Subspace Injections

Joel A. Tropp
Steele Family Professor of
Applied & Computational Mathematics

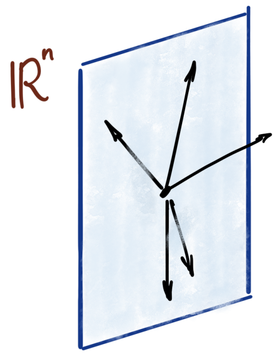
Computing + Mathematical Sciences
California Institute of Technology
jtropp@caltech.edu

Collaborators: Chris Camaño (Caltech), Ethan Epperly (Berkeley), Raphael Meyer (Berkeley)

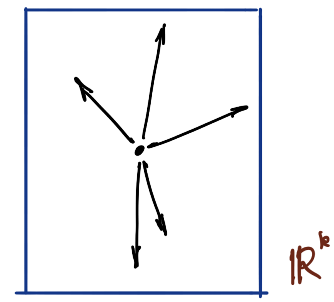
Research supported by ONR and the Caltech Carver Mead Fund and S2I Fund.

Sketching:
Crash Course

Randomized Dimension Reduction



$$\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^k$$



- **Goals:** Reduce dimension, preserve geometry
- **Example:** Johnson–Lindenstrauss
  - Approximate pairwise ℓ_2 distances among N points embedded in $O(\log N)$ dimensions
- **Example:** Subspace embedding
  - Approximate r-dimensional subspace of $\mathbb{R}^n$ embedded in $O(r)$ dimensions

- **How?** Apply a random matrix... Succeeds with high probability
- **Example:** Gaussian matrix
  - Matrix $\mathbf{\Phi} \in \mathbb{R}^{k \times n}$ has iid $\textsc{normal}(0, 1/k)$ entries

Sources: Johnson & Lindenstrauss 1984; Linial et al. 1995; Indyk & Motwani 1998; Frieze et al. 1998; Sarlós 2006; Woodruff 2014; Drineas & Mahoney 2017; Martinsson & Tropp 2020; Kireeva & Tropp 2023; ....

Example: Generalized Nyström

- **Problem:** Find rank-r approximation of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$

Generalized Nyström:

- Draw random embeddings $\boldsymbol{\Psi} : \mathbb{R}^m \to \mathbb{R}^k$ and $\boldsymbol{\Phi} : \mathbb{R}^n \to \mathbb{R}^p$ $\qquad (k = 2r, p = 4r)$
- **Sketch:** Form row $\boldsymbol{X} = \boldsymbol{\Psi}\boldsymbol{A} \in \mathbb{R}^{k \times n}$ and column $\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{A}^* \in \mathbb{R}^{p \times m}$ sketches
- **Solve:** Solve reduced problem:

$$\boldsymbol{M}_\star \quad = \quad (\boldsymbol{\Psi}\boldsymbol{Y})^\dagger \quad \in \quad \arg\min_{\boldsymbol{M} \in \mathbb{R}^{p \times k}} \left\| \boldsymbol{A} - \boldsymbol{Y}^* \boldsymbol{M} \boldsymbol{X} \right\|_{\mathrm{F}}^2$$

- **Approximate:** Construct $\widehat{\boldsymbol{A}} = \boldsymbol{Y}^* \boldsymbol{M}_\star \boldsymbol{X} \in \mathbb{R}^{m \times n}$
- **Goal:** Achieve error bound

$$\| \boldsymbol{A} - \widehat{\boldsymbol{A}} \|_{\mathrm{F}}^2 \quad \leq \quad \mathrm{Const} \cdot \min_{\mathrm{rank}\, \boldsymbol{B} \leq r} \| \boldsymbol{A} - \boldsymbol{B} \|_{\mathrm{F}}^2$$

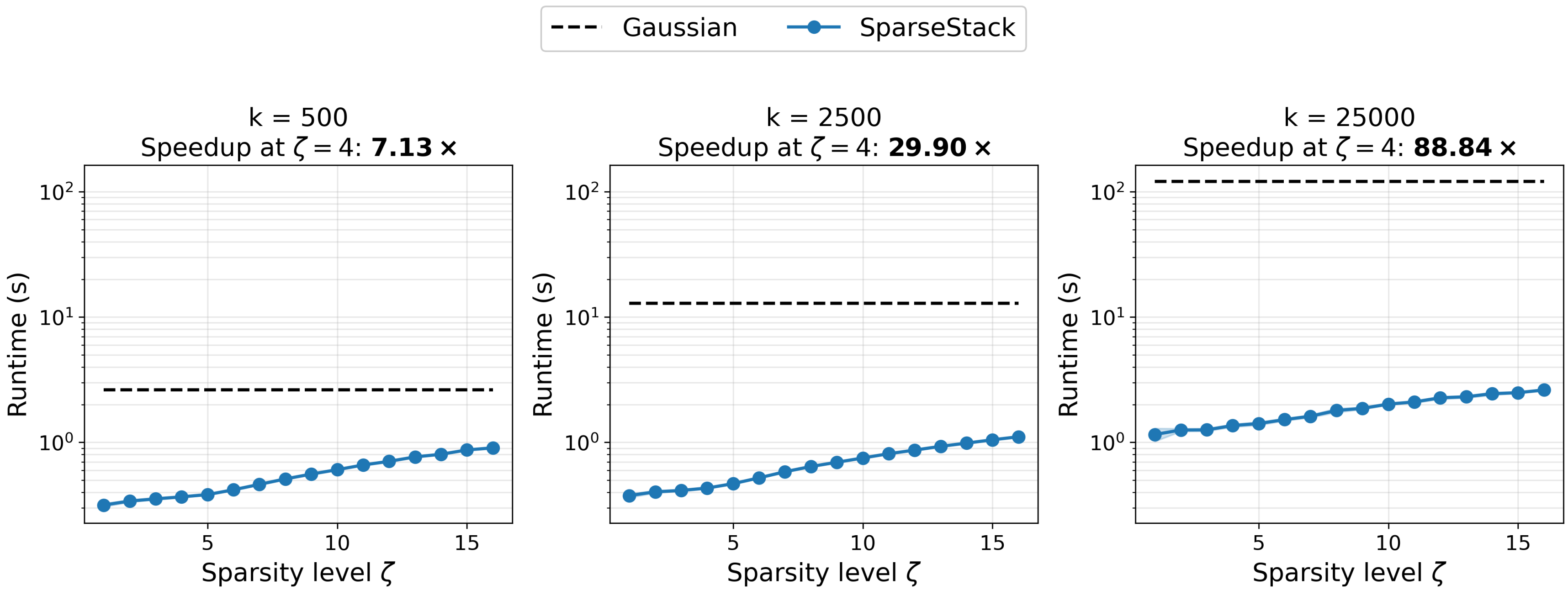
Sources: Woolfe et al. 2008; Clarkson & Woodruff 2009; Halko et al. 2011; Tropp et al. 2017, 2019; Martinsson & Tropp 2020; Nakatsukasa 2020; Nakatsukasa & Tropp 2024; ....

Fast Dimension Reduction

- Dimension reduction requires attention to...
  - **Speed:** Minimize work to reduce problem data [sketching cost]
  - **Compression:** Minimize embedding dimension [postprocessing cost]
  - **Quality:** Preserve relevant geometry [solution quality]
- **Example:** GenNyström with Gaussian matrix
  - **Speed:** Slow... $O(k \cdot \mathrm{nnz}(\boldsymbol{A}))$ for sketching
  - **Compression:** Optimal... $k = 2r$ and $p = 4r$
  - **Quality:** Excellent... $\mathrm{Const} \leq 4$
- **Idea:** Use faster dimension reduction method
- **Example:** GenNyström with SparseStack matrix
  - Sparse embeddings with $\zeta \ll k$ nonzero entries per column
  - **Speed:** Fast... $O(\zeta \cdot \mathrm{nnz}(\boldsymbol{A}))$ for sketching
  - **Sparsity? Compression?**
  - **Quality?**

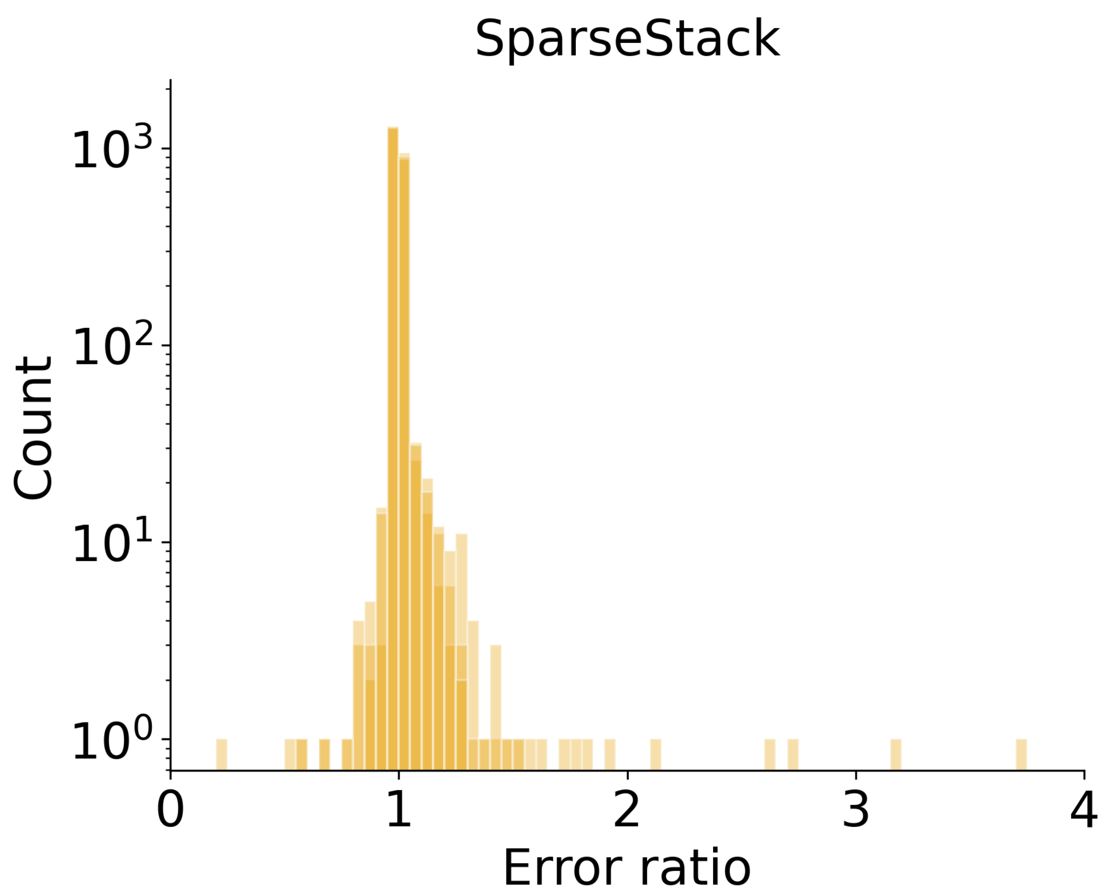
Sources: Achlioptas 2001; Charikar et al. 2003; Ailon & Chazelle 2006; Woolfe et al. 2008; Liberty 2009; Tropp 2011; Clarkson & Woodruff 2011; Nelson & Nguyen 2012, 2013; Kane & Nelson 2014; ....

Speed: Gaussian vs. SparseStack



- Dense input matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ with $n = 50,000$
- Time to form sketch $\boldsymbol{\Phi A} \in \mathbb{R}^{k \times n}$ for embedding dimension $k = 500, 2500, 25000$; median over 10 trials
- Horizontal axis tracks column sparsity ζ of SparseStack; recommended sparsity $\zeta = 4$

Quality: Gaussian vs. SparseStack



- 2,314 matrices from `SparseSuite` with dimensions 300 to 500,000
- RSVD approximation with rank $k = 200$; SparseStack with column sparsity $\zeta = 4$
- Ratio of SparseStack error to Gaussian error (Frobenius norm); 3 trials superimposed

Embeddings vs. Injections

Subspace Embeddings

- Fix an r-dimensional subspace $\mathsf{L} \subseteq \mathbb{R}^n$ (often unknown!)
- Let $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^k$ be a random matrix with *embedding dimension* k

Definition 1 (Subspace embedding)**.** The random matrix $\mathbf{\Phi}$ is an (α, β)*-subspace embedding* for L with *injectivity* $\alpha \in (0, 1]$ and *dilation* $\beta \geq 1$ when it satisfies

$$\alpha \cdot \|\boldsymbol{x}\|^2 \leq \|\mathbf{\Phi}\boldsymbol{x}\|^2 \leq \beta \cdot \|\boldsymbol{x}\|^2 \quad \text{for all } \boldsymbol{x} \in \mathsf{L} \qquad \text{with prob.} \geq 99\%$$

- Can implement linear algebra algorithms with subspace embeddings
- **Challenge:** Establish subspace embedding property for random matrix...
  - With embedding dimension $k = \mathcal{O}(r)$, injectivity $\alpha = \Omega(1)$, and dilation $\beta = \mathcal{O}(1)$
  - With control on other parameters, such as sparsity
  - Gaussian meets these requirements, but what about other examples?

Sources: Sarlós 2006; Clarkson & Woodruff 2009; Woodruff 2014; Drineas & Mahoney 2017; Martinsson & Tropp 2020; Kireeva & Tropp 2023.

Subspace Injections

- Fix an r-dimensional subspace $\mathsf{L} \subseteq \mathbb{R}^n$ (often unknown!)
- Let $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^k$ be a random matrix with *embedding dimension* k

Definition 2 (Subspace injection)**.** The random matrix $\mathbf{\Phi}$ is an *α-subspace injection* for L with *injectivity* $\alpha \in (0, 1]$ when it satisfies

- **Isotropy.** The expectation $\mathbb{E}\,\|\mathbf{\Phi x}\|^2 = \|\boldsymbol{x}\|^2$ for all $\boldsymbol{x} \in \mathbb{R}^n$
- **Injectivity.** With probability at least 99%,

$$\alpha \cdot \|\boldsymbol{x}\|^2 \leq \|\mathbf{\Phi x}\|^2 \quad \text{for all } \boldsymbol{x} \in \mathsf{L}$$

- Want embedding dimension $k = O(r)$ and injectivity $\alpha = \Omega(1)$ for arbitrary L
- Can we implement linear algebra algorithms with subspace injections?
- Why does this concept make our lives better?

Sources: Oymak & Tropp 2018; Martinsson & Tropp 2020; Tropp 2025; [CEMT25].

Linear Algebra with Subspace Injections

Theorem 3 (GenNyström with a Subspace Injection [CEMT25]).
Consider the rank-r approximation problem

$$\text{minimize}_{\operatorname{rank}\boldsymbol{B}\le r} \ \|\boldsymbol{A}-\boldsymbol{B}\|_{\mathrm{F}}^2 \quad \textit{with } \boldsymbol{A}\in\mathbb{R}^{m\times n}$$

Let $\boldsymbol{\Psi}:\mathbb{R}^m\to\mathbb{R}^k$ and $\boldsymbol{\Phi}:\mathbb{R}^n\to\mathbb{R}^p$ be oblivious α-subspace injections for subspaces with dimensions r and k. Then the GenNyström solution $\widehat{\boldsymbol{A}}$ computed with $\boldsymbol{\Phi},\boldsymbol{\Psi}$ satisfies

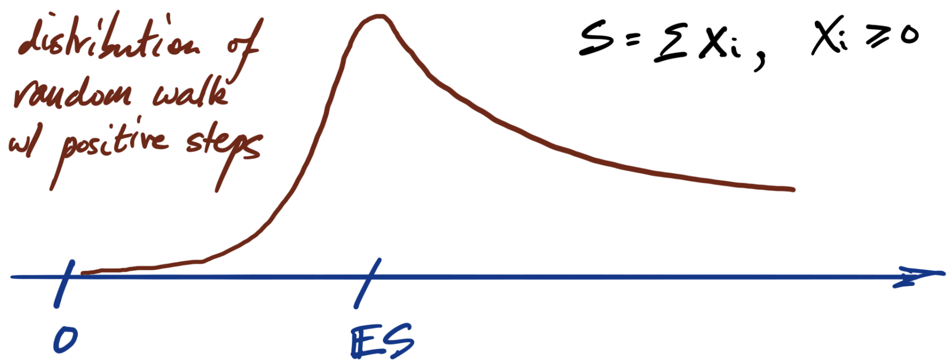
$$\|\boldsymbol{A}-\widehat{\boldsymbol{A}}\|_{\mathrm{F}}^2 \quad \le \quad \frac{\mathrm{Const}}{\alpha^2}\cdot \min_{\operatorname{rank}\boldsymbol{B}\le r} \ \|\boldsymbol{A}-\boldsymbol{B}\|_{\mathrm{F}}^2$$

- Reduce embedding dimension k,p for speed, increase injectivity α for quality
- Can also justify RSVD, Nyström, and Sketch + Solve with subspace injections

- In practice, $(1+\varepsilon)$-optimal error guarantees are overrated

Sources: Martinsson & Tropp 2020; Nakatsukasa & Tropp 2024; [CEMT25].

Injectivity is a Law of Nature



- Assume rows of $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^k$ are iid

$$\|\mathbf{\Phi}\boldsymbol{x}\|^2 = \sum_{i=1}^{k} |\langle \boldsymbol{\varphi}_i,\ \boldsymbol{x}\rangle|^2 = \text{sum of iid positive rvs}$$

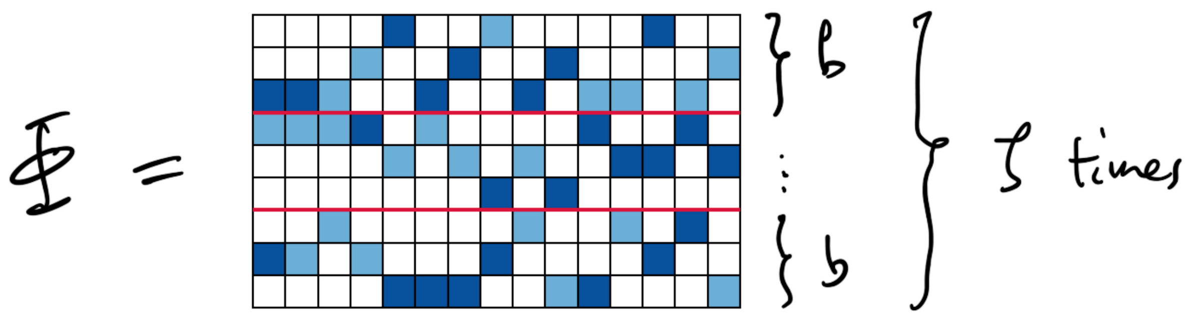
- Norm is large if one summand is large (likely!) $\Rightarrow$ dilation fails
- Norm is small only if all summands are small (unlikely!) $\Rightarrow$ injectivity holds

- Many HDP tools provide control on $\min_{\boldsymbol{x}\in\mathsf{L}} \|\mathbf{\Phi}\boldsymbol{x}\|^2$...
  - Small-ball method [Koltchinskii & Mendelson 2015]
  - Fourth-moment theorem [Oliveira 2016]
  - Gaussian comparison [T25, T26]

Sparse
Stacks

SparseStack Construction

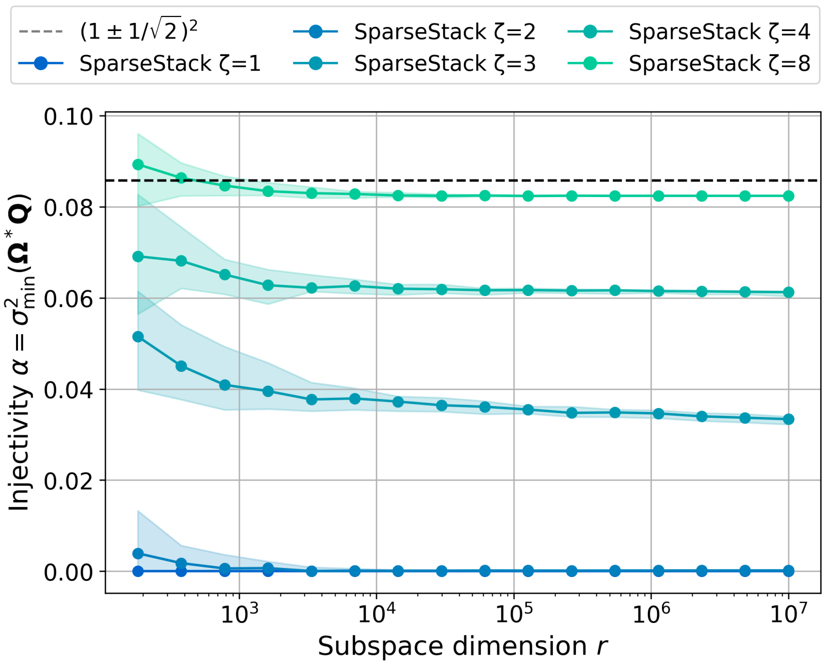
- **SparseStack** matrix $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^k$ with column sparsity ζ
  - Exactly ζ blocks with b rows each, so embedding dimension $k = \zeta b$
  - In each block, each column contains one nonzero entry, in a random position
  - Each nonzero entry is iid UNIFORM$\{\pm\zeta^{-1/2}\}$



- Matvec $\boldsymbol{u} \mapsto \mathbf{\Phi}\boldsymbol{u}$ uses $O(\zeta \cdot \mathrm{nnz}(\boldsymbol{u}))$ arithmetic (with sparse library)
- Extremely effective in practice. But how does it work in theory?

Sources: Nguyen & Nelson 2012, 2013; Kane & Nelson 2014; Cohen 2015; Chenakkod et al. 2023–2025; Tropp 2025; [CEMT25].

SparseStacks are Subspace Injections



Conjecture (Nelson & Nguyen 2013): For any r-dimensional subspace, a SparseStack is a $(1/2, 3/2)$-subspace *embedding* for some $k = O(r)$ and $\zeta = O(\log r)$. Still open!

Theorem 4 (SparseStack [T25, T26]). *For any r-dimensional subspace, a SparseStack serves as a subspace injection with $\alpha = 1/2$ for some $k = O(r)$ and $\zeta = O(\log r)$.*

Sources: Nelson & Nguyen 2013; Cohen 2015; Chenakkod et al. 2023–2025; Tropp 2025, 2026; [CEMT25].

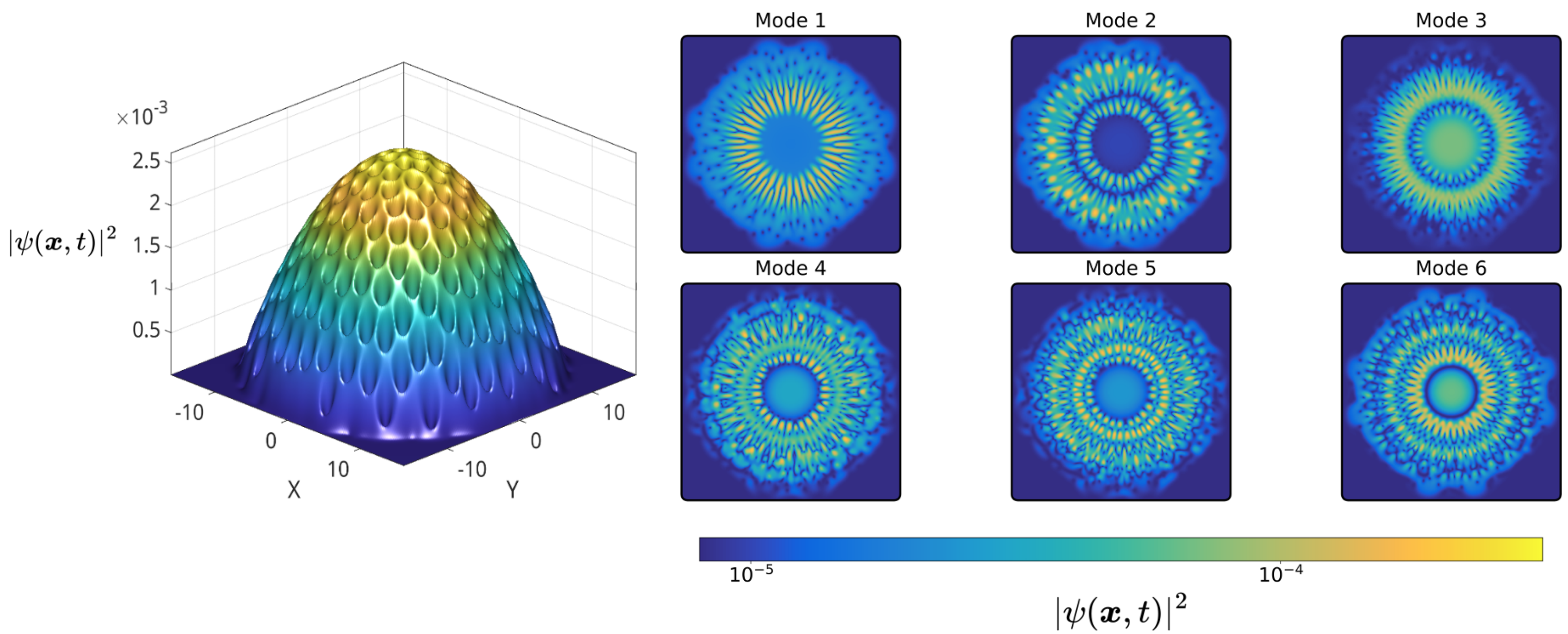
Sparse Linear Algebra with SparseStack

	Best provable runtime	
Algorithm	Gaussian	SparseStack
Sketch + Solve	$\text{nnz}(\boldsymbol{A})\,d + d^3$	$\text{nnz}(\boldsymbol{A})\log(d) + d^3$
GenNyström	$\text{nnz}(\boldsymbol{A})\,r + (n+d)\,r^2$	$\text{nnz}(\boldsymbol{A})\log(r) + (n+d)\,r^2$

- Matrix $\boldsymbol{A}$ has dimension $n \times d$, and r is the approximation rank
- Linear algebra algorithms implemented with Gaussian or with SparseStack matrices
- Runtime to solve problem to constant-factor accuracy
- Big-O symbols suppressed for legibility
- SparseStack results improve over prior work

Sources: Clarkson & Woodruff 2009, 2011; Halko et al. 2011; Nelson & Nguyen 2012; Cohen 2015; Tropp et al. 2017, 2019; Nakatsukasa 2020; Chenakkod et al. 2023–2025; [CEMT25]; ....

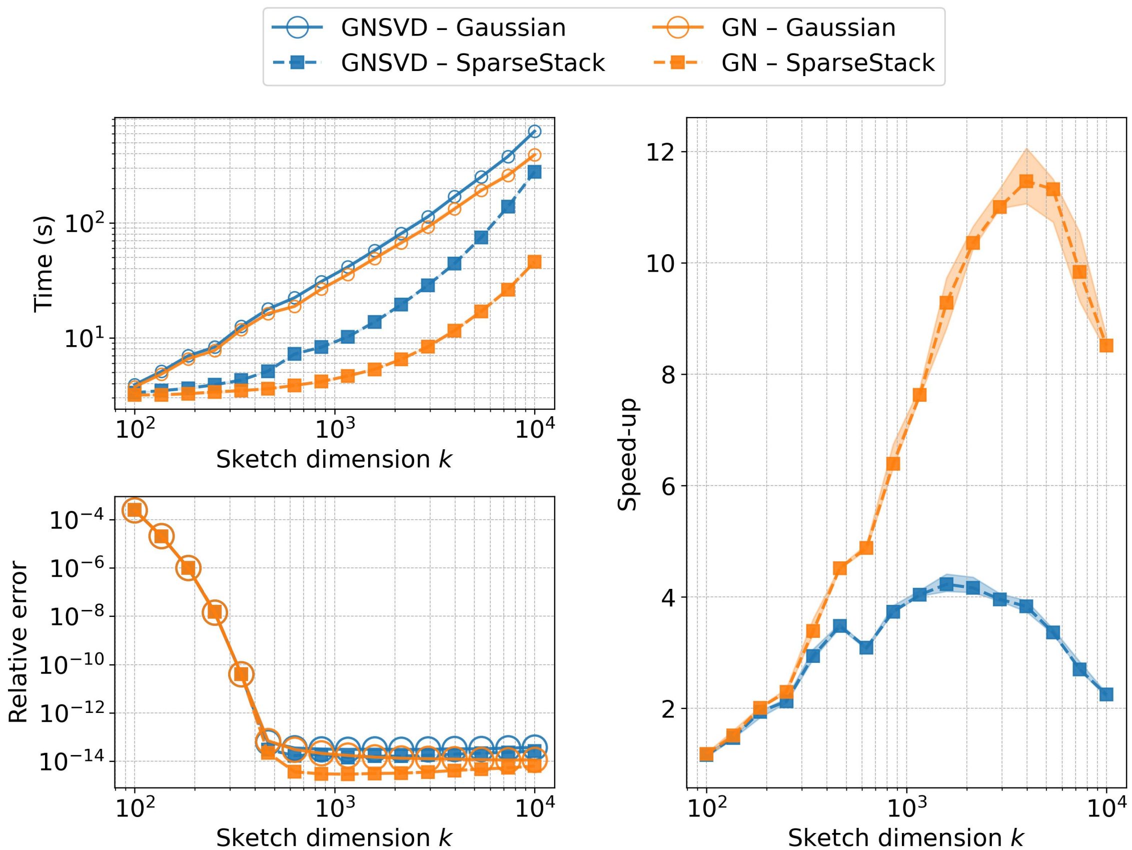
Example: Scientific Simulation with SparseStack



- Simulation of ground state of Bose–Einstein condensate, trapped in potential field, via Gross–Pitaevskii equation
- Use GenNyström SVD for proper orthogonal decomposition (POD)
- Dense matrix $131,072 \times 40,000$; SparseStack with $\zeta = 4$; embedding dimensions $k = 1000$ and $p = 1500$

Source: Tropp et al. 2017, 2019; [CEMT25].

Example: Scientific Simulation with SparseStack



- Simulation of ground state of Bose–Einstein condensate, trapped in potential field, via Gross–Pitaevskii equation
- Use GenNyström for compression and for GenNyström SVD for proper orthogonal decomposition (POD)
- Dense matrix $131,072 \times 40,000$; SparseStack with $\zeta = 4$; embedding dimensions $k = 1000$ and $p = 1.5k$

Khatri–Rao Products

Khatri–Rao Dimension Reduction

- Can we perform efficient dimension reduction for exponentially long vectors?
- **Khatri–Rao** matrix $\mathbf{\Phi} : \bigotimes^{\ell} \mathbb{R}^{d_0} \to \mathbb{R}^k$ with tensor order ℓ
  - Isotropic base distribution $\boldsymbol{\nu} \in \mathbb{R}^{d_0}$ (Spherical, ...)
  - Form $\boldsymbol{\varphi} = \boldsymbol{\omega}^{(1)} \otimes \cdots \otimes \boldsymbol{\omega}^{(\ell)} \in \bigotimes^{\ell} \mathbb{R}^{d_0}$ where $\boldsymbol{\omega}^{(j)} \sim \boldsymbol{\nu}$ iid
  - Matrix $\mathbf{\Phi}$ has iid rows $\boldsymbol{\varphi}_i \sim k^{-1/2} \boldsymbol{\varphi}$
- **Applications:**
  - Bilinear access model (Nonlinear eigenvalues, operator learning, ...)
  - Matrix equations (Sylvester, Lyapunov, Riccati, ...)
  - Tensor algorithms (CP decomposition, ...)
  - Quantum science (Partition functions, ground states, ...)
- Exciting recent research, but not well understood
- How to analyze linear algebra algorithms with Khatri–Rao dimension reduction?

Sources: Rudelson 2012; Sun et al. 2018; Jin et al. 2020; Malik & Becker 2020; Rebrova et al. 2021; Bujanović et al. 2025; Meyer et al. 2023–2025; Saibaba et al. 2025; [CEMT25]; ....

(Some) Khatri–Rao Matrices are Subspace Injections

Theorem 5 (Khatri–Rao [CEMT25]).

- *Let* $\mathbf{\Phi} : \bigotimes^{\ell} \mathbb{R}^{d_0} \to \mathbb{R}^k$ *be a Khatri–Rao matrix*
- *Real spherical base distribution* $\boldsymbol{\nu} \sim \text{UNIFORM}\{\boldsymbol{u} \in \mathbb{R}^{d_0} : \|\boldsymbol{u}\|^2 = d_0\}$
- *For any* r*-dimensional subspace,* $\mathbf{\Phi}$ *is an* α*-subspace injection with*
  - *Injectivity* $\alpha = 1/2$ *for some embedding dimension* $k = O(3^{\ell} r)$
  - *Injectivity* $\alpha = \mathrm{e}^{-O(\ell)}$ *for some embedding dimension* $k = O(r)$
- *Both results are qualitatively correct for worst-case subspaces*

- $\alpha = 1/2$ uses fourth-moment method; improvements for small d_0; extends to many distributions
- $\alpha = \mathrm{e}^{-O(\ell)}$ uses small-ball method; somewhat delicate
- But dilation β expected to grow like $(\log r)^{\ell}$

Sources: Koltchinskii & Mendelson 2015; Oliveira 2016; Guédon et al. 2015; Hu & Paouris 2024; Meyer et al. 2023–2025; Tropp 2025; [CEMT25]; Saibaba 2025; Bandeira et al. 2025; ….

Linear Algebra with Khatri–Rao Matrices

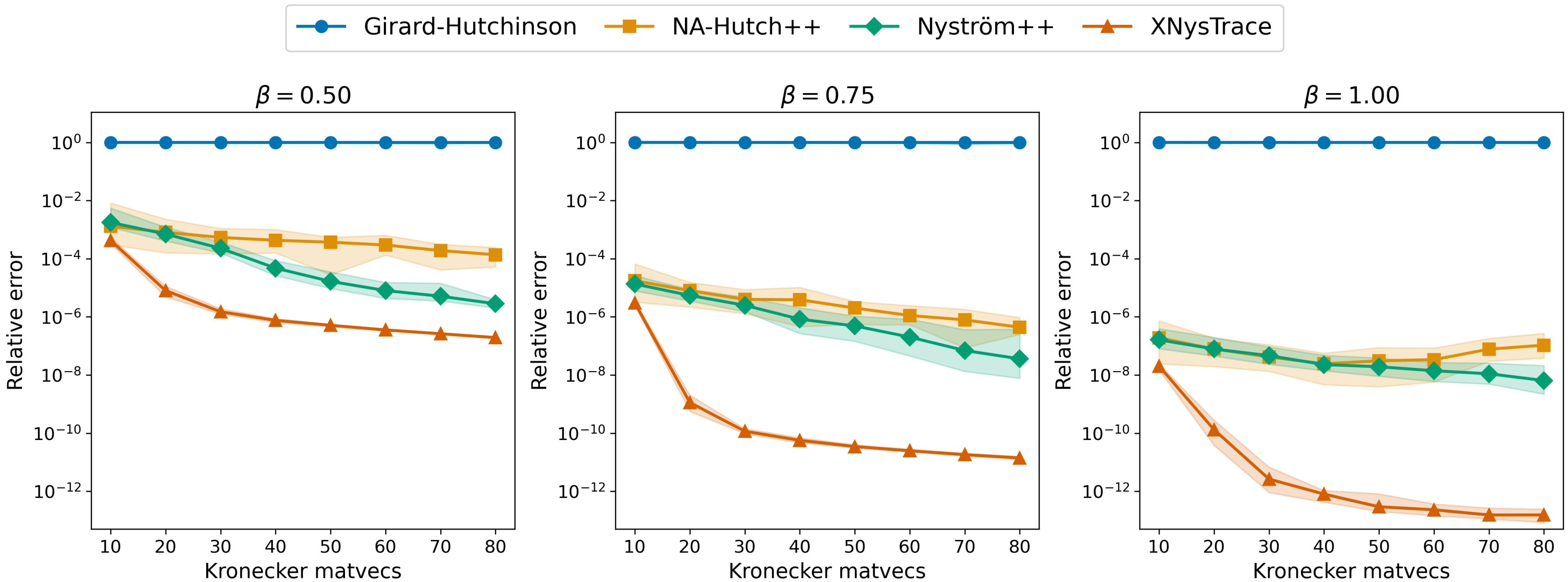
Theorem 6 (Matrix Recovery from Bilinear Queries [CEMT25]).

- *Fix an* r*-dimensional subspace of matrices* $\mathcal{F} := \operatorname{span}\{\boldsymbol{M}_1, \dots, \boldsymbol{M}_r\} \subseteq \mathbb{R}^{n \times n}$
  - *Banded matrices, Toeplitz, Hankel, ...*
- *Fix a target matrix* $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ *with bilinear access* $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x}^{\mathrm{t}} \boldsymbol{B} \boldsymbol{y}$
- *Form a real spherical Khatri–Rao matrix* $\boldsymbol{\Phi} : \mathbb{R}^{n \times n} \to \mathbb{R}^k$ *with* $k = O(r)$
- *Given data* $\boldsymbol{\Phi}(\boldsymbol{B}) \in \mathbb{R}^k$*, use Sketch + Solve to find* $\widetilde{\boldsymbol{B}} \in \mathcal{F}$
- *Then the approximation satisfies the error bound*

$$\|\boldsymbol{B} - \widetilde{\boldsymbol{B}}\|_{\mathrm{F}}^2 \quad \leq \quad \mathrm{Const} \cdot \min_{\boldsymbol{M} \in \mathcal{F}} \|\boldsymbol{B} - \boldsymbol{M}\|_{\mathrm{F}}^2 \quad \textit{with prob.} \geq 99\%$$

- No previous work would imply correct sample complexity $k = \Theta(r)$

Example: Partition Functions via XNysTrace



- Partition function of transverse-field Ising model $\operatorname{trace}[\boldsymbol{A}] = \operatorname{trace}[\mathrm{e}^{-\beta \boldsymbol{H}}]$ with Hamiltonian $\boldsymbol{H} \in (\mathbb{C}^2)^\ell \times (\mathbb{C}^2)^\ell$
- Khatri–Rao matrix $\boldsymbol{\Phi} : (\mathbb{C}^2)^\ell \times (\mathbb{C}^2)^\ell \to \mathbb{C}^k$ with $\ell = 16$
- Baseline: Girard–Hutchinson estimator $\operatorname{trace}[\boldsymbol{A}] \approx \operatorname{trace}[\boldsymbol{\Phi}^*(\boldsymbol{A}\boldsymbol{\Phi})]$
- Comparison: Variance-reduced trace estimators, using low-rank approximation
- Low-rank approximation with Khatri–Rao justified via subspace injection theory

Sources: Girard 1989; Hutchinson 1990; Meyer et al. 2021; Epperly et al. 2024; [CEMT25]; ....

To learn more...

Email: `jtropp@caltech.edu` **Web:** `https://tropp.caltech.edu/`

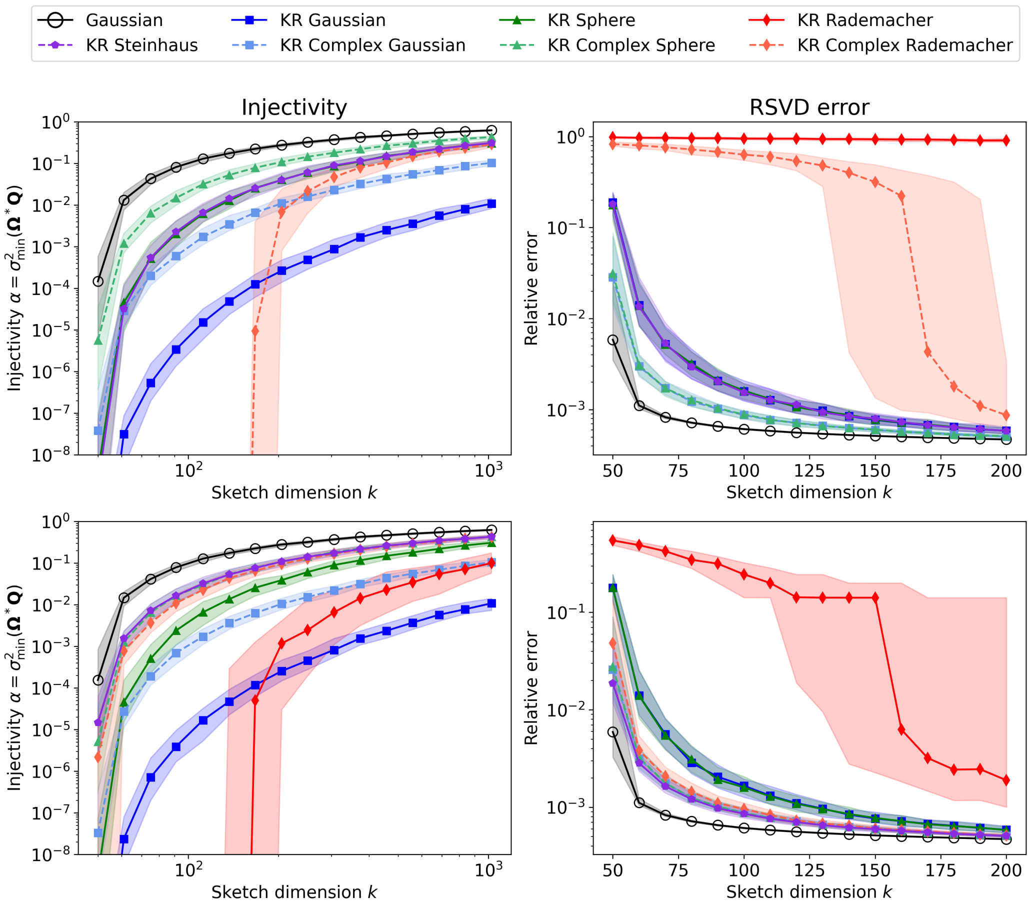
Related Papers:

- Camaño, Epperly, Meyer & Tropp, "Faster linear algebra algorithms with structured random matrices," arXiv:2508.21189
- Camaño, Epperly & Tropp, "Successive randomized compression: A randomized algorithm for the compressed MPO-MPS product," arXiv:2504.06475
- Tropp, "Comparison theorems for the minimum eigenvalue of a random positive-semidefinite matrix," arXiv:2501.16578
- Tropp, "Comparison theorems for the maximum eigenvalue of a random symmetric matrix," forthcoming
- Epperly, Tropp & Webber, "XTrace: Making the most of every sample in stochastic trace estimation," *SIMAX* 2024, arXiv:2301.07825
- Nakatsukasa & Tropp, "Fast & accurate randomized algorithms for linear systems and eigenvalue problems," *SIMAX* 2024, arXiv:2111.00113
- Tropp, Yurtsever, Udell & Cevher, "Streaming low-rank matrix approximation with an application to scientific simulation," *SISC* 2019, arXiv:1902.08651
- Sun, Guo, Tropp & Udell, "Tensor random projection for low-memory dimension reduction," *NeurIPS* 2018, arXiv:2105.00105
- Tropp, Yurtsever, Udell & Cevher, "Practical sketching algorithms for low-rank matrix approximation," *SIMAX* 2017, arXiv:1609.00048

Surveys:

- Kireeva & Tropp, "Randomized matrix computations: Themes and variations," *CIME Lecture Notes*, arXiv:2402.17873
- Tropp & Webber, "Randomized algorithms for low-rank matrix approximation: Design, analysis, and applications," arXiv:2306.12418
- Martinsson & Tropp, "Randomized numerical linear algebra: Foundations and algorithms," *Acta Numerica* 2020, arXiv:2002.01387
- Tropp, "An introduction to matrix concentration inequalities," *Found. Trends Mach. Learning* 2015, arXiv:1501.01571
- Halko, Martinsson & Tropp, "Finding structure with randomness: Probabilistic algorithms for computing approximate matrix decompositions," *SIREV* 2011, arXiv:0909.4061

Role of Khatri–Rao Base Distribution



- Khatri–Rao matrices with base dimension $d_0 = 2$ and tensor order $\ell = 10$, so dimension $n = 2^{10}$
- [left] Injectivity estimate for hard subspace with $r = 50$; [right] RSVD approximation of hard $n \times n$ matrix

Sources: Meyer et al. 2023–2025; [CEMT25].