

Stability and Algebraic Locality In Square Root Iteration

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 relationship between matrix elements and the Cartesian distance between corresponding support functions, or it may involve a generalized distance, *e.g.* a generalized measure between character 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 [1–5], 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 [6, 7] or a related factorization [5]; these transformations correlate local support with a representation independent form, *e.g.* of the eigenproblem. Interestingly, the matrix sign function and the matrix inverse square root function are related by Higham’s identity:

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

A complete overview of matrix function theory and computation is given in Higham’s enjoyable reference [8].

A well conditioned matrix \mathbf{s} may often correspond to matrix sign and inverse square root functions with rapid exponential decay, and be amenable to the sparse matrix approximation $\bar{\mathbf{s}} = \mathbf{s} + \epsilon_{\tau}^{\mathbf{s}}$, where $\epsilon_{\tau}^{\mathbf{s}}$ 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 drop-tolerance, $\epsilon_{\tau}^{\mathbf{s}} = \{-s_{ij} * \hat{e}_i \mid |s_{ij}| < \tau\}$, a radial cutoff, $\epsilon_{\tau}^{\mathbf{s}} = \{-s_{ij} * \hat{e}_i \mid \|\mathbf{r}_i - \mathbf{r}_j\| > \tau\}$, 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) [9–12], yielding fast solutions for multiplication rich iterations and a modulated (**what do you mean with modulated?**) fill-in. These and

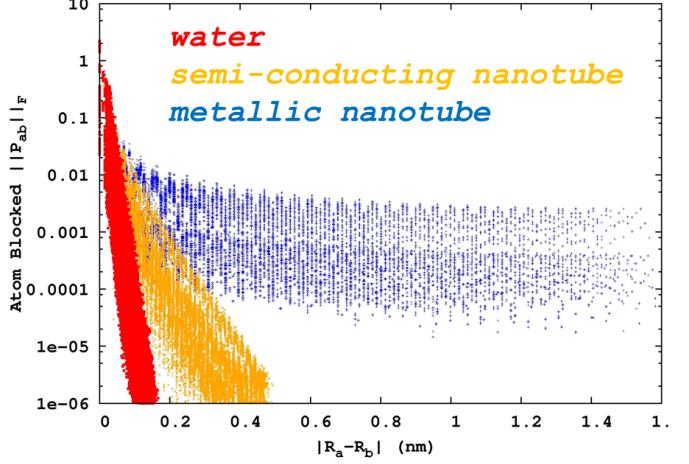


FIG. 1: Examples from electronic structure of decay for the spectral projector (gap shifted sign function) with respect to local (atomic) support. Shown is decay for systems with correlations that are short (insulating water), medium (semi-conducting 4,3 nanotube), and long (metallic 3,3 nanotube) ranged,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

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 [13, 14], and by Bowler [15] and Benzi [16] for electronic structure.

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

$$\overline{\mathbf{a} \cdot \mathbf{b}} = \mathbf{a} \cdot \mathbf{b} + \epsilon_{\tau}^{\mathbf{a}} \cdot \mathbf{b} + \mathbf{a} \cdot \epsilon_{\tau}^{\mathbf{b}} + \mathcal{O}(\tau^2) \quad (2)$$

achieving sparse, stable and rapidly convergent iteration for ill-conditioned problems can be challenging []. In cases of extreme degeneracy, hierarchical semi-separable (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: haut@lanl.gov

[‡]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 [? ?].

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

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. A Bounded Occlusion and Cull

SpAMM has evolved from a row-column oriented skipout [?] to recursive occlusion and culling [?] based on sub-multiplicative norms $\|\cdot\| \equiv \|\cdot\|_F$ and the Cauchy-Schwarz inequality, $\|\mathbf{a} \cdot \mathbf{b}\| < \|\mathbf{a}\| \|\mathbf{b}\|$ [?]. Occlusion involves avoiding work, whilst culling is the collecting of tasks. Here, we amend the previous (naïve) occlusion and cull with the following recursion:

$$\mathbf{a}^i \otimes_{\tau} \mathbf{b}^i = \begin{cases} \emptyset & \text{if } \|\mathbf{a}^i\| \|\mathbf{b}^i\| < \tau \|\mathbf{a}\| \|\mathbf{b}\| \\ \mathbf{a}^i \cdot \mathbf{b}^i & \text{if } (\mathbf{i} = \text{leaf}) \\ \left[\mathbf{a}_{00}^{i+1} \otimes_{\tau} \mathbf{b}_{00}^{i+1} + \mathbf{a}_{01}^{i+1} \otimes_{\tau} \mathbf{b}_{10}^{i+1}, \quad \mathbf{a}_{00}^{i+1} \otimes_{\tau} \mathbf{b}_{01}^{i+1} + \mathbf{a}_{01}^{i+1} \otimes_{\tau} \mathbf{b}_{11}^{i+1} \right] & \text{else} \\ \left[\mathbf{a}_{00}^{i+1} \otimes_{\tau} \mathbf{b}_{01}^{i+1} + \mathbf{a}_{01}^{i+1} \otimes_{\tau} \mathbf{b}_{11}^{i+1}, \quad \mathbf{a}_{00}^{i+1} \otimes_{\tau} \mathbf{b}_{01}^{i+1} + \mathbf{a}_{01}^{i+1} \otimes_{\tau} \mathbf{b}_{11}^{i+1} \right] & \text{else} \end{cases}, \quad (4)$$

which bounds the relative occlusion error,

$$\frac{\|\Delta_{\tau}^{a \cdot b}\|}{N^2} \leq \tau \|\mathbf{a}\| \|\mathbf{b}\|, \quad (5)$$

occurring in the approximate product:

$$\widetilde{\mathbf{a} \cdot \mathbf{b}} \equiv \mathbf{a} \otimes_{\tau} \mathbf{b} = \mathbf{a} \cdot \mathbf{b} + \Delta_{\tau}^{a \cdot b}. \quad (6)$$

B. Proof

We now prove (5).

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

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

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

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

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

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

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

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

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 $\tilde{\alpha}_{l-n/2,ij}^{(2)} = 0$ if $l > n/2$. If, e.g., $\tilde{\alpha}_{l-n/2,ij}^{(2)} = 0$, then $|A_{il}| |B_{lj}| = |(A_{12})_{i,l-n/2}| |(B_{21})_{l-n/2,j}| < \tau_{A,B}$, 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),

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

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

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 [1]. In the SpAMM approach to fast multixplication, the numerically most significant volumes in naïve (ijk) tensor intermediates are culled, with error bounded by Eq. (5). This bound makes \otimes_{τ} 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* [1] of recursive data access [1], enabling range queries, metric queries, higher dimensional queries and so on, with common frameworks, structures and runtimes. Recently for example, we generalized the double (left-right) metric query in Eq. (??) to hextree metric queries of the exact Fock exchange [1].

Also, n -body methods offer well established protocols for turning spatial and metric locality into data and temporal locality [1]; recently, x, y and Yellik 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) [1]. The introduction of algebraic and strong Euclidean locality in this work may further enhance high performance implementaitons.

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 [1] and probing [1]. These methods involve a weighted (probabilistic) and on the fly sampling, with the potential for complexity reduction in the case of random distributions. SpAMM also employs on the fly weighted sampling, but with compression through locality, brought about by algebraic correlations (towards identity) and also in the metric structure, through strong Euclidean locality.

Methods that achieve compression in the stream of product intermediates are different from reduced rank algorithms that achieve matrix compression in a step that precedes multiplication (separability) [1]. However, matrix compressions are generally compatible with the quadtree, as are additional fast (generalized) solvers that add complex functionality (e.g. in electronic structure theory [1]). Thus, generality and interoperability may enable deeply layered, thin and generic solvers easy access to in place data. Further, language support may provide simple (skeletonized) frameworks for generic recursion, offering opportunities to greatly simplify codebase at the systems level, lowering barriers to entry and enhancing concurrence (in the Erlang sense).

Finally, previous work on the scaled NS iteration has heavily influenced this work. Formost is Higham, Mackey, Mackey and T (HMMT; Ref. [1]) masterwork on convergence of NS iteration under all groups, wherein HMMT also develop Fréchet analyses for stable square root iteration at the fixed point. Also, important inspiration comes from Chen and Chow's [1] approach to scaled NS iteration for ill-conditioned problems [1], and from the Helgaker groups work on NS iteration, whose notation we follow in part [1].

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, $\mathbf{I}(\cdot) = \text{sign}^2(\cdot)$. 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 ($\tau \rightarrow 0$) [?].

A. Sign iteration

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

$$\mathbf{I}(\mathbf{s}) = \mathbf{s}^{1/2} \cdot \mathbf{s}^{-1/2}, \quad (8)$$

and the cooresponding canonical (dual) square root iteration [1];

$$\begin{aligned} \mathbf{y}_k &\leftarrow h_\alpha [\mathbf{y}_{k-1} \cdot \mathbf{z}_{k-1}] \cdot \mathbf{y}_{k-1} \\ \mathbf{z}_k &\leftarrow \mathbf{z}_{k-1} \cdot h_\alpha [\mathbf{y}_{k-1} \cdot \mathbf{z}_{k-1}], \end{aligned} \quad (9)$$

with eigenvalues in the proper domain aggregated towards 0 or 1 by the NS map $h_\alpha[\mathbf{x}] = \frac{\sqrt{\alpha}}{2}(3 - \alpha\mathbf{x})$ [1]. Then, starting with $\mathbf{z}_0 = \mathbf{I}$ and

$$\mathbf{s} \leftarrow \mathbf{s}/s_{N-1},$$

$\mathbf{x}_0 = \mathbf{y}_0 = \mathbf{s}$, $\mathbf{y}_k \rightarrow \mathbf{s}^{1/2}$, $\mathbf{z}_k \rightarrow \mathbf{s}^{-1/2}$ and $\mathbf{x}_k \rightarrow \mathbf{I}$. As in the case of sign iteration, this dual iteration was shown by Higham, Mackey, Mackey and Tisseur [35] to remain bounded in the superlinear regime, by idempotent Frechet derivatives about the fixed point $(\mathbf{s}^{1/2}, \mathbf{s}^{-1/2})$, in the direction $(\delta\mathbf{y}_{k-1}, \delta\mathbf{z}_{k-1})$:

$$\delta\mathbf{y}_k = \frac{1}{2}\delta\mathbf{y}_{k-1} - \frac{1}{2}\mathbf{s}^{1/2} \cdot \delta\mathbf{z}_{k-1} \cdot \mathbf{s}^{1/2} \quad (10)$$

$$\delta\mathbf{z}_k = \frac{1}{2}\delta\mathbf{z}_{k-1} - \frac{1}{2}\mathbf{s}^{-1/2} \cdot \delta\mathbf{y}_{k-1} \cdot \mathbf{s}^{-1/2}. \quad (11)$$

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

$$\tilde{\mathbf{x}}_k \leftarrow \tilde{\mathbf{y}}_k (\tilde{\mathbf{x}}_{k-1}) \otimes_\tau \tilde{\mathbf{z}}_k (\tilde{\mathbf{x}}_{k-1}), \quad (12)$$

made difficult by ill-conditioning and a sketchy \otimes_τ .

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 [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 ??.

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

$$\mathbf{x}_{\delta\hat{\mathbf{y}}_{k-1}} = \lim_{\tau \rightarrow 0} \frac{\mathbf{x}(\mathbf{y}_{k-1} + \tau\delta\hat{\mathbf{y}}_{k-1}, \mathbf{z}_{k-1}) - \mathbf{x}_k}{\tau} \quad (13)$$

and

$$\mathbf{x}_{\delta\hat{\mathbf{z}}_{k-1}} = \lim_{\tau \rightarrow 0} \frac{\mathbf{x}(\mathbf{y}_{k-1}, \mathbf{z}_{k-1} + \tau\delta\hat{\mathbf{z}}_{k-1}) - \mathbf{x}_k}{\tau}, \quad (14)$$

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

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

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 [1]. 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 $\|\mathbf{z}_k\| \rightarrow \sqrt{\kappa(\mathbf{s})}$. Also for ill-conditioned systems, scaling is nessessary 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) canonical (dual) square root iteration for ill-conditioned systems and in the strongly non-associative, sketchy \otimes_τ 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,

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

with the corresponding differential;

$$\delta\mathbf{x}_k = \mathbf{x}_{\delta\tilde{\mathbf{z}}_{k-1}} \times \delta\mathbf{z}_{k-1} + \mathcal{O}(\tau^2) . \quad (17)$$

Nominally, \mathbf{y}^{dual} is equivalent to $\mathbf{y}_k^{\text{stab}} \equiv \mathbf{z}_k^\dagger \cdot \mathbf{s}$ is also equivalent to $\mathbf{y}_k^{\text{naive}} \equiv \mathbf{z}_k \cdot \mathbf{s}$. 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 $\tilde{\mathbf{y}}_k^{\text{dual}}$ channel does not retain contact with the eigenvectors, span \mathbf{s} , whilst the stab instance does. In the duals iteration, the $\tilde{\mathbf{y}}_k$ SpAMM update is mild, with errors in the relative product remaining well conditioned. In the stab instance, connection with \mathbf{s} is retained at each step, but at the price of the $\mathbf{y}_k^{\text{stab}}$ update involving magnitudes that vary widely in the SpAMM product.

IV. ERROR FLOWS IN SQUARE ROOT ITERATION

A. The canonical (dual) instance

Referring 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

$$\begin{aligned} \mathbf{x}_{\delta\hat{\mathbf{y}}_{k-1}} &= h_\alpha[\mathbf{x}_{k-1}] \cdot \delta\hat{\mathbf{y}}_{k-1} \cdot \mathbf{z}_k \\ &\quad + h'_\alpha \delta\hat{\mathbf{y}}_{k-1} \cdot \mathbf{z}_{k-1} \cdot \mathbf{y}_{k-1} \cdot \mathbf{z}_k \\ &\quad + \mathbf{y}_k \cdot \mathbf{z}_{k-1} \cdot h'_\alpha \delta\hat{\mathbf{y}}_{k-1} \cdot \mathbf{z}_{k-1} . \end{aligned} \quad (18)$$

$$\begin{aligned} \mathbf{x}_{\delta\hat{\mathbf{z}}_{k-1}} &= \mathbf{y}_{k-1} \cdot h'_\alpha \delta\hat{\mathbf{z}}_{k-1} \cdot \mathbf{y}_{k-1} \cdot \mathbf{z}_k \\ &\quad + \mathbf{y}_k \cdot \delta\hat{\mathbf{z}}_{k-1} \cdot h_\alpha[\mathbf{x}_{k-1}] \\ &\quad + \mathbf{y}_k \cdot \mathbf{z}_{k-1} \cdot \mathbf{y}_{k-1} \cdot h'_\alpha \delta\hat{\mathbf{z}}_{k-1} . \end{aligned} \quad (19)$$

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

$$\mathbf{x}_{\delta\hat{\mathbf{y}}_{k-1}} \rightarrow \delta\hat{\mathbf{y}}_{k-1} \cdot (\mathbf{z}_k - \mathbf{z}_{k-1}) \quad (20)$$

and

$$\mathbf{x}_{\delta\hat{\mathbf{z}}_{k-1}} \rightarrow (\mathbf{y}_k - \mathbf{y}_{k-1}) \cdot \delta\hat{\mathbf{z}}_{k-1} . \quad (21)$$

Thus, contributions along $\delta\hat{\mathbf{y}}_{k-1}$ and $\delta\hat{\mathbf{z}}_{k-1}$ 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 $\delta\mathbf{y}_{k-1}$ and $\delta\mathbf{z}_{k-1}$ accumulate.

In this analysis, we’ve separated 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 separately, at least in this first contribution.

To understand $\delta\mathbf{z}_{k-1}$, we partially unwind the approximate $\tilde{\mathbf{z}}_{k-1}$:

$$\tilde{\mathbf{z}}_{k-1} = \tilde{\mathbf{z}}_{k-2} \otimes_\tau h_\alpha[\tilde{\mathbf{x}}_{k-2}] \quad (22)$$

$$= \Delta_{\tau}^{\tilde{\mathbf{z}}_{k-2} \cdot h_\alpha[\tilde{\mathbf{x}}_{k-2}]} + \tilde{\mathbf{z}}_{k-2} \cdot h_\alpha[\tilde{\mathbf{x}}_{k-2}] \quad (23)$$

Then, using

$$h_\alpha[\tilde{\mathbf{x}}_{k-2}] = h_\alpha[\mathbf{x}_{k-2}] + h'_\alpha \delta\mathbf{x}_{k-2} \quad (24)$$

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

$$\begin{aligned} \delta\mathbf{z}_{k-1} &= \Delta_{\tau}^{\tilde{\mathbf{z}}_{k-2} \cdot h_\alpha[\tilde{\mathbf{x}}_{k-2}]} \\ &\quad + \delta\mathbf{z}_{k-2} \cdot h_\alpha[\tilde{\mathbf{x}}_{k-2}] + \mathbf{z}_{k-2} \cdot h'_\alpha \delta\mathbf{x}_{k-2} , \end{aligned} \quad (25)$$

bounded by

$$\begin{aligned} \delta\mathbf{z}_{k-1} &< \|\mathbf{z}_{k-2}\| (\tau \|h_\alpha[\tilde{\mathbf{x}}_{k-2}]\| + h'_\alpha \delta\mathbf{y}_{k-2} \|\mathbf{z}_{k-2}\|) \\ &\quad + \delta\mathbf{z}_{k-2} (\|h_\alpha[\tilde{\mathbf{x}}_{k-2}]\| + \|\mathbf{y}_{k-2}\|) . \end{aligned} \quad (26)$$

primary error channels contributing to $\delta\mathbf{z}_{k-1}$ are through the first order SpAMM error $\tau \|\mathbf{z}_{k-2}\| \|h_\alpha[\tilde{\mathbf{x}}_{k-2}]\|$ and the volatile term $h'_\alpha \delta\mathbf{y}_{k-2} \|\mathbf{z}_{k-2}\|^2$.

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

$$\begin{aligned} \delta\mathbf{y}_{k-2} &< \|\mathbf{y}_{k-3}\| (\tau_s \|h_\alpha[\tilde{\mathbf{x}}_{k-3}]\| + h'_\alpha \delta\mathbf{z}_{k-3}) \\ &\quad + \delta\mathbf{y}_{k-3} (\|\tilde{\mathbf{z}}_{k-3}\| + \|h_\alpha[\tilde{\mathbf{x}}_{k-3}]\|) . \end{aligned} \quad (27)$$

B. The stabilized (stab) instance

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

$$\mathbf{x}_{\tilde{\mathbf{z}}_{k-1}} = \tilde{\mathbf{z}}_{\tilde{\mathbf{z}}_{k-1}}^\dagger \cdot \mathbf{s} \cdot \mathbf{z}_k + \mathbf{z}_k^\dagger \cdot \mathbf{s} \cdot \mathbf{z}_{\tilde{\mathbf{z}}_{k-1}} \quad (28)$$

$$\begin{aligned} \mathbf{z}_{\tilde{\mathbf{z}}_{k-1}} &= \delta\tilde{\mathbf{z}}_{k-1} \cdot h_\alpha[\tilde{\mathbf{x}}_{k-1}] + \mathbf{z}_{k-1} \cdot (\\ &\quad h'_\alpha \delta\tilde{\mathbf{z}}_{k-1}^\dagger \cdot \mathbf{s} \cdot \mathbf{z}_{k-1} + \mathbf{z}_{k-1}^\dagger \cdot \mathbf{s} \cdot h'_\alpha \delta\tilde{\mathbf{z}}_{k-1}) \end{aligned} \quad (29)$$

$$\tilde{\mathbf{y}}_{k-1}^{\text{stab}} = \tilde{\mathbf{z}}_{k-1}^\dagger \otimes_\tau \mathbf{s} \quad (30)$$

$$= \Delta_{\tau}^{\tilde{\mathbf{z}}_{k-1}^\dagger \cdot \mathbf{s}} + (\tilde{\mathbf{z}}_{k-2} \cdot h_\alpha[\tilde{\mathbf{x}}_{k-2}])^\dagger \cdot \mathbf{s} \quad (31)$$

C. Bifurcations

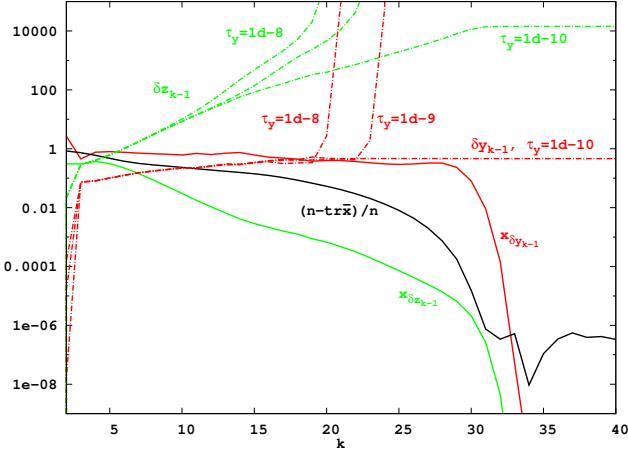


FIG. 2: Derivatives, displacements and the approximate trace of the unscaled, dual NS iteration for a (3,3) nanotube with $\kappa = 10^{10}$. Derivatives are full lines, whilst the displacements cooresponding to $b = 64$, $\tau = 10^{-3}$ and $\tau_y = \{10^{-8}, 10^{-9}, 10^{-10}\}$ are the dashed lines.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

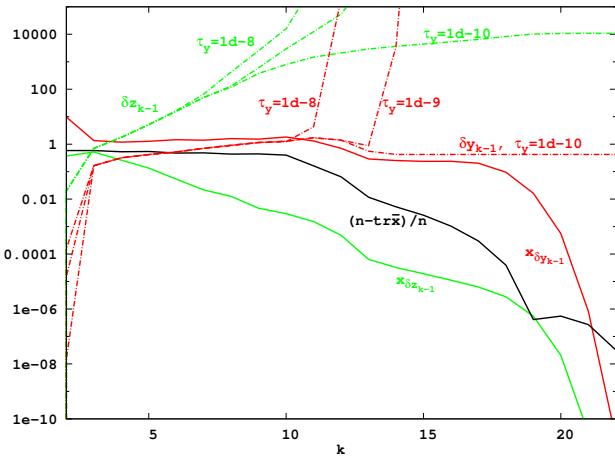


FIG. 3: Derivatives, displacements and the approximate trace of the scaled, stabilized NS iteration for a (3,3) nanotube with $\kappa = 10^{10}$. Derivatives are full lines, whilst the displacements cooresponding to $b = 64$, $\tau = 10^{-3}$ and $\tau_y = \{10^{-3}, 10^{-4}, 10^{-6}\}$ are the dashed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

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).

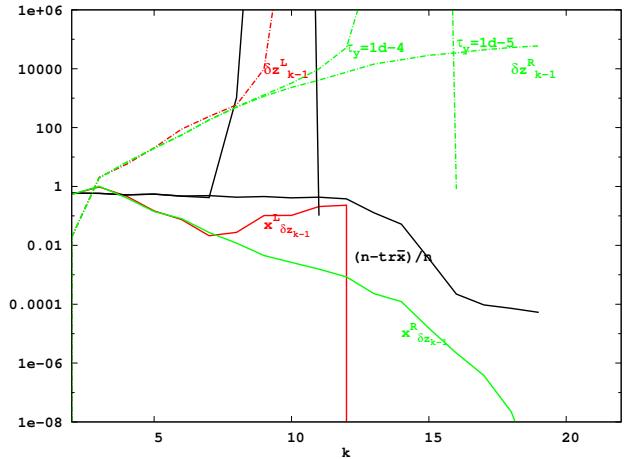


FIG. 4: Derivatives, displacements and the approximate trace of the unscaled, dual NS iteration for a (3,3) nanotube with $\kappa = 10^{10}$. Derivatives are full lines, whilst the displacements cooresponding to $b = 64$, $\tau = 10^{-3}$ and $\tau_y = \{10^{-8}, 10^{-9}, 10^{-10}\}$ are the dashed lines.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

V. TEMPERRY: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

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

Then define $S_{j+1} = R_j \otimes_{\tau_j} S_j \otimes_{\tau_j} R_j$ 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

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

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

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

In general, a defining

$$R_{\text{left}} \equiv R_{j+1} \otimes_{\tau_{j+1}} R_j \otimes_{\tau_j} R_{j-1} \cdots \otimes_{\tau_1} R_0,$$

and

$$R_{\text{right}} \equiv R_0 \otimes_{\tau_1} R_1 \cdots \otimes_{\tau_j} R_j \cdots \otimes_{\tau_{j+1}} R_{j+1},$$

it follows by induction that

$$\|R_{\text{left}} S R_{\text{right}} - I\|_F \lesssim \tau_{j+1}.$$

Since (NOTE: NEED TO CHECK BUT THINK THIS IS TRIVIAL),

$$\|R_{\text{left}} R_{\text{right}}^{-1} - I\|_F \lesssim \tau_{j+1},$$

we see that R_{left} is a τ_{j+1} approximation to $S^{-1/2}$.

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 $S_k = S_{\tau_{k-1}}^{-1/2} \otimes_{\tau_{k-1}} S_{\tau_{k-1}} \otimes_{\tau_j} S_{\tau_{k-1}}^{-1/2}$.

VI. ITERATED REGULARIZATION

Shown in the preceeding section, stability limits application of the NS square root iteration under aggressive SpAMM approximation. These limits can be circumvented through Tikhonov regularization $\|\cdot\|$, involving a small level shift of eigenvalues, $s_\mu \leftarrow s + \mu I$, leading to a more well conditioned matrix with $\kappa(s_\mu) = \frac{\sqrt{s_{N-1}^2 + \mu^2}}{\sqrt{s_0^2 + \mu^2}}$. However, achieving substantial acceleration with severe ill-conditioning may require a large level shift, producing inverse factors of little practical use. One approach to recover a more accurate inverse factor is Riley's method $\|\cdot\|$;

$$s^{-1/2} = s_\mu^{-1/2} \cdot \left(I + \frac{\mu}{2} s_\mu^{-1} + \frac{3\mu^2}{8} s_\mu^{-2} + \dots \right), \quad (32)$$

but this is ineffective when μ is large, and involves powers of the full inverse.

Here, we outline an alternative, nested product representation of the inverse factor and show preliminary results for a first, most approximate solution. 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, 0th preconditioner, $s_{\tau_0 \mu_0}^{-1/2}$. This “thin” iteration may bring spectral resolution into alignment with norm magnitudes

towards the resolvent $I_{\tau_0 \mu_0} \equiv \tilde{I}(s_{\tau_0 \mu_0})$, strengthening Eq. (5).

Culled SpAMM volumes for this most approximate solution are shown with increasing system size in Fig. VI for the stable instance, and in Fig. VI for the dual instance, cooresponding to “thin NS” iteration for the (3,3) $\kappa(s) = 10^{10}$ nanotube series. The behavior of these instances is very different; in the “stable” instance, a stable iteration could not be found at precision $\tau_0 = .1$, even with $\mu_0 = .1$ regularization. Also, this stable iteration sees a weakly convergent trace with inflating cull-volumes. On the other hand, volume of the dual iteration is strongly contracted with resolution of the identity.

most-approximate-yet-effective-by-one-order
(MAYEBOO)

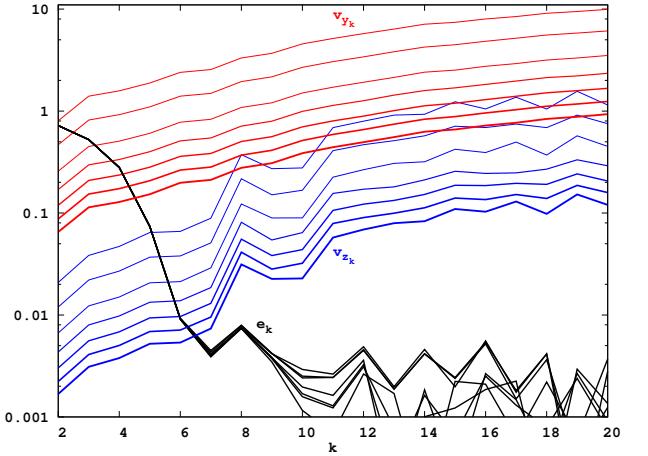


FIG. 5: Culled volumes in the thin slice, stable instance approximation of $s_{\tau_0 \mu_0}^{-1/2}$ for the (3,3) nanotube, $\kappa(s) = 10^{10}$ matrix series described in Section B. In the “stable” instance, it was not possible to achieve stability with $\tau_0 = .1$. In this “stable” case, a thin slice cooresponds to $\mu_0 = .1, \tau_0 = 10^{-2}$ & $\tau_s = 10^{-4}$, and volumes are $v_{\tilde{z}_k} = (\text{vol}_{\tilde{z}_{k-1} \otimes_{\tau_h} [\tilde{x}_{k-1}]}) \times 100\% / N^3$ and $v_{\tilde{y}_k} = (\text{vol}_{s \otimes_{\tau_s} \tilde{z}_k}) \times 100\% / N^3$. Line width increases with increasing system size. Also shown is the trace error, $e_k = (N - \text{tr } \mathbf{x}_k) / N$.

These results reflect very different cull-spaces. In the stable instance, the spectral resolution of powers is not compressive, and $\tilde{y}_k^{\text{stab}} \rightarrow s_{\tau_0 \mu_0}^{-1/2} \otimes_{\tau_0} s_{\mu_0}$ is poorly bound by Eq. (5). In the dual case however, $\tilde{y}_k^{\text{dual}} \rightarrow I_{\tau_0 \mu_0} \otimes_{\tau_0} s_{\tau_0 \mu_0}^{1/2}$ and $\tilde{z}_k^{\text{dual}} \rightarrow s_{\tau_0 \mu_0}^{-1/2} \otimes_{\tau_0} I_{\tau_0 \mu_0}$, with Eq. (5) tightening to

$$\Delta I_{\tau_0 \mu_0} \cdot s_{\tau_0 \mu_0}^{1/2} < \tau n \|s_{\tau_0 \mu_0}^{1/2}\| \quad (33)$$

and

$$\Delta s_{\tau_0 \mu_0}^{-1/2} \cdot I_{\tau_0 \mu_0} < \tau n \|s_{\tau_0 \mu_0}^{-1/2}\|, \quad (34)$$

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 $\mu_0 = .1$.

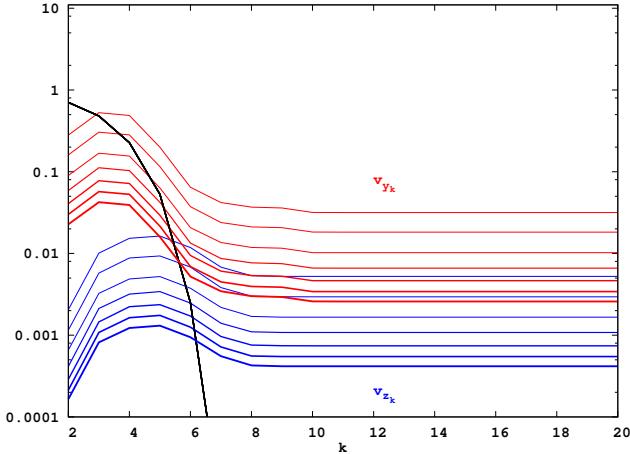


FIG. 6: Culled volumes in the thin slice, dual instance approximation of $\mathbf{s}_{\tau_0 \mu_0}^{-1/2}$ for the (3,3) nanotube, $\kappa(\mathbf{s}) = 10^{10}$ matrix series described in Section B. The thin slice corresponds to $\mu_0 = .1$, $\tau_0 = .1$ & $\tau_s = .001$ with volumes $v_{\tilde{y}_k}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

From this first slice $\mathbf{s}_{\tau_0, \mu_0}^{-1/2}$ then, a next level shifted preconditioner can be found, $\mathbf{s}_{\tau_0 \mu_1}^{-1/2}$, based on the residual $(\mathbf{s}_{\tau_0 \mu_0}^{-1/2})^\dagger \otimes_{\tau_0} (\mathbf{s} + \mu_1 \mathbf{I}) \otimes_{\tau_0} \mathbf{s}_{\tau_0 \mu_0}^{-1/2}$, 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;

$$\mathbf{s}_{\tau_0}^{-1/2} = \mathbf{s}_{\tau_0 \mu_n}^{-1/2} \otimes_{\tau_0} \mathbf{s}_{\tau_0 \mu_{n-1}}^{-1/2} \otimes_{\tau_0} \dots \mathbf{s}_{\tau_0 \mu_0}^{-1/2} \quad (35)$$

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 permissive than otherwise possible. Likewise, it may be possible to obtain the full factor with increasing SpAMM resolution in the product representation:

$$\mathbf{s}^{-1/2} = \mathbf{s}_{\tau_m}^{-1/2} \otimes_{\tau_m} \mathbf{s}_{\tau_{m-1}}^{-1/2} \otimes_{\tau_{m-1}} \dots \mathbf{s}_{\tau_0}^{-1/2} \quad (36)$$

taken over the sequence $1 > \tau_0 > \tau_1 > \dots > \tau_n$. More generally we write

$$\mathbf{s}^{-1/2} \equiv \bigotimes_{\substack{\tau=\tau_0 \\ \mu=\mu_0}} |\tau \mu; \mathbf{s}^{-1/2}\rangle, \quad (37)$$

acknowledging the potential for a flexible path between precision and regularization. The bracket notation marks the potential for asymmetries in the intermediate representation.

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 strengthened, via Eqs. (33-34); (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 amenable to continuous temporal partitioning based on e.g. persistence data.

VII. 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 Cartesian 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 Figures 7 and 8, we demonstrate 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 deflation of the culled volume about the resolvent's plane diagonal. This is important for the product yeilding \mathbf{y}_k , because iteration in the channel begins with the thicker product $\mathbf{s} \otimes_{\tau_s} \mathbf{s}$ and involves the tighter threshold $\tau_s \sim 0.01 \times \tau$.

It is also possible to achieve lensing without regularization as shown in Figure 9. In Figure 9 however, lensing of the \mathbf{y}_k channel is not as strongly contractive as in the regularized case (Figure 8), showing additionally the $i = k$ plane and pillae normal to ik . These additional features and delocalizations coorespond to weakness in Eq. (5), and to the tighter thresholds required to maintain a stable iteration. This effect is even more pronounced in the stable instance (not shown), where delocalizations are exaggerated due to broader spectral resolutions.

The density of pillae observed in Figure 9 is associated with increasing dimensionallity and diminishing Cartesian seperations (nano-tube vs periodic water). For increasingly large systems, the decay of this density is dependent on the underlying Euclidean locality. In Figure 11, we show cull-volumes for periodic water systems ordered by the Hilbert Space Filling Curve (HSFC) [], and also by the Travel Order solution of the end-to-end traveling salesman problem (TO) [] (discussed in Section B).

In Figure 11, we see

In addition, non-Euclidian measures are relevant 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 10 we show development of a first, unregularized preconditioner for such an exam-

ple; the structural matrix $\mathbf{s} = \text{bcssstk14}$ is a $\kappa(\mathbf{s}) = 10^{10}$ 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 \mathbf{y}_k , and hollow accumulation of $\text{vol}_{\mathbf{y}_k \otimes_{\tau} \mathbf{z}_k}$ looking down at \mathbf{y}_k (along bottom).

VIII. CONCLUSIONS

new bound for N-Body approach to fast matrix multiplication: methods that truncate the data do not bound the outcome

stability analyses and bifurcations of ill-conditioned spamm away from basin of stability

show how to go past these bifurcations using regularization. showed how to find a first, most-approximate-yet-effective-by-one-order (MAYEBOO) preconditioner, which brings the norm-bounded and spectral resolutions into close alignment, strengthening Eq. (5).

shows two new kinds of locality: strong Euclidean locality based on Gan and Challacombe's travel order, and algebraic locality associated with lensing, and contracting the culled task volume abount the plane-diagonal of the resolvent (lensing)

product representation of generic MAYEBOO slices with extreme locality of reference, potential for additional acceleration and stabilization of in place data. solution to problem developed from on genericity and concurence of the Erlang kind. Ecosystems impact. algebraic structures cannot be approached by conventional row-coloum orientation 3-dimensional systems that would not be in the conventional (sparse BCSR) regime for n -scaling.

Sandwich approach potential for defered, implicit de-localization of the inverse factor, that can be avoided by incremental application to a target vector or matrix. Potential for incremental preconditionin on multiple time scales most approximate cooresponding to longest time scale. problems in electronic structure,generalized polar decomposition and ill-conditioning in radial basis function problems

Appendix A: Implementation

1. 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 [].

2. scaling

3. stabilization

For these reasons, maintaining connection to the eigenvectors of \mathbf{s} through a tighter first product is necessary. 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

$$\tilde{\mathbf{x}}_k^R \leftarrow \tilde{\mathbf{z}}_k^\dagger \otimes_\tau (\mathbf{s} \otimes_{\tau_s} \tilde{\mathbf{z}}_{k-1}) , \quad (\text{A1})$$

is different from the left first product

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

4. regularization

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

5. convergence

Map switching and etc based on TrX

Appendix B: Data

a. double exponential ill-conditioning

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

b. three-dimensional, periodic

c. Matrix Market

-
- [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. and M. 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/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,

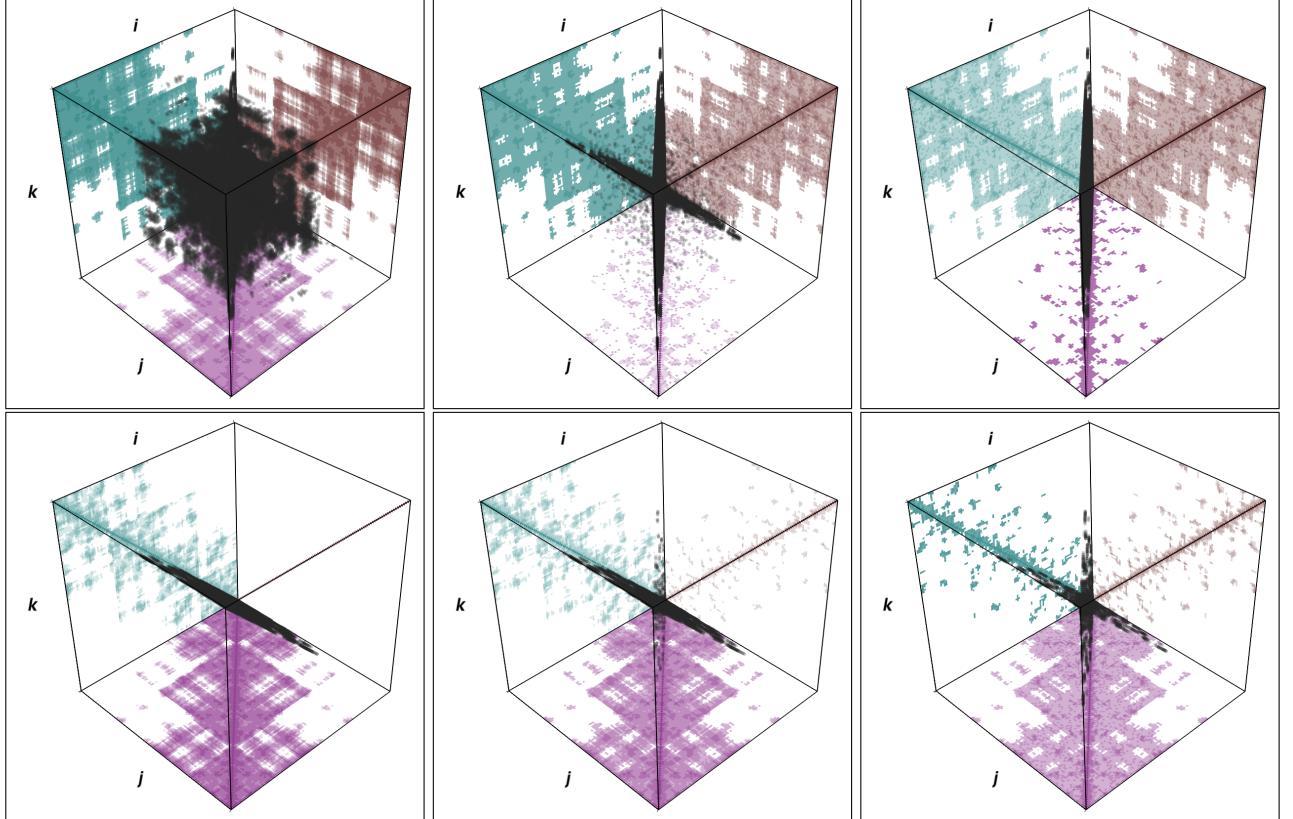


FIG. 7: The ijk task and data space for construction of the MAYEBOO preconditioner $|\tau_0 = .1, \mu_0 = .1; s^{-1/2}\rangle$, with the dual instance of square root iteration and for an $8 \times$ U.C. (3,3) $\kappa(s) = 10^{11}$ nanotube. \mathbf{y}_k appears wider than \mathbf{z}_k because it is computed at a higher precision, $\tau_s = .001$, and because the first multiply involves s^2 . At top its $\mathbf{y}_k = h_\alpha[\mathbf{x}_{k-1}] \otimes_{\tau_s} \mathbf{y}_{k-1}$ for $k = 0, 4, \& 16$, while on the bottom we have $\mathbf{x}_k = \mathbf{y}_k \otimes_\tau \mathbf{z}_k$ for $k = 0, 2, \& 16$. Maroon is \mathbf{a} , purple is \mathbf{b} , green is \mathbf{c} , and black is the volume $\text{vol}_{a \otimes \tau b}$ in the product $\mathbf{c} = \mathbf{a} \otimes_\tau \mathbf{b}$.

- 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).

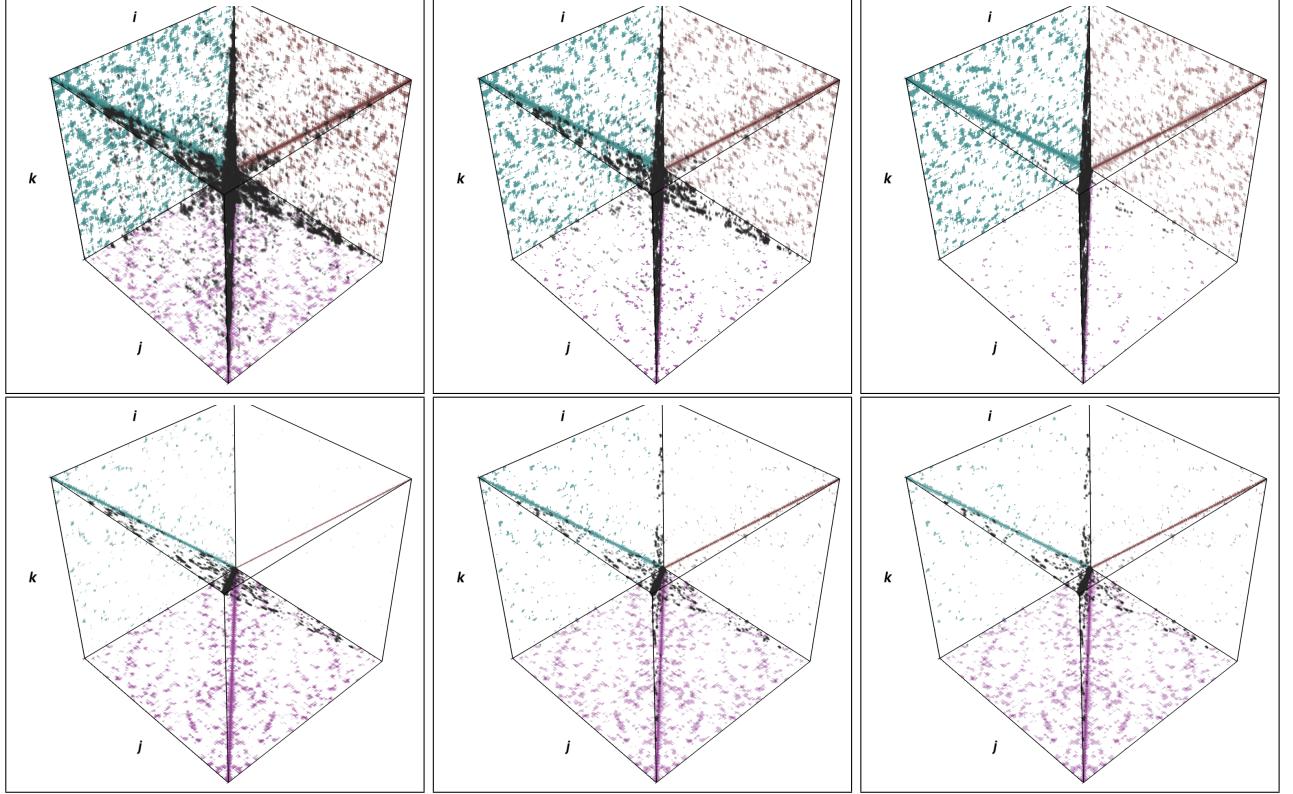


FIG. 8: The ijk task and data space for construction of the MAYEBOO preconditioner $|\tau_0 = .1, \mu_0 = .1; s^{-1/2}\rangle$, with the dual instance of square root iteration and for 6-311G** metric of 100 periodic water molecules at STP. At top its $\mathbf{y}_k = h_\alpha[\mathbf{x}_{k-1}] \otimes_{\tau_s} \mathbf{y}_{k-1}$ for $k = 0, 4, \& 15$, while on the bottom we have $\mathbf{x}_k = \mathbf{y}_k \otimes_{\tau} \mathbf{z}_k$ for $k = 0, 4, \& 15$. Maroon is \mathbf{a} , purple is \mathbf{b} , green is \mathbf{c} , and black is the volume $\text{vol}_{a \otimes \tau b}$ in the product $\mathbf{c} = \mathbf{a} \otimes_{\tau} \mathbf{b}$.

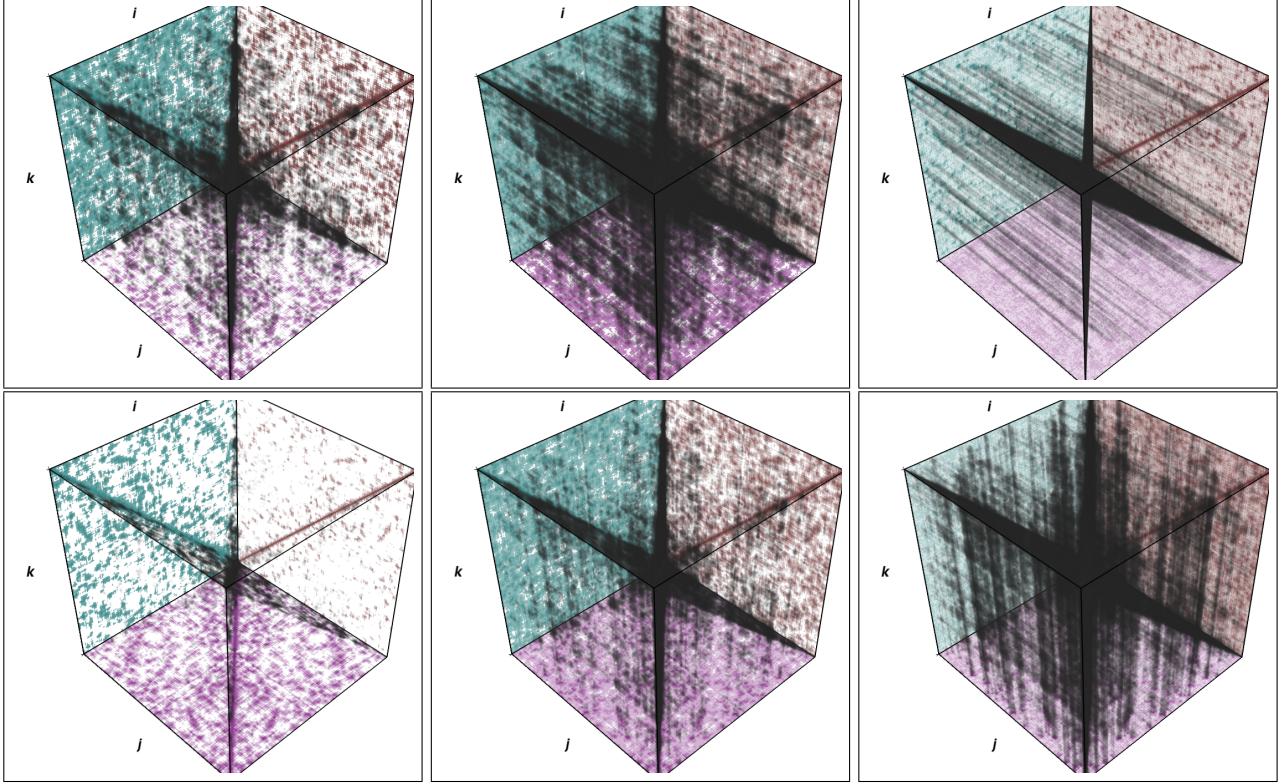


FIG. 9: The ijk task and data space for construction of the unregularized preconditioner $|\tau_0 = .001, \mu_0 = .0; s^{-1/2}\rangle$, with the dual instance of square root iteration and for 6-311G** metric of 100 periodic water molecules at STP. At top its $\mathbf{y}_k = h_\alpha[\mathbf{x}_{k-1}] \otimes_{\tau_s} \mathbf{y}_{k-1}$ for $k = 0, 4, \& 15$, while on the bottom we have $\mathbf{x}_k = \mathbf{y}_{k \otimes \tau} \mathbf{z}_k$ for $k = 0, 4, \& 15$. Maroon is \mathbf{a} , purple is \mathbf{b} , green is \mathbf{c} , and black is the volume $\text{vol}_{a \otimes_{\tau} b}$ in the product $\mathbf{c} = \mathbf{a} \otimes_{\tau} \mathbf{b}$.

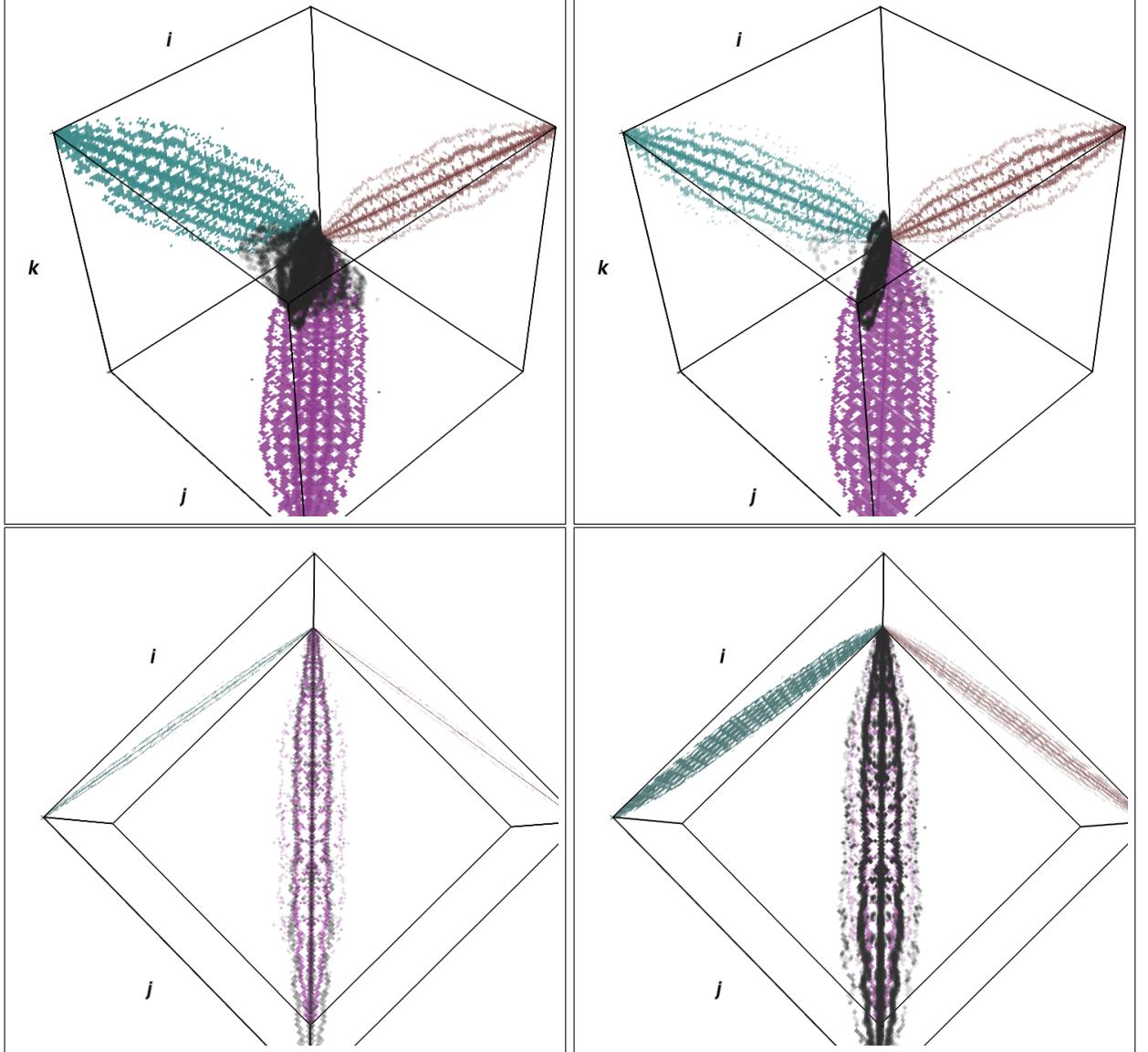


FIG. 10: The ijk task and data space for construction of the unregularized preconditioner $|\tau_0 = .001, \mu_0 = .0; s^{-1/2}\rangle$, with the dual instance of square root iteration and for 6-311G** metric of 100 periodic water molecules at STP. At top its $\mathbf{y}_k = h_\alpha[\mathbf{x}_{k-1}] \otimes_{\tau_s} \mathbf{y}_{k-1}$ for $k = 0, 4, \& 15$, while on the bottom we have $\mathbf{x}_k = \mathbf{y}_k \otimes_{\tau} \mathbf{z}_k$ for $k = 0, 4, \& 15$. Maroon is \mathbf{a} , purple is \mathbf{b} , green is \mathbf{c} , and black is the volume $\text{vol}_{\mathbf{a} \otimes_{\tau} \mathbf{b}}$ in the product $\mathbf{c} = \mathbf{a} \otimes_{\tau} \mathbf{b}$.

FIG. 11: Show percent work for H and TO orderings.