On Stability of the First Order Newton Schulz Iteration in an Approximate Algebra

Matt Challacombe,* Terry Haut, and Nicolas Bock†

Theoretical Division, Los Alamos National Laboratory

I. INTRODUCTION

In many areas of application, finite correlations lead to matrices with decay properties. By decay, we mean an approximate (perhaps bounded []) inverse relationship between matrix elements and an associated distance; this may be a simple inverse exponential relationship between elements and the Cartesian distance between support functions, or it may involve a generalized distance, *e.g.* a statistical measure between strings. In electronic structure, correlations manifest in decay properties of the gap shifted matrix sign function, as projector of the effective Hamiltonian (Fig. I). More broadly, matrix decay properties may correspond to statistical matrices [**? ? ? ? ?**], including learned correlations in a generalized, non-orthogonal metric []. More broadly still, problems with local, non-orthogonal support are often solved with congruence transformations of the matrix inverse square root [**? ?**] or a related factorization [**?**]; these transformations correlate local support with a representation independent form, *eg.* of the eigenproblem. Interestingly, the matrix sign function and the matrix inverse square root function are related by Higham's identity:

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

A complete overivew of matrix function theory and computation is given in Higham's enjoyable reference [**?**].

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 the sparse matrix approximation $\bar{\boldsymbol{s}} = \boldsymbol{s} + \boldsymbol{\epsilon}^{\boldsymbol{s}}_\tau$, where $\boldsymbol{\epsilon}^{\boldsymbol{s}}_\tau$ is the error introduced according to some criterion τ. Supporting this approximation are useful bounds to matrix function elements [**? ?**]. The criterion τ might be a droptolerence, $\epsilon^{\boldsymbol{s}}_\tau = \{-s_{ij} * \hat{\boldsymbol{e}}_i \,|\, |s_{ij}| < \tau\}$, a radial cutoff, $\epsilon^{\boldsymbol{s}}_\tau = \{-s_{ij} * \hat{\boldsymbol{e}}_i \,|\, \|\boldsymbol{r}_i - \boldsymbol{r}_j\| > \tau\}$, or some other approach to truncation, perhaps involving a sparsity pattern chosen *a priori*. Then, conventional computational kernels may be employed, such as the sparse general matrix-matrix multiply (`SpGEMM`) [**? ? ? ?**], yielding fast solutions for multiplication rich iterations and a modulated **(what do you mean with modulated?)** fill-in. These and related incomplete/inexact approaches to the computation of sparse approximate matrix functions often lead to $\mathcal{O}(n)$ algorithms, finding wide use in technologically important preconditioning schemes, the information sciences, electronic structure and many other disciplines. Comprehensive surveys of these methods in the numerical linear algebra are given by Benzi [**? ?**], and by Bowler [**?**] and Benzi [**?**] for electronic structure.

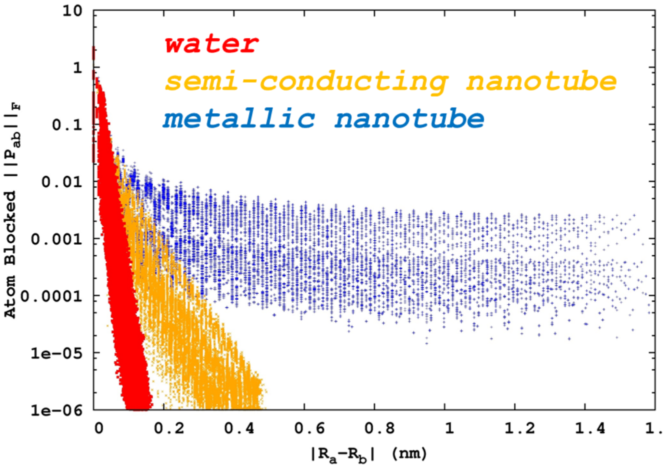

FIG. 1: Examples from electronic structure of decay for the spectral projector (gap shifted sign function) with respect to local (atomic) support. Shown is decay for systems with correlations that are short (insulating water), medium (semi-conducting 4,3 nanotube), and long (metalic 3,3 nanotube) ranged,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

Because the truncated multiplication is controlled only by absolute, additive errors in the product,

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

achieving sparse, stable and rapidly convergent iteration for ill-conditioned problems can be challenging []. In cases of extreme degeneracy, hierarchical semi-seperable (reduced rank) algorithms can offer effective complexity reduction []. However, many practical cases are somewhere in-between sparse and meaningfully degenerate regimes; effectively dense but without an exploitable reduction in rank. This is the case in electronic structure for strong but non-metallic correlation, *e.g.* towards the Mott transition [], and also in the case of local atomic support towards completeness [**? ? ?**].

*Electronic address: `matt.challacombe@freeon.org`; URL: `http://www.freeon.org`

†Electronic address: `nicolasbock@freeon.org`; URL: `http://www.freeon.org`

II. SPARSE APPROXIMATE MATRIX MULTIPLICATION

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

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

(I know we argued about this before, but I still like the normal 1-based notation better :)) and orderings that are locality preserving []. Orderings that preserve data locality are well developed in the database theory [], providing fast spatial and metric queries. Locality enabled, fast data access is central to the N-Body approximation [], and an important problem for enterprise [] and runtime systems [], 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. Stable SpAMM

With blocked-by-magnitude ordering of matrices $\boldsymbol{a}$ and $\boldsymbol{b}$, the Sparse Approximate Matrix Multiplication (SpAMM) kernel, $\otimes_\tau$, carries out fast occlusion culling of insignificant volumes in the product octree:

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

with errors linear in τ bounded by the sub-multiplicative norms $\|\cdot\| \equiv \|\cdot\|_F$ and the Cauchy-Schwarz inequality [**? ?**].

The approximate SpAMM product is

$$\widetilde{\boldsymbol{a} \cdot \boldsymbol{b}} \;\equiv\; \boldsymbol{a} \otimes_\tau \boldsymbol{b} \;=\; \boldsymbol{a} \cdot \boldsymbol{b} + \boldsymbol{\Delta}_\tau^{a \cdot b} \tag{5}$$

with the culled contractions $\boldsymbol{\Delta}_\tau^{a \cdot b}$ obeying the SpAMM bound

$$\|\boldsymbol{\Delta}_\tau^{a \cdot b}\| \;\leq\; \tau \, \|\boldsymbol{a}\| \, \|\boldsymbol{b}\| \, , \tag{6}$$

at each level of recursion. This makes $\otimes_\tau$ *stable*, as defined by Demmel, Dumitriu and Holz (DDH; Ref. [**?**]). However, instead of the roundoff error, we are concerned with a deterministic SpAMM error[1], which leads to a non-associative algebra and error flows with properties of the Lie bracket

$$\widetilde{[\boldsymbol{a}, \boldsymbol{b}]} \equiv \boldsymbol{a} \otimes_\tau \boldsymbol{b} - \boldsymbol{b} \otimes_\tau \boldsymbol{a} = [\boldsymbol{a}, \boldsymbol{b}] + \boldsymbol{\Delta}_\tau^{a \cdot b} - \boldsymbol{\Delta}_\tau^{b \cdot a} \, . \tag{7}$$

B. Related Research

Here, we note overlaping themes and research related to the results presented here. First, its worth noting that the SpAMM concept has evolved, from a row-coloum based occlusion [], to recursive octree occlusion [], and in this work, to the bounded (stable) occlusion satisfying Eq. 6. This evolution cooresponds to an ongoing economization and radical simplification of strongly interacting, high performance solvers in the `freeon` ecosystem []. These economizations include n-body algorithms for the five common electronic structure solvers [], and the (ongoing) development of generic strategies supporting rapid, hierarhcical access of spatial and metric data [**? ?**]. For example, SpAMM has been extended to the *triple* metric querry for hextree occlusion of the exact Fock exchange [].

SpAMM is broadly related to *generalized n-body* methods, poularized by Grey [], that are simply a reflection of *genericity.* algorithms based on fast (hierarchical) spatial and metric querry [**?**] together with local approximation []. Recently, x, y and Yellik showed perfect strong scaling and communication optimality for pairwise n-body methods [**? ?**]. Also Demmel and showed for the related fast matrix multiplicaiton. It may be possible to show similar results, based on the locality of reference developed in this work, encompassing both strong Euclidian locality and with algebraic (diagonal) locality towards identity.

As noted by Aluru [], the top-down n-body model and breadth-first map-reduction are equivalent [], offering the potential for alignment with emergent enterprise frameworks [] and functional programming languages that support generacity []. Language support for generic recur-

[1] A non-deterministic SpAMM occlusion error is also possible, *e.g.* associated with probabilistic stabilization [] or sampling [] methods.

sion may allow very complex solver ecosystems with very simple (skeletonized) frameworks, lowering barriers to entry, enhancing performance towards decentralized memory landscapes, and following a sustainable commodity trend [**?**] that offers increasingly cheap compute cycles over the next few decades [].

`SpAMM` is perhaps most closely related to the Strassen-like branch of fast matrix multiplication [], which has been on fire with recent new developments []. In the Strassen-like approach, disjoint volumes in (abstract) tensor intermediates are omitted recursively []. In the `SpAMM` approach to fast multixplication, a volume of significant contributions is culled from the naive ijk-cube of the intermediate contraction, with error bounded by Eq. 6. These approaches would seem to be compatible and synergystic.

`SpAMM` is also related to technologies for the non-deterministic, compresive sampling of the product. These technologies have also seen exciting developments, including sketching [**? ?**], joining, sensing and probing []. These methods involve a weighted (probablistic) and on the fly sampling with the potential for complexity reduction under certain assumptions (random distributions) [**is this true for all? what about MAD?**]. `SpAMM` also employs an on the fly wheighted sampling, based on the product of matrix norms, but opperating under the contrary assumption of compresion through locality in the naive ijk space, brought about via correlations in the algebra (towards identity) and in the underlying data (blocked-by-magnitude).

Methods that achieve compression in the product stream are different from reduced rank algorithms that achieve matrix compression [] and/or sparsificiation [] of the matrix in a step preceeding multiplication. However, these approaches are not incompatible, with the quadtree data structure supportive of most approaches to matrix compression [] and sparsification [**?**], as well as most fast solvers they might interact with. With little deflation in the cost of fast memory, solver ecosystems that can bring multiple levels of approximation to in place data may enjoy significant cumulitive advantages.

Finally, the mathematical developments in Higham, Mackey, Mackey and T (HMMT; Ref. []) demonstrate the convergence of NS iteration under all groups, addressing potential failure due to the development of pathalogical symmetries related to *e.g.* Eq. 7. Also in this key paper, HMMT develop the fixed point stability analyses (about 2/3 of the way in), which this work draws heavily upon. This work also draws on the scaling in Chen and Chow's [] approach to a scaled NS iteration for ill-conditioned problems.

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. Our interest is to access this basin with the most permissive τ possible, building a foundation for future refinement at a reduced cost and with a higher precision ($\tau\to 0$) [**?**].

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 [**?**], while Byers, He and Mehrmann [] developed asymptotic bounds. The automatic stability of sign iteration is a well developed theme in Ref.[**?**].

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 the cooresponding canonical (dual) square root iteration [];

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

with eigenvalues in the proper domain aggregated towards 0 or 1 by the NS map $h_\alpha[\boldsymbol{x}]=\frac{\sqrt{\alpha}}{2}\left(3-\alpha\boldsymbol{x}\right)$ []. Then, starting with $\boldsymbol{z}_0=\boldsymbol{I}$ and $\boldsymbol{x}_0=\boldsymbol{y}_0=\boldsymbol{s}$, $\boldsymbol{y}_k\to\boldsymbol{s}^{1/2}$, $\boldsymbol{z}_k\to\boldsymbol{s}^{-1/2}$ and $\boldsymbol{x}_k\to\boldsymbol{I}$. As in the case of sign iteration, this dual iteration was shown by Higham, Mackey, Mackey and Tisseur [**?**] to remain 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 this contribution, we consider another aspect of convergence, namely the (hopefully) linear approach towards stability of the iteration

$$\widetilde{\boldsymbol{x}}_k\leftarrow\widetilde{\boldsymbol{y}}_k\left(\widetilde{\boldsymbol{x}}_{k-1}\right)\otimes_\tau\widetilde{\boldsymbol{z}}_k\left(\widetilde{\boldsymbol{x}}_{k-1}\right)\ , \tag{12}$$

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

C. the NS map

Initially, h'_α at the smallest eigenvalue x_0 controls the rate of progress towards idempotence. As recently shown by Jie and Chen [**?**], 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 [**?**], Jie and Chen [**?**], 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 **??**.

D. Ill-conditioning, Stability and Implementation

There are a number of nominally equivalent instances of the square root iteration, related by commutations and transpositions. However, these instances may have very different stability properties, controled to first order by the Frechet derivatives

$$\boldsymbol{x}_{\delta\widehat{\boldsymbol{y}}_{k-1}} = \lim_{\tau\to 0} \frac{\boldsymbol{x}\left(\boldsymbol{y}_{k-1} + \tau\delta\widehat{\boldsymbol{y}}_{k-1}, \boldsymbol{z}_{k-1}\right) - \boldsymbol{x}_k}{\tau} \qquad (13)$$

and

$$\boldsymbol{x}_{\delta\widehat{\boldsymbol{z}}_{k-1}} = \lim_{\tau\to 0} \frac{\boldsymbol{x}\left(\boldsymbol{y}_{k-1}, \boldsymbol{z}_{k-1} + \tau\delta\widehat{\boldsymbol{z}}_{k-1}\right) - \boldsymbol{x}_k}{\tau}\,, \qquad (14)$$

along the unit directions of the previous errors $\delta\widehat{\boldsymbol{y}}_{k-1}$ and $\delta\widehat{\boldsymbol{z}}_{k-1}$, corresponding to the associated displacement magnitudes $\delta y_{k-1} = \|\delta\boldsymbol{y}_{k-1}\|$ and $\delta z_{k-1} = \|\delta\boldsymbol{z}_{k-1}\|$. Then, the differential

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

determines the total first order stability.

This formulation allows to consider orientational effects involving eigenvector fidelity and convergence of derivatives towards zero seperately from displacement effects involving accumulation and `SpAMM` source errors. In some cases, instabilities may be associated with derivatives that do not vanish towards identity, yeilding an unbounded iteration []. In other instances, an instability may be associated with rapidly increasing displacements, due to a too large τ. Instability may also arize due to the numerical corruption of the eigenvectors, also resulting in derivatives that vanish too slowly (or blow up altogether).

The potential for instability is increased with ill-conditioning through the terms $\|\boldsymbol{z}_k\| \to \sqrt{\kappa\left(\boldsymbol{s}\right)}$. Also for ill-conditioned systems, scaling is nessesary to accelerate convergence. However with scaling, increasing the map derivative h'_α can also further enhance the rate of error accumulation.

In following sections, we'll examine how these effects differ from the ideal (double precision) cannonical (dual) square root iteration for ill-conditioned systems and in the strongly non-associative, sketchy $\otimes_\tau$ regime corresponding to permisive values of τ. At this early stage, we are interested in hazzards and opportunities associated with different formulations and implementational details. In addition to deviations from the full precision dual instance, we will develop the "stabilized" instance,

$$\begin{aligned} \boldsymbol{z}_k \;&\leftarrow\; \boldsymbol{z}_{k-1} \cdot h_\alpha\left[\boldsymbol{x}_{k-1}\right]\,, \\ \boldsymbol{x}_k \;&\leftarrow\; \boldsymbol{z}_k^\dagger \cdot \boldsymbol{s} \cdot \boldsymbol{z}_{k-1}\,, \end{aligned} \qquad (16)$$

with the corresponding differential;

$$\delta\boldsymbol{x}_k = \boldsymbol{x}_{\delta\widehat{\boldsymbol{z}}_{k-1}} \times \delta z_{k-1} + \mathcal{O}(\tau^2)\,. \qquad (17)$$

Nominally, $\boldsymbol{y}^{\mathrm{dual}}$ is equivalent to $\boldsymbol{y}_k^{\mathrm{stab}} \equiv \boldsymbol{z}_k^\dagger \cdot \boldsymbol{s}$ is also equivalent to $\boldsymbol{y}_k^{\mathrm{naive}} \equiv \boldsymbol{z}_k \cdot \boldsymbol{s}$. However, with ill-conditioning and in only double precision, these two instances may diverge due to non-associative errors that rapidly compound. In the case of the duals iteration under `SpAMM` approximation, the $\widetilde{\boldsymbol{y}}_k^{\mathrm{dual}}$ channel does not retain contact with the eigenvectors, span $\boldsymbol{s}$, whilst the stab instance does. In the duals iteration, the $\widetilde{\boldsymbol{y}}_k$ `SpAMM` update is mild, with errors in the relative product remaining well conditioned. In the stab instance, conection with $\boldsymbol{s}$ is retained at each step, but at the price of the $\boldsymbol{y}_k^{\mathrm{stab}}$ update involving magnitudes that vary widely in the `SpAMM` product.

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

$$\widetilde{\boldsymbol{x}}_k^R \leftarrow \widetilde{\boldsymbol{z}}_k^\dagger \otimes_\tau \left(\boldsymbol{s} \otimes_{\tau_s} \widetilde{\boldsymbol{z}}_{k-1}\right)\,, \qquad (18)$$

is different from the left first product

$$\widetilde{\boldsymbol{x}}_k^L \leftarrow \left(\widetilde{\boldsymbol{z}}_k^\dagger \otimes_{\tau_s} \boldsymbol{s}\right) \otimes_\tau \widetilde{\boldsymbol{z}}_{k-1}\,. \qquad (19)$$

IV. IMPLEMENTATION

A. programming

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 [].

B. scaling and stabilization

C. regularization

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

D. convergence

Map switching and etc based on TrX

V. DATA

1. double exponential ill-conditioning

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

2. three-dimensional, periodic

3. Matrix Market

VI. STABILITY (PROOF)

VII. ERROR FLOWS IN SQUARE ROOT ITERATION

A. The cannonical (dual) instance

Refering back to Eq. (15), we develop the Fréchet analyses [] with the goal of understanding the contractive approach to identity in competition with error accumulations and SpAMM sources. Of interest are the derivatives

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

$$\begin{aligned}\boldsymbol{x}_{\delta\widehat{\boldsymbol{z}}_{k-1}} = \boldsymbol{y}_{k-1}\cdot h'_\alpha\delta\widehat{\boldsymbol{z}}_{k-1}\cdot\boldsymbol{y}_{k-1}\cdot\boldsymbol{z}_k \\ + \boldsymbol{y}_k\cdot\delta\widehat{\boldsymbol{z}}_{k-1}\cdot h_\alpha\left[\boldsymbol{x}_{k-1}\right] \\ + \boldsymbol{y}_k\cdot\boldsymbol{z}_{k-1}\cdot\boldsymbol{y}_{k-1}\cdot h'_\alpha\delta\widehat{\boldsymbol{z}}_{k-1}\,. \end{aligned}\tag{21}$$

Closer to a 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,

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

and

$$\boldsymbol{x}_{\delta\widehat{\boldsymbol{z}}_{k-1}}\to\left(\boldsymbol{y}_k-\boldsymbol{y}_{k-1}\right)\cdot\delta\widehat{\boldsymbol{z}}_{k-1}.\tag{23}$$

Thus, contributions along $\delta\widehat{\boldsymbol{y}}_{k-1}$ and $\delta\widehat{\boldsymbol{z}}_{k-1}$ are tightly shut down in the region of superlinear convergence. Achieving a contractive fixed point orbit, however requires that the three terms in Eq. (**??**), with potentially different error accumulations and SpAMM sources, must cancel faster than δy_{k-1} and δz_{k-1} accumulate.

In this analysis, we've seperated the directional component of the error from its distance, because in addition to the previous compounding error, each displacement contains also a first order SpAMM source error. Its simpler to consider these effects serpately, at least in this first contribution.

To understand $\delta\boldsymbol{z}_{k-1}$, we partially unwind the approxinate $\widetilde{\boldsymbol{z}}_{k-1}$;

$$\begin{aligned}\widetilde{\boldsymbol{z}}_{k-1} &= \widetilde{\boldsymbol{z}}_{k-2}\otimes_\tau h_\alpha[\widetilde{\boldsymbol{x}}_{k-2}] & (24)\\ &= \Delta_\tau^{\widetilde{\boldsymbol{z}}_{k-2}\cdot h_\alpha[\widetilde{\boldsymbol{x}}_{k-2}]}+\widetilde{\boldsymbol{z}}_{k-2}\cdot h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right] & (25)\end{aligned}$$

Then, using

$$h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right]=h_\alpha\left[\boldsymbol{x}_{k-2}\right]+h'_\alpha\delta\boldsymbol{x}_{k-2}\tag{26}$$

and taking $\boldsymbol{z}_{k-1}$ from both sides, we find

$$\begin{aligned}\delta\boldsymbol{z}_{k-1} = \Delta_\tau^{\widetilde{\boldsymbol{z}}_{k-2}\cdot h_\alpha[\widetilde{\boldsymbol{x}}_{k-2}]} \\ + \delta\boldsymbol{z}_{k-2}\cdot h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right]+\boldsymbol{z}_{k-2}\cdot h'_\alpha\delta\boldsymbol{x}_{k-2}\,,\end{aligned}\tag{27}$$

bounded by

$$\begin{aligned}\delta z_{k-1} < \|\boldsymbol{z}_{k-2}\|\left(\tau\left\|h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right]\right\|+h'_\alpha\delta y_{k-2}\|z_{k-2}\|\right) \\ + \delta z_{k-2}\left(\left\|h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right]\right\|+\|y_{k-2}\|\right).\end{aligned}\tag{28}$$

primary error channels contibuting to δz_{k-1} are through the first order SpAMM error $\tau\|\boldsymbol{z}_{k-2}\|\left\|h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right]\right\|$ and the volatile term $h'_\alpha\delta y_{k-2}\|\boldsymbol{z}_{k-2}\|^2$.

corresponding to basis corruption and controlled by $\otimes_{\tau_s}$, with $\tau_s\ll\tau$. As above, we can unwind this sensitive term, to find

$$\begin{aligned}\delta y_{k-2} < \|\boldsymbol{y}_{k-3}\|\left(\tau_s\|h_\alpha[\widetilde{\boldsymbol{x}}_{k-3}]\|+h'_\alpha\delta z_{k-3}\right) \\ + \delta y_{k-3}\left(\|\widetilde{\boldsymbol{z}}_{k-3}]\|+\|h_\alpha[\widetilde{\boldsymbol{x}}_{k-3}]\|\right).\end{aligned}\tag{29}$$

B. The stabilized (stab) instance

Here, we carry on from Eq. (17) in the "stabilized" instance, with the single channel differential

$$\boldsymbol{x}_{\widehat{\boldsymbol{z}}_{k-1}}=\boldsymbol{z}^\dagger_{\widehat{\boldsymbol{z}}_{k-1}}\cdot\boldsymbol{s}\cdot\boldsymbol{z}_k+\boldsymbol{z}^\dagger_k\cdot\boldsymbol{s}\cdot\boldsymbol{z}_{\widehat{\boldsymbol{z}}_{k-1}}\tag{30}$$

$$\begin{aligned}\boldsymbol{z}_{\widehat{\boldsymbol{z}}_{k-1}} = \delta\widehat{\boldsymbol{z}}_{k-1}\cdot h_\alpha[\widetilde{\boldsymbol{x}}_{k-1}]+\boldsymbol{z}_{k-1}\cdot(\\ h'_\alpha\delta\widehat{\boldsymbol{z}}^\dagger_{k-1}\cdot\boldsymbol{s}\cdot\boldsymbol{z}_{k-1}+\boldsymbol{z}^\dagger_{k-1}\cdot\boldsymbol{s}\cdot h'_\alpha\delta\widehat{\boldsymbol{z}}_{k-1}\Big)\end{aligned}\tag{31}$$

$$\begin{aligned}\widetilde{\boldsymbol{y}}^{\rm stab}_{k-1} &= \widetilde{\boldsymbol{z}}^\dagger_{k-1}\otimes_\tau\boldsymbol{s} & (32)\\ &= \boldsymbol{\Delta}^{\widetilde{\boldsymbol{z}}^\dagger_{k-1}\cdot\boldsymbol{s}}+\left(\widetilde{\boldsymbol{z}}_{k-2}\cdot h_\alpha[\widetilde{\boldsymbol{x}}_{k-2}]\right)^\dagger\cdot\boldsymbol{s} & (33)\end{aligned}$$

C. Bifurcations

Differences in occlusion between stab and dual magnified as bounds for s.z not as tight as bounds for h.y.

lot of overlap too (reproducing hilberts etc).

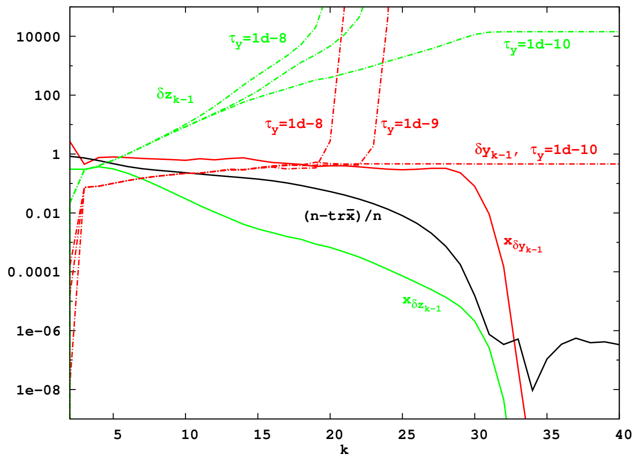

FIG. 2: Derivatives, displacements and the approximate trace of the unscaled, dual NS iteration for a (3,3) nanotube with $\kappa = 10^{10}$. Derivatives are full lines, whilst the displacements cooresponding to $b = 64$, $\tau = 10^{-3}$ and $\tau_y = \{10^{-8}, 10^{-9}, 10^{-10}\}$ are the dashed lines. The trace expectation is shown as a full black line.

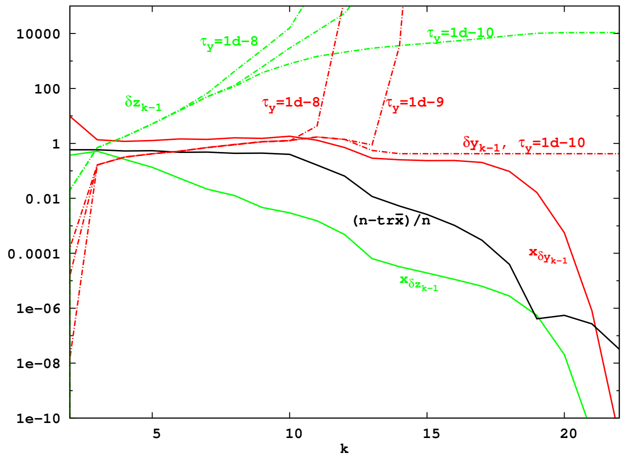

FIG. 3: Derivatives, displacements and the approximate trace of the scaled, stablized NS iteration for a (3,3) nanotube with $\kappa = 10^{10}$. Derivatives are full lines, whilst the displacements cooresponding to $b = 64$, $\tau = 10^{-3}$ and $\tau_y = \{10^{-3}, 10^{-4}, 10^{-6}\}$ are the dashed lines. The trace expectation is shown as a full black line.

VIII. ITERATED REGULARIZATION

Limits of by the constraints on stability, imposed by *e.g.* ill-conditioning. These limits can be circumscribed through Tikhonov regularization [], involving a level shift of the original matrix, $\boldsymbol{s}_\mu \leftarrow \boldsymbol{s}+\mu\boldsymbol{I}$, leading to an effective condition number for $\boldsymbol{s}_\mu$ by the ammount ... []. However, $\boldsymbol{s}_\mu^{-1/2}$ may still be too far from $\boldsymbol{s}^{-1/2}$ to be of much use.

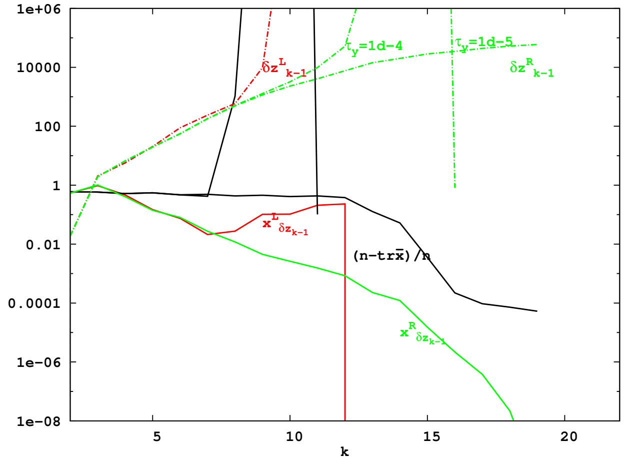

FIG. 4: Derivatives, displacements and the approximate trace of the unscaled, dual NS iteration for a (3,3) nanotube with $\kappa = 10^{10}$. Derivatives are full lines, whilst the displacements cooresponding to $b = 64$, $\tau = 10^{-3}$ and $\tau_y = \{10^{-8}, 10^{-9}, 10^{-10}\}$ are the dashed lines. The trace expectation is shown as a full black line.

One approach is to use Riley's method []

$$\boldsymbol{s}^{-1/2} = \boldsymbol{s}_\mu^{-1/2} \cdot \left(\boldsymbol{I} + \frac{\mu}{2}\boldsymbol{s}_\mu^{-1} + \frac{3\mu^2}{8}\boldsymbol{s}_\mu^{-2} + \dots \right) . \quad (34)$$

Alternatively, we consider Tikhonov regularization and its implementatation with the `SpAMM` algebra. In this approach, a most permisive (but still effective) preconditioner is found; $\boldsymbol{s}_{\tau_0,\mu_0}^{-1/2}$. By permisive, we mean τ_0 is large but enables stable NS iterations. By effective we mean $\mu_0 < .1$ (*ie* taking $\boldsymbol{s} \leftarrow \boldsymbol{s}/s_0$, conditioning is improved by at least one order).

In this exploritory contribution, we examine the NS square root iteration with a generic thin slice, cooresponding to $\tau_0 \sim .1$ and $\mu_0 \sim .1$, and with unregularized (thick) `SpAMM` iteration at the edge of stability. First up, we show volumes of the stable instance from Fig (**??**) as a percentage of the total (conventional) N^3 volume. Next up in Fig. **??** is the same problem in the dual instance.

These stable and dual instances behave very differently in this case. In the first instance, the $\boldsymbol{y}$ channel is unimproved and retains sensitivity of the unregularized problem. This is because $\widetilde{\boldsymbol{y}}_k^{\text{stab}} \rightarrow \boldsymbol{s}^{1/2_{\tau_0\mu_0}} \equiv \boldsymbol{s}_{\tau_0\mu_0}^{-1/2} \otimes_{\tau_0} \boldsymbol{s}_{\mu_0}$ is a product not helped by regularization. In the dual case the situation is quite different with $\widetilde{\boldsymbol{y}}_k^{\text{dual}} \rightarrow \boldsymbol{I}_{\tau_0\mu_0} \otimes_{\tau_0} \boldsymbol{s}_{\tau_0\mu_0}^{1/2}$, which is reflective about the cube diagonal, and with a complexity approaching the cost of quadtree copy in place.

In the dual case then, τ_0 can be brought to extreme permisve values, $\tau_0 \sim .1$, for *e.g.* $\mu_0 \sim .1$. Then, from a first generic slice $\boldsymbol{s}_{\tau_0,\mu_0}^{-1/2}$, dual NS maybe employed again to find a next effective preconditioner, $\boldsymbol{s}_{\tau_0\mu_1}^{-1/2}$, based on the residual $\left(\boldsymbol{s}_{\tau_0\mu_0}^{-1/2}\right)^\dagger \otimes_{\tau_0} (\boldsymbol{s} + \mu_1\boldsymbol{I}) \otimes_{\tau_0} \boldsymbol{s}_{\tau_0\mu_0}^{-1/2}$, again

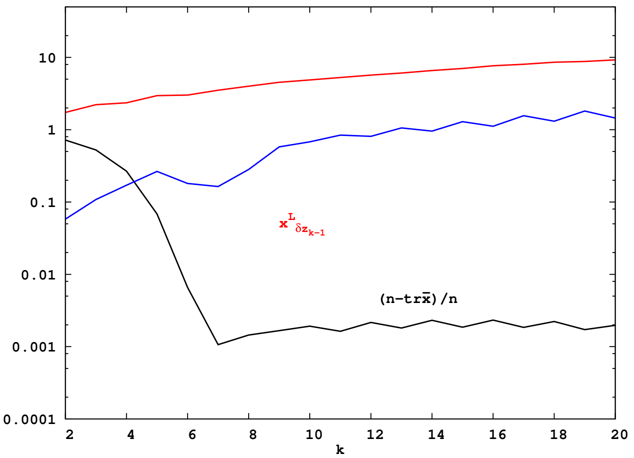

FIG. 5: Culled volumes for the XXx U.C. (3,3) $\kappa(\boldsymbol{s}) = 10^{10}$ nanotube overlap matrix examined in the previous section, in the stable instance and with regularization ($\mu_0 = .1$): $\mathrm{v}_{\widetilde{\boldsymbol{z}_\mathrm{k}}} = \left(\mathrm{vol}_{\widetilde{\boldsymbol{z}}_{\mathrm{k}-1} \otimes_\tau \mathrm{h}[\widetilde{\boldsymbol{x}}_{\mathrm{k}-1}]}\right) \times 100\%/\mathrm{N}^3$, $\mathrm{v}_{\widetilde{\boldsymbol{y}}_k} = \left(\mathrm{vol}_{\boldsymbol{s} \otimes_{\tau_s} \widetilde{\boldsymbol{z}}_k}\right) \times 100\%/N^3$. Also shown is the trace error, $\mathrm{e}_k = \left(N - \mathrm{tr}\, \boldsymbol{x}_k\right)/N$.

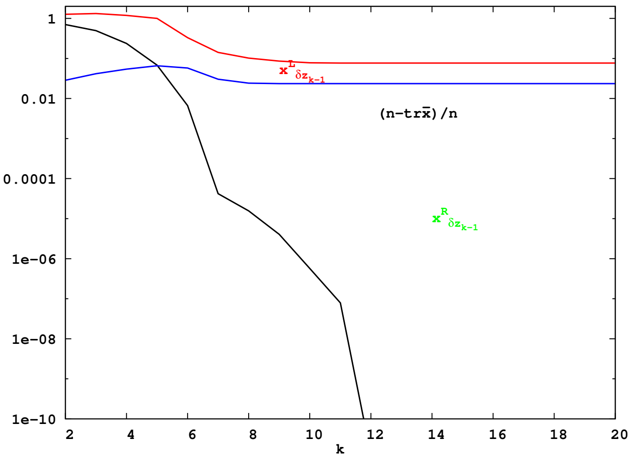

FIG. 6: Culled volumes for the XXx U.C. (3,3) $\kappa(\boldsymbol{s}) = 10^{10}$ nanotube overlap matrix examined in the previous section, in the dual instance and with regularization ($\mu_0 = .1$): $\mathrm{v}_{\widetilde{\boldsymbol{y}}_k} = \left(\mathrm{vol}_{h[\widetilde{\boldsymbol{x}}_{k-1}] \otimes_{\tau_s} \widetilde{\boldsymbol{y}}_k}\right) \times 100\%/N^3$, $\mathrm{v}_{\widetilde{\boldsymbol{z}}_k} = \left(\mathrm{vol}_{\widetilde{\boldsymbol{z}}_{k-1} \otimes_\tau h[\widetilde{\boldsymbol{x}}_{k-1}]}\right) \times 100\%/N^3$, Also shown is the trace error, $\mathrm{e}_k = \left(N - \mathrm{tr}\, \boldsymbol{x}_k\right)/N$.

with an effective increment $\mu_1 < .1\mu_0$. Generally then, it may be possible to find the full factor as the nested product of generic slices (`SpAMM` sandwitch):

$$\boldsymbol{s}_{\tau_0}^{-1/2} = \boldsymbol{s}_{\tau_0 \mu_n}^{-1/2} \otimes_{\tau_0} \boldsymbol{s}_{\tau_0 \mu_{n-1}}^{-1/2} \otimes_{\tau_0} \ldots \boldsymbol{s}_{\tau_0 \mu_0}^{-1/2} \tag{35}$$

$$= \bigotimes_{\mu=\mu_0}^{\mu_n} \boldsymbol{s}_{\tau_0 \mu} \,. \tag{36}$$

In this way, iterated 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, a nested product representation based on increasing `SpAMM` resolution may build the complete factor digit by digit:

$$\boldsymbol{s}^{-1/2} = \boldsymbol{s}_{\tau_m}^{-1/2} \otimes_{\tau_m} \boldsymbol{s}_{\tau_{m-1}}^{-1/2} \otimes_{\tau_{m-1}} \ldots \boldsymbol{s}_{\tau_0}^{-1/2} \tag{37}$$

$$= \bigotimes_{\tau=\tau_0}^{\tau_m} \boldsymbol{s}_{\tau} \,, \tag{38}$$

with $1 > \tau_0 > \tau_1 > \cdots > \tau_n$. Certainly, there are many ways to combine `SpAMM` resolutions with Tikhonov regularization.

The naive scheme requires $m + n$ total NS solves, and potentially as many multiplies to incrementally apply the full factor $\boldsymbol{s}^{-1/2}$. However, a thin (~ 0.1) product representation may have a number of advantages: (**1**) Each NS solve can be reduced to a generic, constant and reduced number of well behaved steps; (**2**) each NS solve can be brought into the regime of strongly contractive identity iteration; (**3**) Equation (6) tightly bounds products about the cube-diagonal at each resolution $\boldsymbol{I}_{\tau,\mu}\left(\boldsymbol{s}\right)$, and normalizing gaps between τ and τ_s; (**4**) Resolution of the identity cooresponds to a sharply culled ijk-volume, $\mathrm{vol}_{\otimes_\tau}$, that is compressive and tending towards copy in place complexities for $\boldsymbol{y}$ and $\boldsymbol{z}$ updates; (**5**) Each generic solve may achieve additional computational and mathematical acceleration for the next solve through persistence (Section **??**).

A. Locality

A premier feature of the n-body algorithm is the ability to exploit data locality efficiently. In the ijk space, there are multiple locality principles that can be exploited by `SpAMM`. Fir persistence copy in place, tightness of the bound

curse of dimensionality Each generic slice may be associated with different *temporal locality*

Metric locality is locality with respect to a Euclidean or generalized distance, *e.g.* of the basis.

increasing the Euclidean locality, "wrings the zeros" out of the product space for systems with decay. On moving from locality provided by the Hilbert order (space filling curve) to annealed sollutions to the end-to-end TSP.

1. *Lensing*

A feature of square root iteration with the $\otimes_\tau$ kernel is localization of the culled octree towards identity iteration, $\widetilde{\boldsymbol{x}}_k \rightarrow \boldsymbol{I}\left(\widetilde{\boldsymbol{x}}_{k-1}\right)$. Towards convergence, the product $\widetilde{\boldsymbol{y}}_{k \otimes_\tau} \widetilde{\boldsymbol{z}}_k$ involves the product of large and small eigenvalues, and large and small norms, which are recursively brought towards unity along the $i = k$ diagonal. Likewise, application of the NS map, Eq. (9), tend towards reflection about the ijk cube-diagonal. Because the `SpAMM` error obeys the multiplicative Cauchy-Schwarz bound, Eq. (), the cooresponding culled-octree can likewise follow the $i = j$ plane about the ijk cube-diagonal, resolving the *relative* error in identity to within τ. This effect is shown in Figure **??**. We call this identity related, plane-wise concentration of the culled octree about the cube-diagonal *lensing*.

Lensing is an algebraic localization offering compression beyond ,

complexity reduction relative to the naive (full) volume of the cube, and also relative to sparsification strategies that preserve only absolute errors, as in Eq. 2. The lensed task space offers an enhanced locality of reference, and may also afford fast methods with costs approaching an in-place scalar multiply and copy, *e.g.* as $h_\alpha \to \boldsymbol{I}$ in Eq. 9. Our thesis is that many problems in physical and information sciences can be brought to this lensed state, *e.g.* through preconditioning as described here, and maintained as the NS residual is brought to a higher level of precision with a more complete $\otimes_\tau$, and also with respect to an outer simulation loop, *e.g.* cooresponding to time iteration.

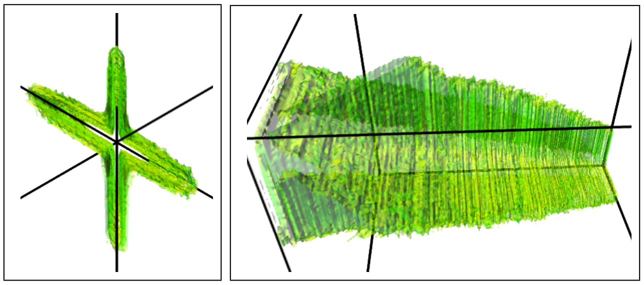

FIG. 7: Views of the $\tau = 0.03$ sign occlusion surface, for the 128x u.c. nanotube, at $\sim 14k \times 14k$ and $\kappa(\boldsymbol{s}) = 10^6$. This surface envelopes the ijk volume of the $\otimes_\tau$ kernel, cooresponding to the unscaled dual iteration step $\widetilde{\boldsymbol{x}}_{19} \leftarrow \widetilde{\boldsymbol{y}}_{19} \otimes_\tau \widetilde{\boldsymbol{z}}_{19}$ at $b = 64$, $\tau = 0.03$ and $\tau_y = 10^{-3}\tau$. The first pannel looks straight down the cube-diagonal $i = j = k$, from the upper bound towards (1,1,1). Remarkably, this surface forms an elongated ×, closely following intersection of the $i = j$ and $i = k$ planes along the cube-diagonal. The second pannel looks along the cube-diagonal, with the upper bound at upper left, and (1,1,1) at lower right.

In this section, we present numerical experiments that highlight the effects of ill-conditioning, dimensionality, and the stability of different first order NS approaches to iteration with `SpAMM`. We turn first to complexity reduction for $\otimes_\tau$ in the basin of stability, where we find a novel, compressive effect in the product octree. This effect is shown in Fig. VIII A 1, for unscaled, inverse square root duals iteration, Eqs. (**??**), on the 3,3 carbon nanotube metric at $\kappa = 10^6$.

In this example, the `SpAMM` octree culled from the ijk-cube is outlined by its occlusion surface, enclosing a volume that closely follows the $i = j$ and $i = k$ planes, forming an ×. The banded distribution of large norms along matrix diagonals leads to cube-diagonal dominance, with plane-following a consequence of moderate ill-conditioning, large norms along the plane-diagonals and their overlap in ijk via the multiplicative bound, Eq. (6). The tightness of this bound, and the compression gained relative to methods that control only the absolute error, *e.g.* as given by Eq. (2), will hopefully be quantified in future work.

2. Dimensionality

3. Strong Metric Locality

unscaled, with hilbert order
unscaled, with salesman's order

4. Misc

B. Appendix

In this section, we prove the following stability bound for SpAMM.

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. To prove 1, we first need the following technical result: it is possible to choose $\alpha_{lij} \in \{0, 1\}$ such that

$$(A \otimes_\tau B)_{ij} = \sum_{l=1}^{n} A_{il} B_{lj} \alpha_{lij}, \tag{39}$$

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 (39) 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 (39) 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

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

where we used the induction hypothesis in the second equality.

Now suppose that $\alpha_{lij} = 0$ for some l. Then $\tilde{\alpha}_{lij}^{(1)} = 0$ if $l \leq n/2$ or $\tilde{\alpha}_{l-n/2,ij}^{(2)} = 0$ $l > n/2$. If, e.g., $\tilde{\alpha}_{l-n/2,ij}^{(2)} = 0$, then $\left|A_{il}\right| \left|B_{lj}\right| = \left|\left(A_{12}\right)_{i,l-n/2}\right| \left|\left(B_{21}\right)_{l-n/2,j}\right| < \tau_{A,B}$, 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 (39),

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

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

□

IX. CONCLUSIONS AND OUTLOOK

Reflecting about the cube diagonal is copy in place. Cooresponds to lensing.

These synnergistic effects can be combined, in place, with additional tecniques for accelerated computaion.