, so that

$$\|R_{j+1} \otimes_{\tau_K} S_{j+1} \otimes_{\tau_K} R_{j+1} - I\|_F \lesssim \tau_{j+1}.$$

Then since $S_{j+1} = R_j \otimes_{\tau_K} S_j \otimes_{\tau_K} R_j$,

$$\|R_{j+1} \otimes_{\tau_K} (R_j \otimes_{\tau_K} S_j \otimes_{\tau_K} R_j) \otimes_{\tau_K} R_{j+1} - I\|_F \lesssim \tau_{j+1}.$$

In general, defining

$$R_{\text{left}} \equiv R_{j+1} \otimes_{\tau_K} R_j \otimes_{\tau_K} R_{j-1} \cdots \otimes_{\tau_K} R_0,$$

and

$$R_{\text{right}} \equiv R_0 \otimes_{\tau_K} R_1 \cdots \otimes_{\tau_K} R_j \cdots \otimes_{\tau_K} R_{j+1},$$

it follows by induction that

$$\|R_{\text{left}} \otimes_{\tau_K} S \otimes_{\tau_K} R_{\text{right}} - I\|_F \lesssim \tau_K.$$

Now, by stability of SpAMM,

$$\|R_{\text{left}} S R_{\text{right}} - I\|_F \lesssim \tau_K.$$

Also,

$$R_{j+1} \otimes_{\tau_K} R_j \otimes_{\tau_K} R_{j-1} \cdots \otimes_{\tau_K} R_0$$

can be written as

$$(R_0 \otimes_{\tau_K} R_1 \cdots \otimes_{\tau_K} R_j \cdots \otimes_{\tau_K} R_{j+1})^{\mathrm{T}} + \mathcal{O}(\tau_K).$$

Therefore, $R_{\text{left}} = R_{\text{right}}^{\mathrm{T}} + \mathcal{O}(\tau_K)$, and so

$$\left\|R_{\text{left}} S R_{\text{left}}^{\mathrm{T}} - I\right\|_F \lesssim \tau_K. \tag{C1}$$

We can therefore write the following symbolic representation

$$S^{-1/2} = S_{\tau_{j+1}}^{-1/2} \otimes_{\tau_{j+1}} S_{\tau_j}^{-1/2} \otimes_{\tau_j} S_{\tau_{j-1}}^{-1/2} \cdots \otimes_{\tau_1} S_{\tau_0}^{-1/2} + \mathcal{O}(\tau_{j+1}),$$

where $S_{\tau_k}^{-1/2}$ is a τ_k approximation to the inverse square root of $S_k = S_{\tau_{k-1}}^{-1/2} \otimes_{\tau_{k-1}} S_{k-1} \otimes_{\tau_j} S_{\tau_{k-1}}^{-1/2}$.

[1] O. Penrose and J. Lebowitz, Commun. Math. Phys. **184** (1974).

[2] J. Voit, *The Statistical Mechanics of Financial Markets*, Theoretical and Mathematical Physics (Springer Berlin Heidelberg, 2006), ISBN 9783540262893, URL `https://books.google.com/books?id=6zUlh\_TkWSwC`.

[3] L. Anselin, Int. Reg. Sci. Rev. **26**, 153 (2003), ISSN 01600176, URL `http://irx.sagepub.com/cgi/doi/10.1177/0160017602250972`.

[4] J. Hardin, S. R. Garcia, and D. Golan, Ann. Appl. Stat. **7**, 1733 (2013), ISSN 1932-6157, arXiv:1106.5834v4.

[5] I. Krishtal, T. Strohmer, and T. Wertz, Found. Comput. . . . (2013), ISSN 1615-3375.

[6] P. O. Lowdin, Advances in Physics **5**, 1 (1956).

[7] A. R. Naidu, p. 1 (2011), 1105.3571.

[8] N. J. Higham, *Functions of Matrices* (Society for Industrial & Applied Mathematics,, 2008).

[9] F. G. Gustavson, ACM Transactions on Mathematical Software (TOMS) **4**, 250 (1978).

[10] S. Toledo, IBM J. Res. Dev. **41**, 711 (1997).

[11] M. Challacombe, Comput. Phys. Commun. **128**, 93 (2000).

[12] D. R. Bowler and T. M. andM. J. Gillan, Comp. Phys. Comm. **137**, 255 (2000).

[13] M. Benzi and M. Tuma, Appl. Numer. Math. **30**, 305 (1999).

[14] M. Benzi, J. Comput. Phys. **182**, 418 (2002).
[15] D. R. Bowler and T. Miyazaki, Reports Prog. Phys. **75**, 36503 (2012).
[16] M. Benzi, P. Boito, and N. Razouk, SIAM Rev. **55**, 3 (2013).
[17] V. Strassen, Numer. Math. **13**, 354 (1969).
[18] G. Ballard, J. Demmel, O. Holtz, and O. Schwartz, Communications of the ACM **57**, 107 (2014).
[19] A. G. Gray and A. W. Moore, in *Advances in Neural Information Processing Systems* (MIT Press, 2001), vol. 4, pp. 521–527, URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.7138`.
[20] A. Gray (2003), URL `http://reports-archive.adm.cs.cmu.edu/anon/anon/2003/abstracts/03-222.html`.
[21] T. Sarlos, In Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS) (2006).
[22] P. Drineas, R. Kannan, and M. W. Mahoney, SIAM Journal on Computing **36**, 132 (2006).
[23] M. W. Mahoney, pp. 1–53 (2012), arXiv:1502.03032v1.
[24] R. Pagh, ACM Trans. Comput. Theory **5**, 1 (2013).
[25] J. V. T. Sivertsen (2014).
[26] D. P. Woodruff, **10**, 1 (2015), arXiv:1411.4357v3.
[27] P. Mishra and M. H. Eich, ACM Comput. Surv. **24**, 63 (1992).
[28] E. G. Hoel and H. Samet, in *Parallel Processing, 1994. ICPP 1994. International Conference on* (1994), vol. 3, pp. 227 –234.
[29] E. H. Jacox and H. Samet, ACM Trans. Database Syst. **28**, 230 (2003).
[30] S. Chen, A. Ailamaki, P. B. Gibbons, and T. C. Mowry, ACM Trans. Database Syst. **32** (2007).
[31] R. R. Amossen and R. Pagh, in *Proceedings of the 12th International Conference on Database Theory* (ACM, 2009), pp. 121–126.
[32] M. D. Lieberman, J. Sankaranarayanan, and H. Samet, in *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering* (IEEE Computer Society, Washington, DC, USA, 2008), pp. 1111–1120, ISBN 978-1-4244-1836-7.
[33] C. Kim, T. Kaldewey, V. W. Lee, E. Sedlar, A. D. Nguyen, N. Satish, J. Chhugani, A. Di Blas, and P. Dubey, Proc. VLDB Endow. **2**, 1378 (2009).
[34] Z. Bai and J. Demmel, SIAM J. Matrix Anal. Appl **19**, 205 (1998).
[35] N. J. . Higham, N. M. D . Steven Mackey, and F. . coise Tisseur, SIAM J. Matrix Anal. Appl. **26**, 849 (2005).
[36] J. Chen and E. Chow, Preprint ANL/MCS-P5059-0114, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439 (2014).
[37] V. Pan and R. Schreiber, SIAM J. Sci. Stat. Comput. (1991).