

Overwhelming Orthogonality

and the Mysteries of Tensor-Structured Sketching

Presented by Raphael Meyer

Joint work with Haim Avron, Chris Camaño, Ethan Epperly, Will Swartworth, Joel Tropp, and David Woodruff

Agenda

1. Upper bounds that look bad

Trace Estimation

2. Lower bounds saying that the world is mean

Information-Theoretic Lower Bounds

3. Upper bounds that give a bit more hope

Low-Rank Approximation

Fast Numerical Linear Algebra

The Usual Story:

Someone has a large matrix $A \in \mathbb{R}^{d \times d}$

They want to compute something about A
top eigenvectors, log determinant

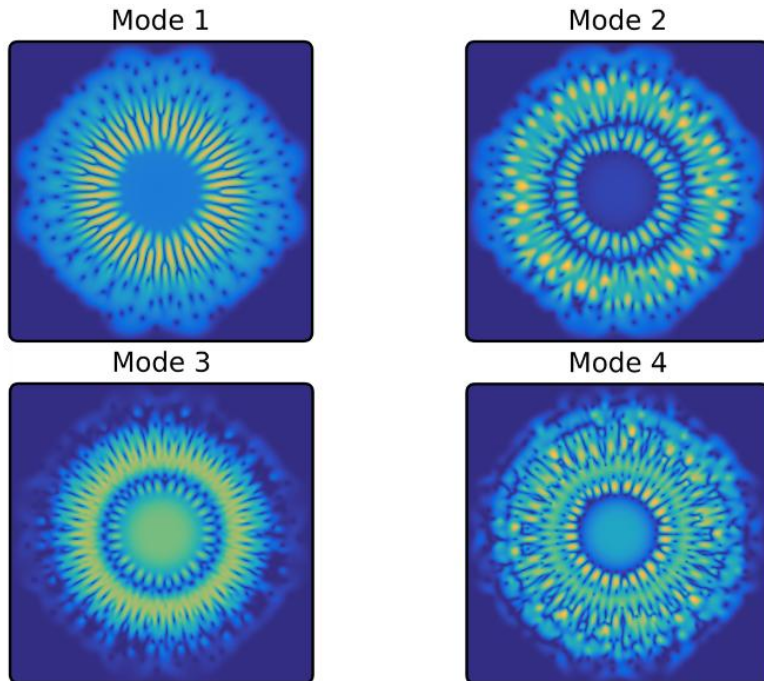
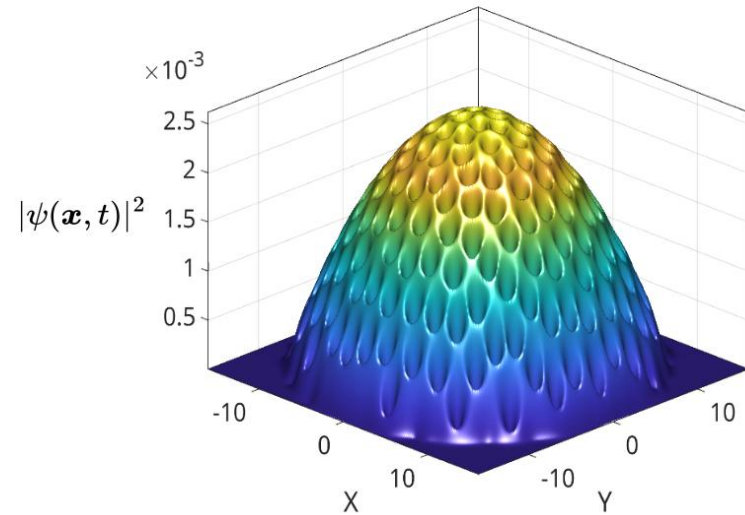
$O(d^3)$ time is always possible!

Goal: quadratic dependence on d : $O(d^2 \text{poly}(1/\epsilon))$

Today:

What if A is impossibly large?

What still works? What goes wrong?



Linear Algebra Review

Let $A \in \mathbb{R}^{d \times d}$ be symmetric. We say A is PSD if all eigenvalues $\lambda_i \geq 0$. A will always be PSD.

Kronecker / Tensor product

Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$. Then $A \otimes B \in \mathbb{R}^{mp \times nq}$

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$$

Some examples:

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 10 \\ 100 \end{bmatrix} = \begin{bmatrix} 2 \\ 20 \\ 200 \\ 3 \\ 30 \\ 300 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Mixed Product Property:

$$(A \otimes B)(x \otimes y) = Ax \otimes By$$

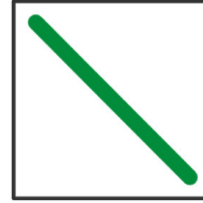
Act 1: Trace Estimation

Trace Estimation

Trace Estimation

The trace is the sum of the diagonal

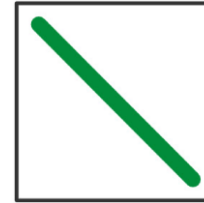
$$\text{tr}(\mathbf{A}) = \sum_{i=1}^d a_{ii} = \sum_{i=1}^d \lambda_i$$



Trace Estimation

The trace is the sum of the diagonal

$$tr(\mathbf{A}) = \sum_{i=1}^d a_{ii} = \sum_{i=1}^d \lambda_i$$



In applications, $\mathbf{A}$ is not in memory.

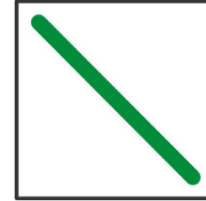
Instead, $\mathbf{B}$ is in memory and $\mathbf{A} = f(\mathbf{B})$.

No. Triangles	Partition Function	Log Determinant
$tr(\frac{1}{6}\mathbf{B}^3)$	$tr(e^{-\mathbf{B}})$	$tr(\ln(\mathbf{B}))$

Trace Estimation

The trace is the sum of the diagonal

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^d a_{ii} = \sum_{i=1}^d \lambda_i$$



In applications, $\mathbf{A}$ is not in memory.

Instead, $\mathbf{B}$ is in memory and $\mathbf{A} = f(\mathbf{B})$.

No. Triangles	Partition Function	Log Determinant
$\text{tr}(\frac{1}{6}\mathbf{B}^3)$	$\text{tr}(e^{-\mathbf{B}})$	$\text{tr}(\ln(\mathbf{B}))$

Computing $\mathbf{B}^3$ takes $O(d^3)$ time. *Slow!*

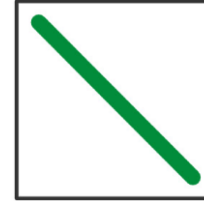
Computing $\mathbf{B}^3 x = \mathbf{B}(\mathbf{B}(\mathbf{B}x))$ takes $O(d^2)$ time. *Fast!*

If $\mathbf{A} = f(\mathbf{B})$ we can compute $f(\mathbf{B})x$ quickly.

Trace Estimation

The trace is the sum of the diagonal

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^d a_{ii} = \sum_{i=1}^d \lambda_i$$



In applications, $\mathbf{A}$ is not in memory.

Instead, $\mathbf{B}$ is in memory and $\mathbf{A} = f(\mathbf{B})$.

No. Triangles	Partition Function	Log Determinant
$\text{tr}(\frac{1}{6}\mathbf{B}^3)$	$\text{tr}(e^{-\mathbf{B}})$	$\text{tr}(\ln(\mathbf{B}))$

Computing $\mathbf{B}^3$ takes $O(d^3)$ time. *Slow!*

Computing $\mathbf{B}^3 x = \mathbf{B}(\mathbf{B}(\mathbf{B}x))$ takes $O(d^2)$ time. *Fast!*

If $\mathbf{A} = f(\mathbf{B})$ we can compute $f(\mathbf{B})x$ quickly.

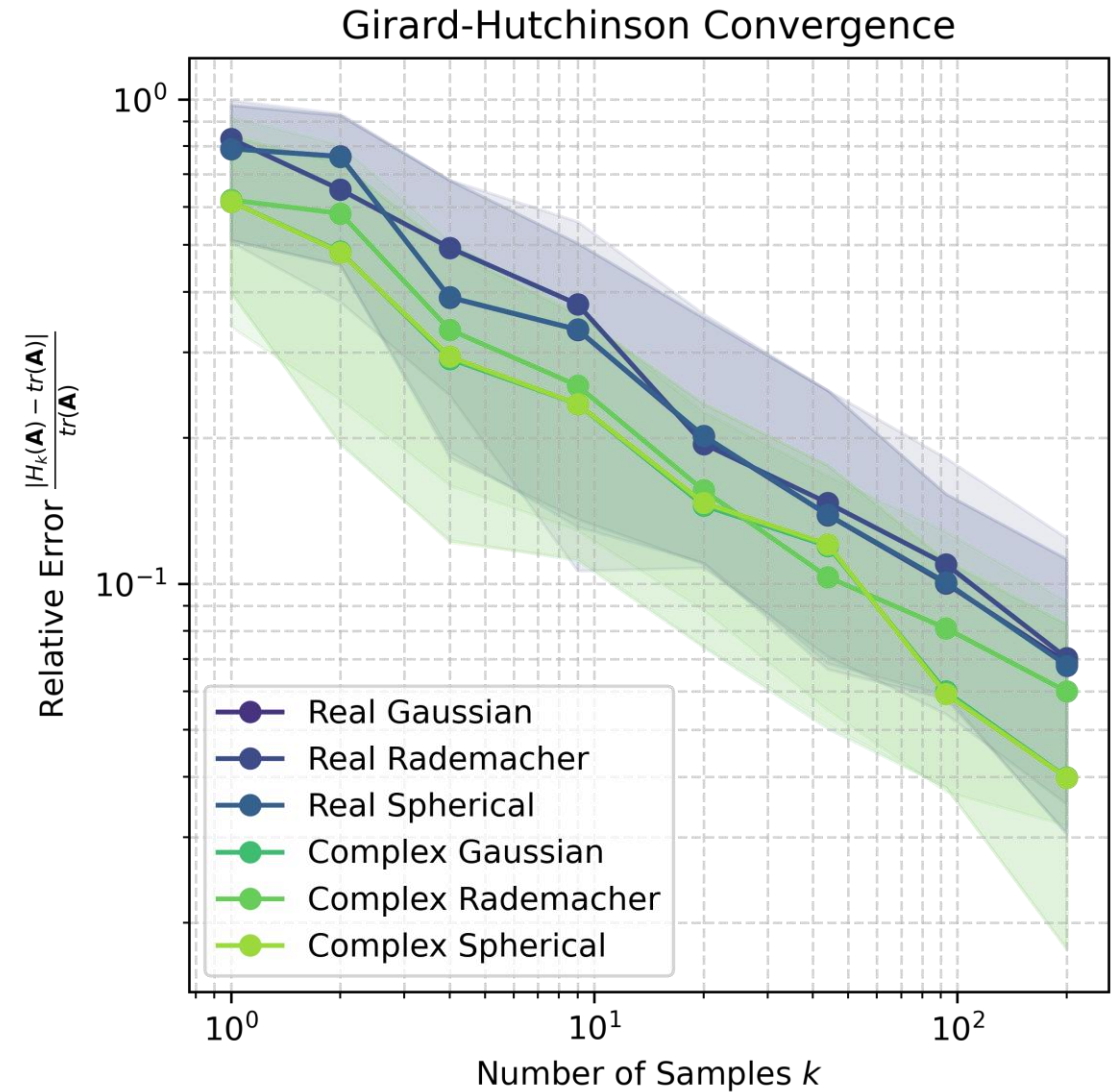
Can we estimate $\text{tr}(\mathbf{A})$ by computing $\mathbf{A}x^{(1)}, \dots, \mathbf{A}x^{(k)}$?

The Girard-Hutchinson Estimator

Sample $x \sim \mathcal{N}(0, \mathbf{I})$. Then,

$$\mathbb{E}[x^\top \mathbf{A} x] = \text{tr}(\mathbf{A}) \quad \text{Var}[x^\top \mathbf{A} x] \leq 2(\text{tr}(\mathbf{A}))^2$$

for PSD matrices $\mathbf{A}$.



The Girard-Hutchinson Estimator

Sample $x \sim \mathcal{N}(0, \mathbf{I})$. Then,

$$\mathbb{E}[x^\top \mathbf{A} x] = \text{tr}(\mathbf{A}) \quad \text{Var}[x^\top \mathbf{A} x] \leq 2(\text{tr}(\mathbf{A}))^2$$

for PSD matrices $\mathbf{A}$.

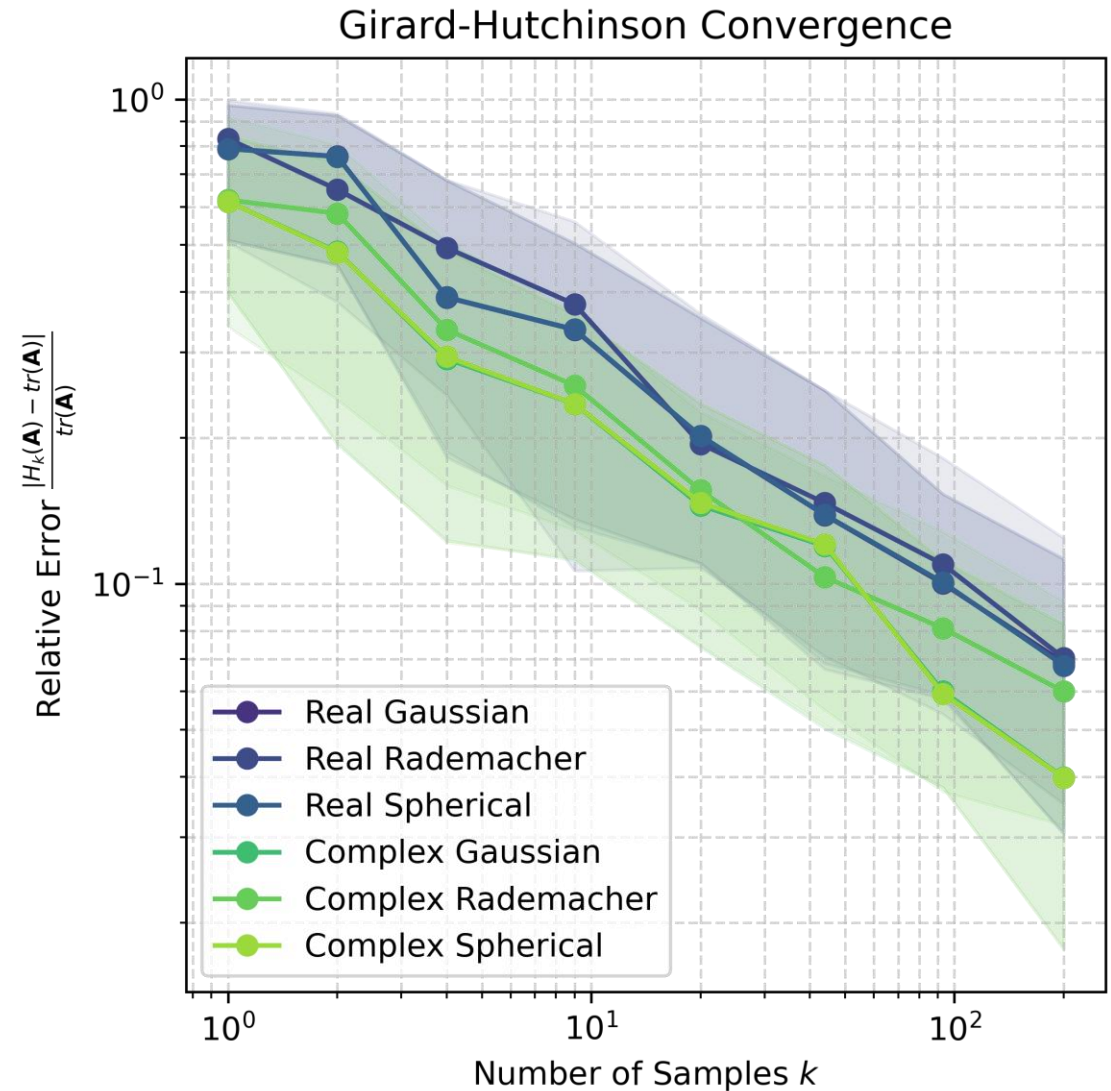
The Girard-Hutchinson Estimator

$$H_k(\mathbf{A}) = \frac{1}{k} \sum_{i=1}^k x^{(i)\top} \mathbf{A} x^{(i)}$$

with $k = O(1/\varepsilon^2)$ samples gets

$$|H_k(\mathbf{A}) - \text{tr}(\mathbf{A})| \leq \varepsilon \text{tr}(\mathbf{A})$$

with prob > 0.99 for PSD matrices $\mathbf{A}$.



But this assumes we can compute
 Ax for any vector x .

Kronecker Matrix-Vector Products

Before: $A \in \mathbb{R}^{d \times d}$. We can compute Ax for any $x \in \mathbb{R}^d$

Now: $A \in \mathbb{R}^{d^n \times d^n}$. We can compute Ax for any $x = x_1 \otimes \cdots \otimes x_n$ for any $x_i \in \mathbb{R}^d$
 $n = \text{tensor order}$

Can we still solve linear algebra problems efficiently?

$\text{poly}(n, d, 1/\varepsilon)$

Immediate concern: dn parameters vs. d^n parameters.

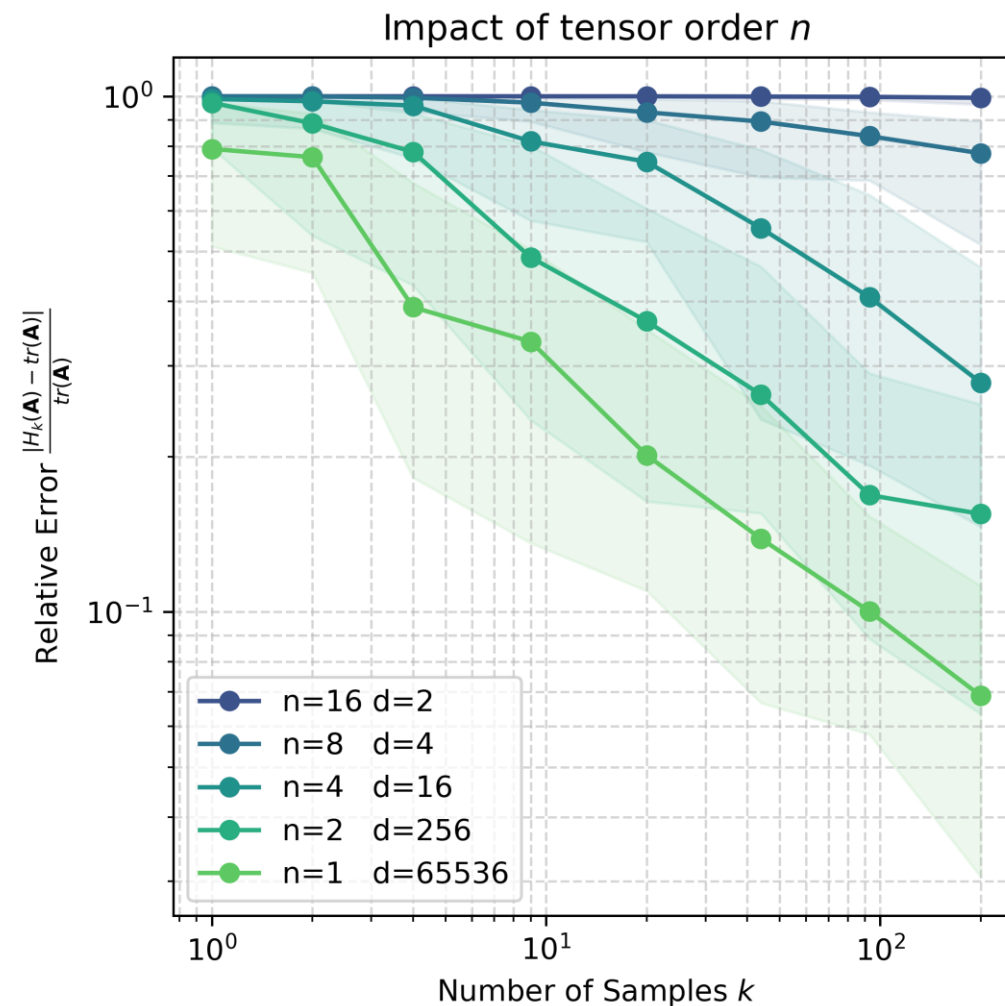
Kronecker Girard-Hutchinson

Let $x_1, \dots, x_n \sim \mathcal{N}(0, \mathbf{I})$ and $x = x_1 \otimes \dots \otimes x_n$. Then,

$$\mathbb{E}[x^\top A x] = \text{tr}(A) \quad \text{Var}[x^\top A x] \leq 3^n (\text{tr}(A))^2$$

for PSD A . We need $k = O(3^n / \varepsilon^2)$ samples!

So much worse.... why?



The Culprit

Claim: there is a PSD matrix with

$$\text{Var}[x^\top A x] = (3^n - 1)(\text{tr}(A))^2$$

$$(A \otimes B)(x \otimes y) = Ax \otimes By$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \mathbf{1}\mathbf{1}^\top$$

The Culprit

The All-Ones Matrix

Claim: there is a PSD matrix with
 $\text{Var}[x^\top A x] = (3^n - 1)(\text{tr}(A))^2$

$$(A \otimes B)(x \otimes y) = Ax \otimes By$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \mathbf{1}\mathbf{1}^\top$$

The Culprit

Claim: there is a PSD matrix with
 $\text{Var}[x^\top Ax] = (3^n - 1)(\text{tr}(A))^2$

The All-Ones Matrix

Let $A = \mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{d^n \times d^n}$ be the all-ones matrix. Then

$$(A \otimes B)(x \otimes y) = Ax \otimes By$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \mathbf{1}\mathbf{1}^\top$$

The Culprit

Claim: there is a PSD matrix with
 $\text{Var}[x^\top A x] = (3^n - 1)(\text{tr}(A))^2$

The All-Ones Matrix

Let $A = \mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{d^n \times d^n}$ be the all-ones matrix. Then
 $x^\top A x = \langle x, \mathbf{1} \rangle^2$

$$(A \otimes B)(x \otimes y) = Ax \otimes By$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \mathbf{1}\mathbf{1}^\top$$

The Culprit

Claim: there is a PSD matrix with
 $\text{Var}[x^\top A x] = (3^n - 1)(\text{tr}(A))^2$

The All-Ones Matrix

Let $A = \mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{d^n \times d^n}$ be the all-ones matrix. Then
 $x^\top A x = \langle x, \mathbf{1} \rangle^2$

Decompose the elements

$$(A \otimes B)(x \otimes y) = Ax \otimes By$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \mathbf{1}\mathbf{1}^\top$$

The Culprit

Claim: there is a PSD matrix with

$$\text{Var}[x^\top A x] = (3^n - 1)(\text{tr}(A))^2$$

The All-Ones Matrix

Let $A = \mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{d^n \times d^n}$ be the all-ones matrix. Then

$$x^\top A x = \langle x, \mathbf{1} \rangle^2$$

Decompose the elements

We can write

$$(A \otimes B)(x \otimes y) = Ax \otimes By$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \mathbf{1}\mathbf{1}^\top$$

The Culprit

Claim: there is a PSD matrix with
 $\text{Var}[x^\top Ax] = (3^n - 1)(\text{tr}(A))^2$

The All-Ones Matrix

Let $A = \mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{d^n \times d^n}$ be the all-ones matrix. Then
 $x^\top Ax = \langle x, \mathbf{1} \rangle^2$

Decompose the elements

We can write

$$\begin{aligned}x &= (x_1 \otimes \cdots \otimes x_n) \\ \mathbf{1} &= (\mathbf{1} \otimes \cdots \otimes \mathbf{1})\end{aligned}$$

So $\langle x, \mathbf{1} \rangle = \prod_{i=1}^n \langle x_i, \mathbf{1} \rangle$.

$$(A \otimes B)(x \otimes y) = Ax \otimes By$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \mathbf{1}\mathbf{1}^\top$$

The Culprit

Claim: there is a PSD matrix with
 $\text{Var}[x^\top Ax] = (3^n - 1)(\text{tr}(A))^2$

The All-Ones Matrix

Let $A = \mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{d^n \times d^n}$ be the all-ones matrix. Then
 $x^\top Ax = \langle x, \mathbf{1} \rangle^2$

Decompose the elements

We can write

$$\begin{aligned}x &= (x_1 \otimes \cdots \otimes x_n) \\ \mathbf{1} &= (\mathbf{1} \otimes \cdots \otimes \mathbf{1})\end{aligned}$$

So $\langle x, \mathbf{1} \rangle = \prod_{i=1}^n \langle x_i, \mathbf{1} \rangle$.

Product of independent variables

$$(A \otimes B)(x \otimes y) = Ax \otimes By$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \mathbf{1}\mathbf{1}^\top$$

The Culprit

Claim: there is a PSD matrix with
 $\text{Var}[x^\top A x] = (3^n - 1)(\text{tr}(A))^2$

The All-Ones Matrix

Let $A = \mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{d^n \times d^n}$ be the all-ones matrix. Then
 $x^\top A x = \langle x, \mathbf{1} \rangle^2$

Decompose the elements

We can write

$$\begin{aligned} x &= (x_1 \otimes \cdots \otimes x_n) \\ \mathbf{1} &= (\mathbf{1} \otimes \cdots \otimes \mathbf{1}) \end{aligned}$$

So $\langle x, \mathbf{1} \rangle = \prod_{i=1}^n \langle x_i, \mathbf{1} \rangle$.

Product of independent variables

$$\mathbb{E}[\langle x, \mathbf{1} \rangle^4] = \prod_{i=1}^n \mathbb{E}[\langle x_i, \mathbf{1} \rangle^4] = 3^n (\text{tr}(A))^2$$

$$(A \otimes B)(x \otimes y) = Ax \otimes By$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \mathbf{1}\mathbf{1}^\top$$

The Culprit

Claim: there is a PSD matrix with
 $\text{Var}[x^\top Ax] = (3^n - 1)(\text{tr}(A))^2$

The All-Ones Matrix

Let $A = \mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{d^n \times d^n}$ be the all-ones matrix. Then
 $x^\top Ax = \langle x, \mathbf{1} \rangle^2$

Decompose the elements

We can write

$$\begin{aligned}x &= (x_1 \otimes \cdots \otimes x_n) \\ \mathbf{1} &= (\mathbf{1} \otimes \cdots \otimes \mathbf{1})\end{aligned}$$

So $\langle x, \mathbf{1} \rangle = \prod_{i=1}^n \langle x_i, \mathbf{1} \rangle$.

Product of independent variables

$$\mathbb{E}[\langle x, \mathbf{1} \rangle^4] = \prod_{i=1}^n \mathbb{E}[\langle x_i, \mathbf{1} \rangle^4] = 3^n (\text{tr}(A))^2$$

Conclusion

$$(A \otimes B)(x \otimes y) = Ax \otimes By$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \mathbf{1}\mathbf{1}^\top$$

The Culprit

Claim: there is a PSD matrix with
 $\text{Var}[x^\top Ax] = (3^n - 1)(\text{tr}(A))^2$

The All-Ones Matrix

Let $A = \mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{d^n \times d^n}$ be the all-ones matrix. Then
 $x^\top Ax = \langle x, \mathbf{1} \rangle^2$

Decompose the elements

We can write

$$\begin{aligned}x &= (x_1 \otimes \cdots \otimes x_n) \\ \mathbf{1} &= (\mathbf{1} \otimes \cdots \otimes \mathbf{1})\end{aligned}$$

So $\langle x, \mathbf{1} \rangle = \prod_{i=1}^n \langle x_i, \mathbf{1} \rangle$.

Product of independent variables

$$\mathbb{E}[\langle x, \mathbf{1} \rangle^4] = \prod_{i=1}^n \mathbb{E}[\langle x_i, \mathbf{1} \rangle^4] = 3^n (\text{tr}(A))^2$$

Conclusion

$$\text{Var}[x^\top Ax] = (3^n - 1)(\text{tr}(A))^2$$

$$(A \otimes B)(x \otimes y) = Ax \otimes By$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \mathbf{1}\mathbf{1}^\top$$

The Culprit

Claim: there is a PSD matrix with
 $\text{Var}[x^\top A x] = (3^n - 1)(\text{tr}(A))^2$

The All-Ones Matrix

Let $A = \mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{d^n \times d^n}$ be the all-ones matrix. Then
 $x^\top A x = \langle x, \mathbf{1} \rangle^2$

Decompose the elements

We can write

$$\begin{aligned} x &= (x_1 \otimes \cdots \otimes x_n) \\ \mathbf{1} &= (\mathbf{1} \otimes \cdots \otimes \mathbf{1}) \end{aligned}$$

So $\langle x, \mathbf{1} \rangle = \prod_{i=1}^n \langle x_i, \mathbf{1} \rangle$.

Product of independent variables

$$\mathbb{E}[\langle x, \mathbf{1} \rangle^4] = \prod_{i=1}^n \mathbb{E}[\langle x_i, \mathbf{1} \rangle^4] = 3^n (\text{tr}(A))^2$$

Conclusion

$$\text{Var}[x^\top A x] = (3^n - 1)(\text{tr}(A))^2$$

But what if x_i wasn't Gaussian?

$$(A \otimes B)(x \otimes y) = Ax \otimes By$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \mathbf{1}\mathbf{1}^\top$$

Changing the Base Distribution

Let $x_i \sim \mathcal{D}_{base}$ such that:

$$\mathbb{E}[x_i x_i^\top] = \mathbf{I}, \quad \text{Var}[x_i^\top B x_i] < C(\text{tr}(\mathbf{B}))^2 \text{ for all PSD } \mathbf{B} \in \mathbb{R}^{d \times d}$$

Changing the Base Distribution

Let $x_i \sim \mathcal{D}_{base}$ such that:

$$\mathbb{E}[x_i x_i^\top] = \mathbf{I}, \quad \text{Var}[x_i^\top B x_i] < C(\text{tr}(\mathbf{B}))^2 \text{ for all PSD } \mathbf{B} \in \mathbb{R}^{d \times d}$$

Then $x_1, \dots, x_n \sim \mathcal{D}_{base}$ and $x = x_1 \otimes \dots \otimes x_n$ gives

$$\mathbb{E}[x^\top \mathbf{A} x] = \text{tr}(\mathbf{A}) \quad \text{Var}[x^\top \mathbf{A} x] \leq (C + 1)^n (\text{tr}(\mathbf{A}))^2$$

Changing the Base Distribution

Let $x_i \sim \mathcal{D}_{base}$ such that:

$$\mathbb{E}[x_i x_i^\top] = \mathbf{I}, \quad \text{Var}[x_i^\top B x_i] < C(\text{tr}(\mathbf{B}))^2 \text{ for all PSD } \mathbf{B} \in \mathbb{R}^{d \times d}$$

Then $x_1, \dots, x_n \sim \mathcal{D}_{base}$ and $x = x_1 \otimes \dots \otimes x_n$ gives

$$\mathbb{E}[x^\top A x] = \text{tr}(\mathbf{A}) \quad \text{Var}[x^\top A x] \leq (C + 1)^n (\text{tr}(\mathbf{A}))^2$$

$x_i \sim \mathcal{D}_{base}$	Complexity	($n = 2$)
$\mathcal{N}(0, \mathbf{I})$	3^n	3^n

Changing the Base Distribution

Let $x_i \sim \mathcal{D}_{base}$ such that:

$$\mathbb{E}[x_i x_i^\top] = \mathbf{I}, \quad \text{Var}[x_i^\top B x_i] < C(\text{tr}(B))^2 \text{ for all PSD } B \in \mathbb{R}^{d \times d}$$

Then $x_1, \dots, x_n \sim \mathcal{D}_{base}$ and $x = x_1 \otimes \dots \otimes x_n$ gives

$$\mathbb{E}[x^\top A x] = \text{tr}(A) \quad \text{Var}[x^\top A x] \leq (C + 1)^n (\text{tr}(A))^2$$

$x_i \sim \mathcal{D}_{base}$	Complexity	($n = 2$)
$\mathcal{N}(0, \mathbf{I})$	3^n	3^n
$\{\pm 1\}^d$	$(3 - 2/d)^n$	2^n

Changing the Base Distribution

Let $x_i \sim \mathcal{D}_{base}$ such that:

$$\mathbb{E}[x_i x_i^\top] = \mathbf{I}, \quad \text{Var}[x_i^\top B x_i] < C(\text{tr}(\mathbf{B}))^2 \text{ for all PSD } \mathbf{B} \in \mathbb{R}^{d \times d}$$

Then $x_1, \dots, x_n \sim \mathcal{D}_{base}$ and $x = x_1 \otimes \dots \otimes x_n$ gives

$$\mathbb{E}[x^\top A x] = \text{tr}(\mathbf{A}) \quad \text{Var}[x^\top A x] \leq (C + 1)^n (\text{tr}(\mathbf{A}))^2$$

$x_i \sim \mathcal{D}_{base}$	Complexity	($n = 2$)
$\mathcal{N}(0, \mathbf{I})$	3^n	3^n
$\{\pm 1\}^d$	$(3 - 2/d)^n$	2^n
$\text{Unif} \{ \ x_i\ = \sqrt{d} \}$	$(3 - \frac{6}{d+2})^n$	1.5^n

Changing the Base Distribution

Let $x_i \sim \mathcal{D}_{base}$ such that:

$$\mathbb{E}[x_i x_i^\top] = \mathbf{I}, \quad \text{Var}[x_i^\top B x_i] < C(\text{tr}(\mathbf{B}))^2 \text{ for all PSD } \mathbf{B} \in \mathbb{R}^{d \times d}$$

Then $x_1, \dots, x_n \sim \mathcal{D}_{base}$ and $x = x_1 \otimes \dots \otimes x_n$ gives

$$\mathbb{E}[x^\top A x] = \text{tr}(\mathbf{A}) \quad \text{Var}[x^\top A x] \leq (C + 1)^n (\text{tr}(\mathbf{A}))^2$$

$x_i \sim \mathcal{D}_{base}$	Complexity	($n = 2$)
$\mathcal{N}(0, \mathbf{I})$	3^n	3^n
$\{\pm 1\}^d$	$(3 - 2/d)^n$	2^n
$\text{Unif} \{ \ x_i\ = \sqrt{d} \}$	$(3 - \frac{6}{d+2})^n$	1.5^n
$\mathcal{N}_{\mathbb{C}}(0, \mathbf{I})$	2^n	2^n

Changing the Base Distribution

Let $x_i \sim \mathcal{D}_{base}$ such that:

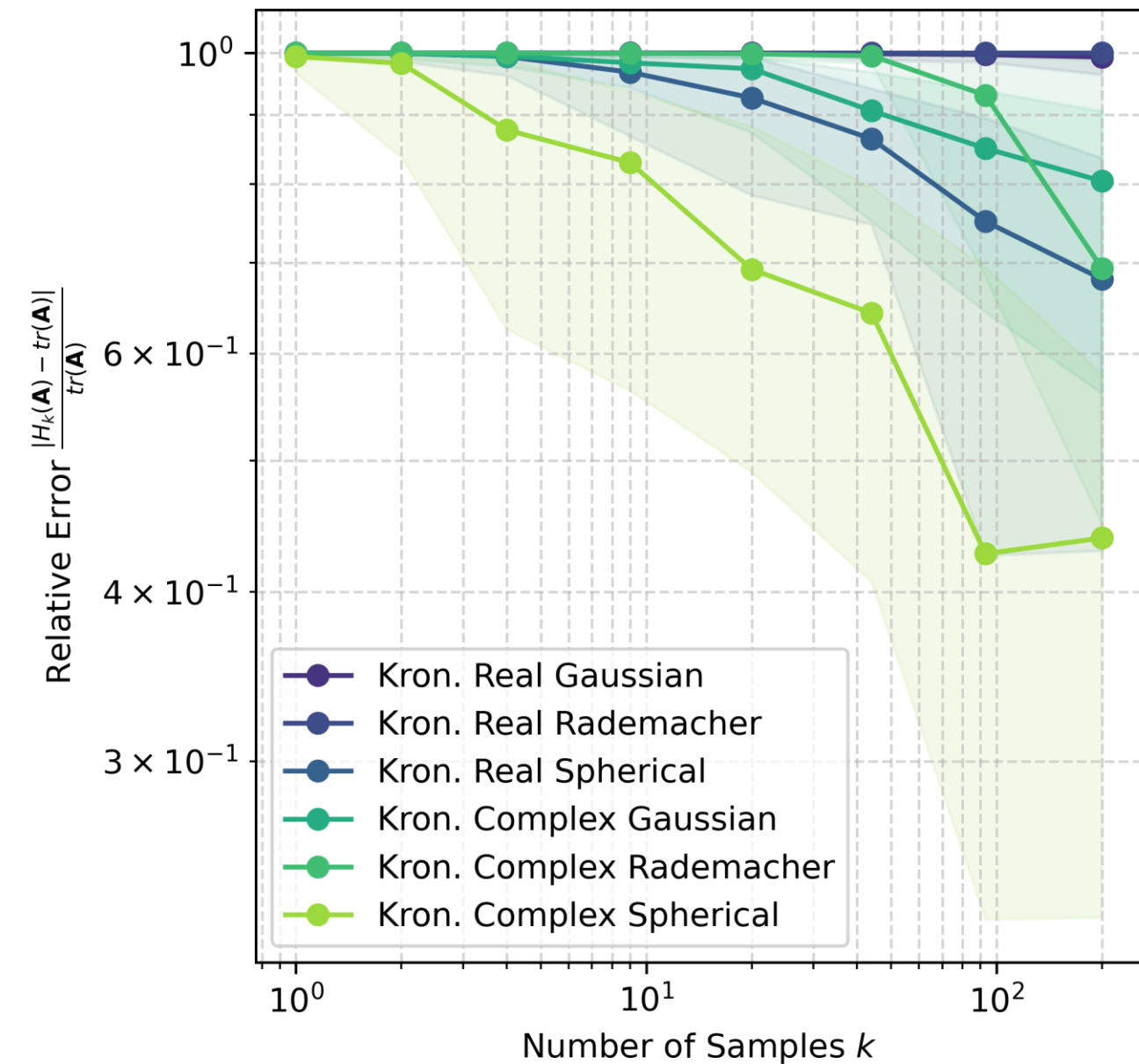
$$\mathbb{E}[x_i x_i^\top] = \mathbf{I}, \quad \text{Var}[x_i^\top B x_i] < C(\text{tr}(B))^2 \text{ for all PSD } B \in \mathbb{R}^{d \times d}$$

Then $x_1, \dots, x_n \sim \mathcal{D}_{base}$ and $x = x_1 \otimes \dots \otimes x_n$ gives

$$\mathbb{E}[x^\top A x] = \text{tr}(A) \quad \text{Var}[x^\top A x] \leq (C + 1)^n (\text{tr}(A))^2$$

$x_i \sim \mathcal{D}_{base}$	Complexity	($n = 2$)
$\mathcal{N}(0, \mathbf{I})$	3^n	3^n
$\{\pm 1\}^d$	$(3 - 2/d)^n$	2^n
$\text{Unif} \left\{ \ x_i\ = \sqrt{d} \right\}$	$\left(3 - \frac{6}{d+2}\right)^n$	1.5^n
$\mathcal{N}_{\mathbb{C}}(0, \mathbf{I})$	2^n	2^n
$\{\pm 1, \pm i\}^d$	$(2 - d)^n$	1.5^n
$\text{Unif} \left\{ \ x_i\ = \sqrt{d} \right\}_{\mathbb{C}}$	$\left(2 - \frac{2}{d+1}\right)^n$	$1.3\overline{3}^n$

Kronecker: Base Distribution Matters ($d^n = 2^{16}$)



This matters.

A real matrix $A = e^{-\frac{1}{2}B}$ from quantum physics. $d^n = 2^{16}$.
 Matrix-vector products $x \rightarrow Ax$ computed using
`scipy.sparse.linalg.expm_multiply`
 Median error of 100 trials, with 25th – 75th quantiles

We just went from $3^n/\varepsilon^2$ to $1.33^n/\varepsilon^2$...

...why not $\text{poly}(n/\varepsilon)$?

Act 2: The Curse of Overwhelming Orthogonality

Things are weird

Things that change our complexity:

1. Exponential Dependence on n
2. Swapping out different subgaussian variables
3. Complex numbers

Things are weird

Things that change our complexity:

1. Exponential Dependence on n
2. Swapping out different subgaussian variables
3. Complex numbers

But... what if our algorithm is bad/dumb?

Can we find natural linear algebra **problems** such that the same effects must occur?

Things are weird

Things that change our complexity:

1. Exponential Dependence on n
2. Swapping out different subgaussian variables
3. Complex numbers

But... what if our algorithm is bad/dumb?

Can we find natural linear algebra **problems** such that the same effects must occur?

Yes!

Zero Testing

Determine if $A = 0$ with as few queries $Ax^{(1)}, \dots, Ax^{(k)}$ as possible.

A single Gaussian query suffices...

Zero Testing

Determine if $A = 0$ with as few queries $Ax^{(1)}, \dots, Ax^{(k)}$ as possible.

A single Gaussian query suffices...

Any algorithm which uses vectors

$x_i \in \{\pm 1\}^d$
must use 2^n queries to succeed
with prob $2/3$!

Zero Testing

A random basis vector

Determine if $A = 0$ with as few queries $Ax^{(1)}, \dots, Ax^{(k)}$ as possible.

A single Gaussian query suffices...

Any algorithm which uses vectors

$x_i \in \{\pm 1\}^d$
must use 2^n queries to succeed
with prob $2/3$!

Zero Testing

Determine if $A = 0$ with as few queries $Ax^{(1)}, \dots, Ax^{(k)}$ as possible.

A single Gaussian query suffices...

Any algorithm which uses vectors

$x_i \in \{\pm 1\}^d$
must use 2^n queries to succeed
with prob $2/3$!

A random basis vector

Let $A = uu^\top \in \mathbb{R}^{d^n \times d^n}$, where:

Zero Testing

Determine if $A = 0$ with as few queries $Ax^{(1)}, \dots, Ax^{(k)}$ as possible.

A single Gaussian query suffices...

Any algorithm which uses vectors

$x_i \in \{\pm 1\}^d$
must use 2^n queries to succeed
with prob $2/3$!

A random basis vector

Let $A = uu^\top \in \mathbb{R}^{d^n \times d^n}$, where:

$$u = (u_1 \otimes \dots \otimes u_n)$$

$$u_i = [\pm 1 \quad \pm 1 \quad 0 \quad \dots \quad 0] \text{ iid}$$

Zero Testing

Determine if $A = 0$ with as few queries $Ax^{(1)}, \dots, Ax^{(k)}$ as possible.

A single Gaussian query suffices...

Any algorithm which uses vectors

$x_i \in \{\pm 1\}^d$
must use 2^n queries to succeed
with prob $2/3$!

A random basis vector

Let $A = uu^\top \in \mathbb{R}^{d^n \times d^n}$, where:

$$u = (u_1 \otimes \dots \otimes u_n)$$

$$u_i = [\pm 1 \quad \pm 1 \quad 0 \quad \dots \quad 0] \text{ iid}$$

Local Orthogonality

Zero Testing

Determine if $A = 0$ with as few queries $Ax^{(1)}, \dots, Ax^{(k)}$ as possible.

A single Gaussian query suffices...

Any algorithm which uses vectors

$x_i \in \{\pm 1\}^d$
must use 2^n queries to succeed
with prob $2/3$!

A random basis vector

Let $A = uu^\top \in \mathbb{R}^{d^n \times d^n}$, where:

$$u = (u_1 \otimes \dots \otimes u_n)$$

$$u_i = [\pm 1 \quad \pm 1 \quad 0 \quad \dots \quad 0] \text{ iid}$$

Local Orthogonality

For any $x_i \in \{\pm 1\}^d$,

Zero Testing

Determine if $A = 0$ with as few queries $Ax^{(1)}, \dots, Ax^{(k)}$ as possible.

A single Gaussian query suffices...

Any algorithm which uses vectors $x_i \in \{\pm 1\}^d$ must use 2^n queries to succeed with prob $2/3$!

A random basis vector

Let $A = uu^\top \in \mathbb{R}^{d^n \times d^n}$, where:

$$u = (u_1 \otimes \dots \otimes u_n)$$

$$u_i = [\pm 1 \quad \pm 1 \quad 0 \quad \dots \quad 0] \text{ iid}$$

Local Orthogonality

For any $x_i \in \{\pm 1\}^d$,

$$\langle x_i, u_i \rangle \neq 0 \text{ with prob } \frac{1}{2}$$

Zero Testing

Determine if $A = 0$ with as few queries $Ax^{(1)}, \dots, Ax^{(k)}$ as possible.

A single Gaussian query suffices...

Any algorithm which uses vectors

$x_i \in \{\pm 1\}^d$
must use 2^n queries to succeed
with prob $2/3$!

A random basis vector

Let $A = uu^\top \in \mathbb{R}^{d^n \times d^n}$, where:

$$u = (u_1 \otimes \dots \otimes u_n)$$

$$u_i = [\pm 1 \quad \pm 1 \quad 0 \quad \dots \quad 0] \text{ iid}$$

Local Orthogonality

For any $x_i \in \{\pm 1\}^d$,

$$\langle x_i, u_i \rangle \neq 0 \text{ with prob } \frac{1}{2}$$

Global orthogonality

For any $x = x_1 \otimes \dots \otimes x_n$,

Zero Testing

Determine if $A = 0$ with as few queries $Ax^{(1)}, \dots, Ax^{(k)}$ as possible.

A single Gaussian query suffices...

Any algorithm which uses vectors $x_i \in \{\pm 1\}^d$ must use 2^n queries to succeed with prob $2/3$!

A random basis vector

Let $A = uu^\top \in \mathbb{R}^{d^n \times d^n}$, where:

$$u = (u_1 \otimes \dots \otimes u_n)$$

$$u_i = [\pm 1 \quad \pm 1 \quad 0 \quad \dots \quad 0] \text{ iid}$$

Local Orthogonality

For any $x_i \in \{\pm 1\}^d$,

$$\langle x_i, u_i \rangle \neq 0 \text{ with prob } \frac{1}{2}$$

Global orthogonality

For any $x = x_1 \otimes \dots \otimes x_n$,

$$\langle x, u \rangle = \prod_{i=1}^n \langle x_i, u_i \rangle \neq 0 \text{ with prob } 2^{-n}$$

Zero Testing

Determine if $A = 0$ with as few queries $Ax^{(1)}, \dots, Ax^{(k)}$ as possible.

A single Gaussian query suffices...

Any algorithm which uses vectors $x_i \in \{\pm 1\}^d$ must use 2^n queries to succeed with prob $2/3$!

A random basis vector

Let $A = uu^\top \in \mathbb{R}^{d^n \times d^n}$, where:

$$u = (u_1 \otimes \dots \otimes u_n)$$

$$u_i = [\pm 1 \quad \pm 1 \quad 0 \quad \dots \quad 0] \text{ iid}$$

Local Orthogonality

For any $x_i \in \{\pm 1\}^d$,

$$\langle x_i, u_i \rangle \neq 0 \text{ with prob } \frac{1}{2}$$

Global orthogonality

For any $x = x_1 \otimes \dots \otimes x_n$,

$$\langle x, u \rangle = \prod_{i=1}^n \langle x_i, u_i \rangle \neq 0 \text{ with prob } 2^{-n}$$

Conclusion

Zero Testing

Determine if $A = 0$ with as few queries $Ax^{(1)}, \dots, Ax^{(k)}$ as possible.

A single Gaussian query suffices...

Any algorithm which uses vectors $x_i \in \{\pm 1\}^d$ must use 2^n queries to succeed with prob $2/3$!

A random basis vector

Let $A = uu^\top \in \mathbb{R}^{d^n \times d^n}$, where:

$$u = (u_1 \otimes \dots \otimes u_n)$$

$$u_i = [\pm 1 \quad \pm 1 \quad 0 \quad \dots \quad 0] \text{ iid}$$

Local Orthogonality

For any $x_i \in \{\pm 1\}^d$,

$$\langle x_i, u_i \rangle \neq 0 \text{ with prob } \frac{1}{2}$$

Global orthogonality

For any $x = x_1 \otimes \dots \otimes x_n$,

$$\langle x, u \rangle = \prod_{i=1}^n \langle x_i, u_i \rangle \neq 0 \text{ with prob } 2^{-n}$$

Conclusion

$$Ax = uu^\top x \neq 0 \text{ with prob } 2^{-n}$$

Zero Testing

Determine if $A = 0$ with as few queries $Ax^{(1)}, \dots, Ax^{(k)}$ as possible.

A single Gaussian query suffices...

Any algorithm which uses vectors $x_i \in \{\pm 1\}^d$ must use 2^n queries to succeed with prob $2/3$!

A random basis vector

Let $A = uu^\top \in \mathbb{R}^{d^n \times d^n}$, where:

$$u = (u_1 \otimes \dots \otimes u_n)$$

$$u_i = [\pm 1 \quad \pm 1 \quad 0 \quad \dots \quad 0] \text{ iid}$$

Local Orthogonality

For any $x_i \in \{\pm 1\}^d$,

$$\langle x_i, u_i \rangle \neq 0 \text{ with prob } \frac{1}{2}$$

Global orthogonality

For any $x = x_1 \otimes \dots \otimes x_n$,

$$\langle x, u \rangle = \prod_{i=1}^n \langle x_i, u_i \rangle \neq 0 \text{ with prob } 2^{-n}$$

Conclusion

$$Ax = uu^\top x \neq 0 \text{ with prob } 2^{-n}$$

2^n queries are needed to prove $A \neq 0$

Overwhelming Orthogonality

Kronecker-structured vectors are, almost always, very orthogonal!

Overwhelming Orthogonality

Kronecker-structured vectors are, almost always, very orthogonal!

There exists a random vector $u \in \mathbb{R}^{d^n}$ such that $\mathbb{E}[uu^\top] = \mathbf{I}$, and

For all unit vectors $x = x_1 \otimes \cdots \otimes x_n$,

$$\Pr[\langle x, u \rangle^2 > \tau^{-n}] < \tau^{-n}$$

for some constant $\tau \in (0,1)$

Overwhelming Orthogonality

Kronecker-structured vectors are, almost always, very orthogonal!

There exists a random vector $u \in \mathbb{R}^{d^n}$ such that $\mathbb{E}[uu^\top] = \mathbf{I}$, and

For all unit vectors $x = x_1 \otimes \cdots \otimes x_n$,

$$\Pr[\langle x, u \rangle^2 > \tau^{-n}] < \tau^{-n}$$

for some constant $\tau \in (0,1)$

Reference: if $u \sim \mathcal{N}(0_{d^n}, \mathbf{I}_{d^n})$ then $\langle x, u \rangle^2 \approx 1$.

Unconditional* Lower Bounds

Bury a huge Kronecker-vector under a small Gaussian

Unconditional* Lower Bounds

Bury a huge Kronecker-vector under a small Gaussian

These random variables are exponentially hard to distinguish:

$$r_1 = \langle x, g \rangle \quad r_2 = \langle x, g + \lambda u \rangle$$

Where x is a kron vector, $g \sim \mathcal{N}(0_{d^n}, \mathbf{I}_{d^n})$, u from last slide, even if $\lambda = \tau^{n/2}$.

Unconditional* Lower Bounds

Bury a huge Kronecker-vector under a small Gaussian

These random variables are exponentially hard to distinguish:

$$r_1 = \langle x, g \rangle \quad r_2 = \langle x, g + \lambda u \rangle$$

Where x is a kron vector, $g \sim \mathcal{N}(0_{d^n}, \mathbf{I}_{d^n})$, u from last slide, even if $\lambda = \tau^{n/2}$.

These matrices cannot be distinguished but very different trace

$$\begin{aligned} A_1 &= gg^\top \\ A_2 &= gg^\top + \tau^{n/2}uu^\top \end{aligned}$$

Unconditional* Lower Bounds

Bury a huge Kronecker-vector under a small Gaussian

These random variables are exponentially hard to distinguish:

$$r_1 = \langle x, g \rangle \quad r_2 = \langle x, g + \lambda u \rangle$$

Where x is a kron vector, $g \sim \mathcal{N}(0_{d^n}, \mathbf{I}_{d^n})$, u from last slide, even if $\lambda = \tau^{n/2}$.

Theorem 1

No algorithm can determine if $\text{tr}(\mathbf{A}) \leq 1$ or $\text{tr}(\mathbf{A}) \geq \exp(n)$ with fewer than $\exp(n)$ bilinear queries $x \rightarrow x^\top \mathbf{A} x$

These matrices cannot be distinguished but very different trace

$$\begin{aligned} \mathbf{A}_1 &= g g^\top \\ \mathbf{A}_2 &= g g^\top + \tau^{n/2} u u^\top \end{aligned}$$

Unconditional* Lower Bounds

Bury a huge Kronecker-vector under a small Gaussian

These random variables are exponentially hard to distinguish:

$$r_1 = \langle x, g \rangle \quad r_2 = \langle x, g + \lambda u \rangle$$

Where x is a kron vector, $g \sim \mathcal{N}(0_{d^n}, \mathbf{I}_{d^n})$, u from last slide, even if $\lambda = \tau^{n/2}$.

Theorem 1

No algorithm can determine if $\text{tr}(\mathbf{A}) \leq 1$ or $\text{tr}(\mathbf{A}) \geq \exp(n)$ with fewer than $\exp(n)$ bilinear queries $x \rightarrow x^\top \mathbf{A} x$

These matrices cannot be distinguished but very different trace

$$\begin{aligned} \mathbf{A}_1 &= g g^\top \\ \mathbf{A}_2 &= g g^\top + \tau^{n/2} u u^\top \end{aligned}$$

These matrices cannot be distinguished but very different $\lambda_{\max}$

$$\begin{aligned} \mathbf{A}_1 &= G \\ \mathbf{A}_2 &= G + \tau^{n/2} u u^\top \end{aligned}$$

Unconditional* Lower Bounds

Bury a huge Kronecker-vector under a small Gaussian

These random variables are exponentially hard to distinguish:

$$r_1 = \langle x, g \rangle \quad r_2 = \langle x, g + \lambda u \rangle$$

Where x is a kron vector, $g \sim \mathcal{N}(0_{d^n}, \mathbf{I}_{d^n})$, u from last slide, even if $\lambda = \tau^{n/2}$.

Theorem 1

No algorithm can determine if $\text{tr}(\mathbf{A}) \leq 1$ or $\text{tr}(\mathbf{A}) \geq \exp(n)$ with fewer than $\exp(n)$ bilinear queries $x \rightarrow x^\top \mathbf{A} x$

Theorem 2

No algorithm can determine if $\lambda_{\max}(\mathbf{A}) \leq 1$ or $\lambda_{\max}(\mathbf{A}) \geq \exp(n)$ with fewer than $\exp(n)$ matrix-vector queries $x \rightarrow \mathbf{A} x$

These matrices cannot be distinguished but very different trace

$$\begin{aligned} \mathbf{A}_1 &= g g^\top \\ \mathbf{A}_2 &= g g^\top + \tau^{n/2} u u^\top \end{aligned}$$

These matrices cannot be distinguished but very different $\lambda_{\max}$

$$\begin{aligned} \mathbf{A}_1 &= G \\ \mathbf{A}_2 &= G + \tau^{n/2} u u^\top \end{aligned}$$

Unconditional* Lower Bounds

Bury a huge Kronecker-vector under a small Gaussian

These random variables are exponentially hard to distinguish:

$$r_1 = \langle x, g \rangle \quad r_2 = \langle x, g + \lambda u \rangle$$

Where x is a kron vector, $g \sim \mathcal{N}(0_{d^n}, \mathbf{I}_{d^n})$, u from last slide, even if $\lambda = \tau^{n/2}$.

Theorem 1

No algorithm can determine if $\text{tr}(\mathbf{A}) \leq 1$ or $\text{tr}(\mathbf{A}) \geq \exp(n)$ with fewer than $\exp(n)$ bilinear queries $x \rightarrow x^\top \mathbf{A} x$

Theorem 2

No algorithm can determine if $\lambda_{\max}(\mathbf{A}) \leq 1$ or $\lambda_{\max}(\mathbf{A}) \geq \exp(n)$ with fewer than $\exp(n)$ matrix-vector queries $x \rightarrow \mathbf{A} x$

*Caveat

Requires mild assumption on the “conditioning” of the queries.

These matrices cannot be distinguished but very different trace

$$\begin{aligned} \mathbf{A}_1 &= g g^\top \\ \mathbf{A}_2 &= g g^\top + \tau^{n/2} u u^\top \end{aligned}$$

These matrices cannot be distinguished but very different $\lambda_{\max}$

$$\begin{aligned} \mathbf{A}_1 &= G \\ \mathbf{A}_2 &= G + \tau^{n/2} u u^\top \end{aligned}$$

Things are bad

Things that change our complexity:

1. Exponential Dependence on n
2. Swapping out different subgaussian variables
3. Complex numbers help?

Things are bad

Things that change our complexity:

1. Exponential Dependence on n

This is unavoidable*

2. Swapping out different subgaussian variables

3. Complex numbers help?

Things are bad

Things that change our complexity:

1. Exponential Dependence on n

This is unavoidable*

2. Swapping out different subgaussian variables

Do not analyze an algorithm using only subgaussianity.

3. Complex numbers help?

Things are bad

Things that change our complexity:

1. Exponential Dependence on n

This is unavoidable*

2. Swapping out different subgaussian variables

Do not analyze an algorithm using only subgaussianity.

Don't use Rademacher vectors, specifically.

3. Complex numbers help?

Things are bad

Things that change our complexity:

1. Exponential Dependence on n

This is unavoidable*

2. Swapping out different subgaussian variables

Do not analyze an algorithm using only subgaussianity.

Don't use Rademacher vectors, specifically.

3. Complex numbers help?

Actually, not really! You can just use reals!

Act 3: Algorithms beyond Trace Estimation

Beyond Trace Estimation

Low-Rank Approximation

Given r, ε , find matrix $\tilde{A}$ such that $\text{rank}(\tilde{A}) = O(r)$ and

$$\|A - \tilde{A}\| \leq (1 + \varepsilon) \min_{\text{rank}(\mathbf{B}) \leq r} \|A - \mathbf{B}\|$$



Beyond Trace Estimation



Low-Rank Approximation

Given r, ε , find matrix $\tilde{A}$ such that $\text{rank}(\tilde{A}) = O(r)$ and

$$\|A - \tilde{A}\| \leq (1 + \varepsilon) \min_{\text{rank}(\mathbf{B}) \leq r} \|A - \mathbf{B}\|$$

Nyström Approximation

Let $\Omega \in \mathbb{R}^{d^n \times k}$ with iid columns

$$\Omega = \frac{1}{\sqrt{k}} \begin{bmatrix} | & & | \\ x^{(1)} & \dots & x^{(k)} \\ | & & | \end{bmatrix}$$

Where $x^{(i)}$ are iid copies of $x = x_1 \otimes \dots \otimes x_n$ for $x_i \sim \mathcal{D}_{base}$.

Beyond Trace Estimation



Low-Rank Approximation

Given r, ε , find matrix $\tilde{A}$ such that $\text{rank}(\tilde{A}) = O(r)$ and

$$\|A - \tilde{A}\| \leq (1 + \varepsilon) \min_{\text{rank}(\mathbf{B}) \leq r} \|A - \mathbf{B}\|$$

Nyström Approximation

Let $\Omega \in \mathbb{R}^{d^n \times k}$ with iid columns

$$\Omega = \frac{1}{\sqrt{k}} \begin{bmatrix} | & & | \\ x^{(1)} & \dots & x^{(k)} \\ | & & | \end{bmatrix}$$

Where $x^{(i)}$ are iid copies of $x = x_1 \otimes \dots \otimes x_n$ for $x_i \sim \mathcal{D}_{base}$.

Compute $Y = A\Omega$, and return

$$\tilde{A} = Y(\Omega^\top Y)^{-1} Y^\top$$

How many Kronecker matrix-vector products k are needed?



Two Regimes, Two Theorems



Two Regimes, Two Theorems

Many queries, Small error

Suppose that the base vectors $x_i \sim \mathcal{D}_{base}$ satisfy

$$\mathbb{E}[x_i x_i^\top] = \mathbf{I}, \quad \mathbb{E}[\langle x_i, u \rangle^4] \leq C$$

for all unit vectors $u \in \mathbb{R}^d$.



Two Regimes, Two Theorems

Many queries, Small error

Suppose that the base vectors $x_i \sim \mathcal{D}_{base}$ satisfy

$$\mathbb{E}[x_i x_i^\top] = \mathbf{I}, \quad \mathbb{E}[\langle x_i, u \rangle^4] \leq C$$

for all unit vectors $u \in \mathbb{R}^d$.

Then, $k = O(C^n r)$ sketching dimension suffices for

$$\|A - \tilde{A}\|_* \leq 3 \min_{\text{rank}(\mathbf{B}) \leq r} \|A - \mathbf{B}\|_*$$



Two Regimes, Two Theorems

Many queries, Small error

Suppose that the base vectors $x_i \sim \mathcal{D}_{base}$ satisfy

$$\mathbb{E}[x_i x_i^\top] = \mathbf{I}, \quad \mathbb{E}[\langle x_i, u \rangle^4] \leq C$$

for all unit vectors $u \in \mathbb{R}^d$.

Then, $k = O(C^n r)$ sketching dimension suffices for

$$\|A - \tilde{A}\|_* \leq 3 \min_{\text{rank}(\mathbf{B}) \leq r} \|A - \mathbf{B}\|_*$$

Few queries, Large error

Suppose that the base vectors $x_i \sim \mathcal{D}_{base}$ have iid entries with continuous bounded pdf, and that $\mathbb{E}[x_i x_i^\top] = \mathbf{I}$.



Two Regimes, Two Theorems

Many queries, Small error

Suppose that the base vectors $x_i \sim \mathcal{D}_{base}$ satisfy

$$\mathbb{E}[x_i x_i^\top] = \mathbf{I}, \quad \mathbb{E}[\langle x_i, u \rangle^4] \leq C$$

for all unit vectors $u \in \mathbb{R}^d$.

Then, $k = O(C^n r)$ sketching dimension suffices for

$$\|A - \tilde{A}\|_* \leq 3 \min_{\text{rank}(\mathbf{B}) \leq r} \|A - \mathbf{B}\|_*$$

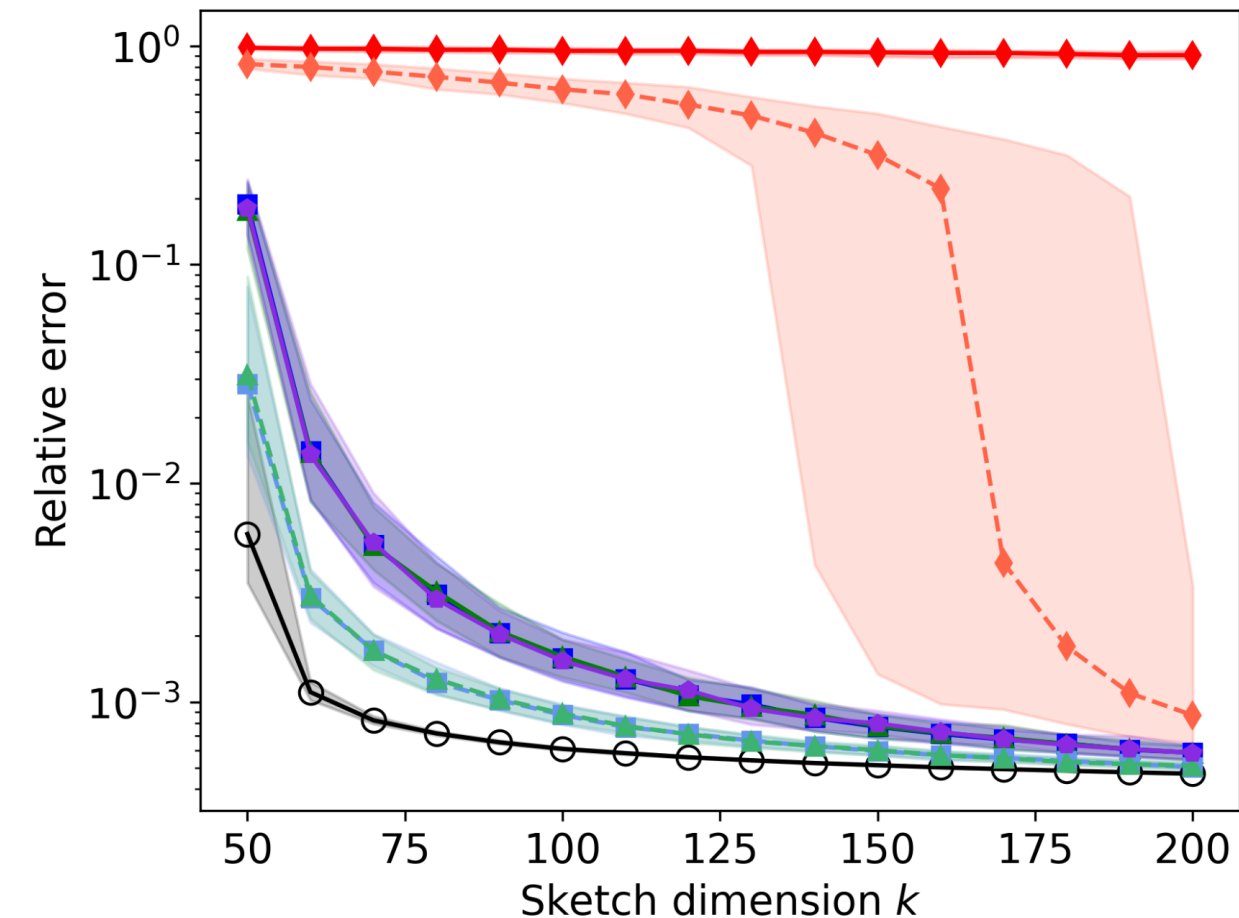
Few queries, Large error

Suppose that the base vectors $x_i \sim \mathcal{D}_{base}$ have iid entries with continuous bounded pdf, and that $\mathbb{E}[x_i x_i^\top] = \mathbf{I}$.

Then, $k = 3r$ sketching dimension suffices for

$$\|A - \tilde{A}\|_* \leq C^n \min_{\text{rank}(\mathbf{B}) \leq r} \|A - \mathbf{B}\|_*$$

RSVD error



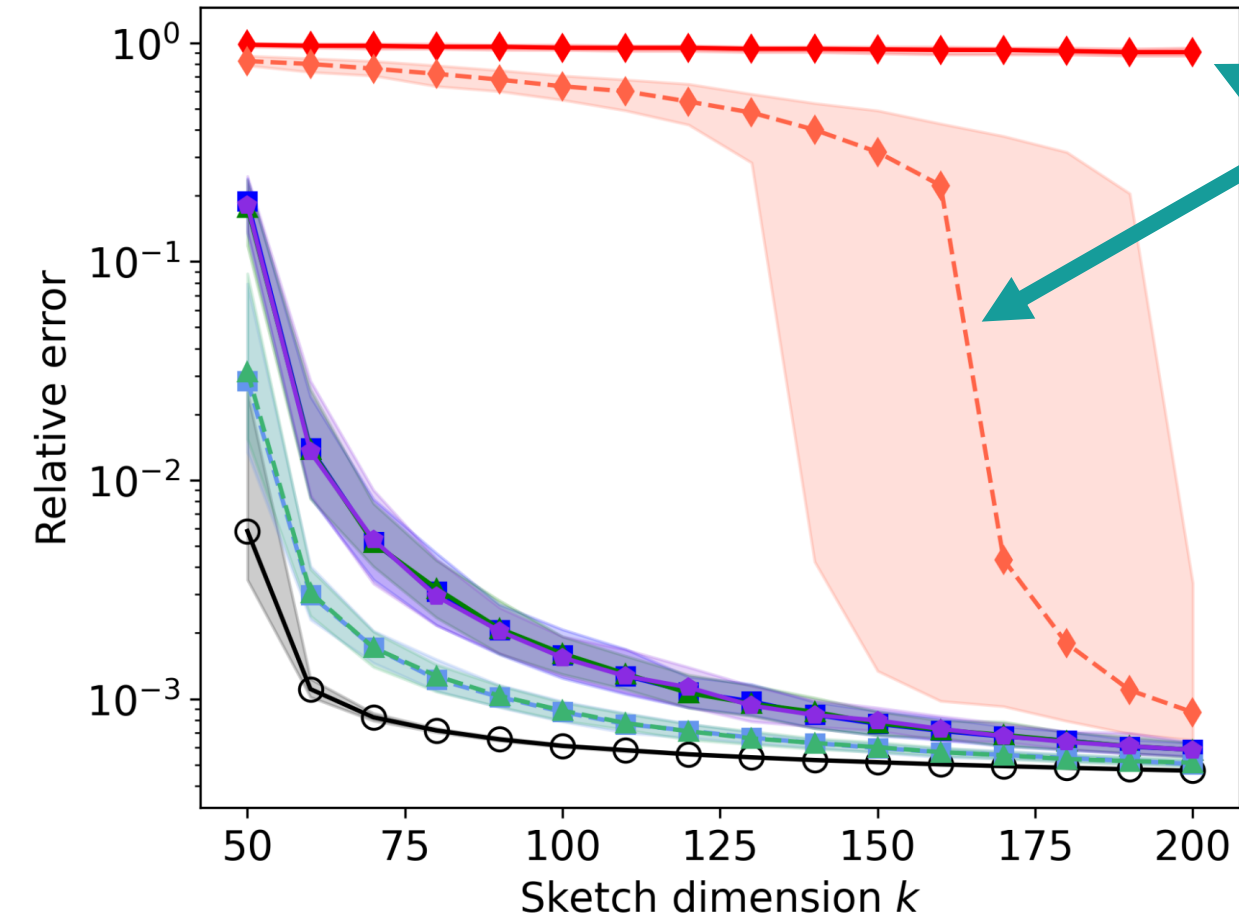
- Gaussian
- KR Gaussian
- ◆ KR Steinhaus
- ◆ KR Complex Gaussian
- ▲ KR Sphere
- ▲ KR Complex Sphere
- ◆ KR Rademacher
- ◆ KR Complex Rademacher

This works!

Actually, it works better than the theory predicts....

*Plot not actually generated using the Nyström approx.
Adversarial input matrix A , tensor order $n = 10$
Y-axis is relative reconstruction error $\|A - \tilde{A}\|_* / \|A\|_*$*

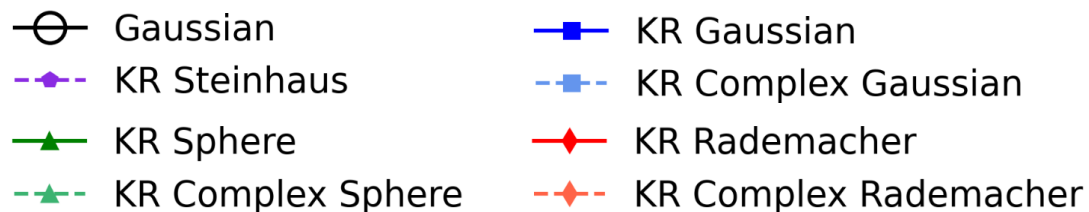
RSVD error



**Don't use
Rademacher!**

This works!

Actually, it works better than the theory predicts....



*Plot not actually generated using the Nyström approx.
Adversarial input matrix A , tensor order $n = 10$
Y-axis is relative reconstruction error $\|A - \tilde{A}\|_* / \|A\|_*$*

Main open problem: Resolve the caveat for the information-theoretic lower bound.

Let $u = u_1 \otimes \cdots \otimes u_n$ where $u_i \sim \text{Unif}\{\|z\|_2 = \sqrt{d}\}$.

Let $X \in \mathbb{R}^{d^n \times k}$ have arbitrary Kronecker-structured columns.

Let $P \in \mathbb{R}^{d^n \times d^n}$ project onto the column space of X .

Then, for some $\tau \in (0,1)$,

$$\|Pu\|_2^2 \leq k \cdot \tau^{-n} \text{ with prob at least } 1 - k \cdot \tau^{-n}$$

Q & A

Thanks for listening! Any questions?