Understanding the Kronecker Matrix-Vector Complexity of Linear Algebra

Raphael A. Meyer, William Swartworth, David P. Woodruff

Setup

Motivation Part 1: Tensors

Vector	$\vec{a} \in \mathbb{R}^d$	$[\vec{a}]_i \in \mathbb{R}$
Matrix	$A \in \mathbb{R}^{d \times d}$	$[A]_{ij} \in \mathbb{R}$
Tensor	$\mathcal{A} \in (\mathbb{R}^d)^{\otimes n}$	$[\mathcal{A}]_{(i_1,\ldots,i_n)} \in \mathbb{R}$

We will think about linear maps $\mathcal{A}: (\mathbb{R}^d)^{\otimes n} \to (\mathbb{R}^d)^{\otimes n}$

Goal: compute trace, spectral norm, etc of the linear map

$\mathcal{A}$ is linear so we can represent it as a matrix $A \in \mathbb{R}^{d^n \times d^n}$

Formally, $[A]_{(i_1,\ldots,i_n),(j_1,\ldots,j_n)} = \left[\mathcal{A}\vec{e}_{(i_1,\ldots,i_n)}\right]_{(j_1,\ldots,j_n)}$

Motivation Part 2: Matricization

The matrix $A \in \mathbb{R}^{d^n \times d^n}$ is obscenely huge

But we represent the entries of $\mathcal{A}$ is a compact format (e.g. tensor network)

So, building $A \in \mathbb{R}^{d^n \times d^n}$ explicitly is expensive

Computing *tensor-structured products* with A is cheap though

Matrix-Vector Product

Computing Ax where $x = x_1 \otimes \cdots \otimes x_n$ for $x_i \in \mathbb{R}^d$

Same as $\mathcal{A}\mathcal{X}$

Bilinear Product

Computing $x^T A y$ where $x = x_1 \otimes \cdots \otimes x_n$ for $x_i \in \mathbb{R}^d$

Same as $\langle \mathcal{A}, \mathcal{X} \otimes \mathcal{Y} \rangle$

Query Complexity

Fundamental question:

How many Kronecker matrix-vector products
(or bilinear products)
to estimate properties of $\mathcal{A}$?

Prior work requires exponentially many products in the worst case.

Is this necessary?

In this paper:

Yes, exponentially many queries are needed! ☹

Main Results

Trace Estimation

Estimate $tr(A)$ from few matrix-vector products with PSD A

Find $\tilde{t}$ such that:

$$(1-\varepsilon)\,\mathrm{tr}(A) \le \tilde{t} \le (1+\varepsilon)\,\mathrm{tr}(A) \qquad \text{w.h.p.}$$

Classically, Hutchinson's Estimator uses $\ell = O(\frac{1}{\varepsilon^2})$ matvecs

$$\mathbb{E}[x^T A x] = \mathrm{tr}(A) \qquad if \qquad \mathbb{E}[xx^T] = I$$

$$H_\ell(A) = \tfrac{1}{\ell}\textstyle\sum_{i=1}^{\ell} x^{(i)T} A x^{(i)} \qquad for \qquad x^{(i)} \sim \mathcal{D}$$

[Meyer Avron '23] Take $x = x_1 \otimes \cdots \otimes x_k$ where $x_i \sim \mathcal{D}$

For $\varepsilon = O(1)$,

	$x_i \sim \mathcal{D}$	Worst Case	$d = 2$	$d = \Omega(k)$
	$\mathcal{N}(0,I)$	3^k	3^k	3^k
$\mathbb{R}^d$	$\{\pm1\}^d$	$(3-\frac{2}{d})^k$	2^k	3^k
	Unit Vector	$(3-\frac{6}{d+2})^k$	1.5^k	3^k
	$\frac{1}{\sqrt{2}}\mathcal{N}(0,I) + \frac{i}{\sqrt{2}}\mathcal{N}(0,I)$	2^k	2^k	2^k
$\mathbb{C}^d$	$\{\pm1, \pm i\}^d$	$(2-\frac{1}{d})^k$	1.5^k	2^k
	$\mathbb{C}$ *Unit Vector*	$(2-\frac{2}{d+1})^k$	1.33^k	2^k

Kronecker Model sensitive to things we used to not care about*

Distribution choices: Gaussian vs Rademacher vs Unit Vecs

Real versus Complex

Questions:

Are these sensitivities an artifact of Hutchinson? (No!)

Are they fundamental to the Kronecker Matvec Model? (Yes!)

Is $poly(d, k, \frac{1}{\varepsilon})$ possible? (No!)

Main Results

Poly(n) dependency is not* possible!

- Cannot estimate $tr(A)$ with less than $\exp(n)$ bilinear products
- Cannot estimate $\|A\|_2$ with less than $\exp(n)$ matrix-vector products

* See open problem

Subgaussianity does not suffice to understand runtime!

Zero Testing Problem: Determine if $A = 0$ using only Kron-Mat.Vecs.

- 1 MatVec suffices: Let $x = g_1 \otimes \cdots \otimes g_n$ for $g_i \sim \mathcal{N}(0, I)$
- If we force $x_i \in \{\pm1\}^d$, then $\Theta(2^n)$ matrix-vector products needed

Lower Bounds

Main Observation

Mixed Product Property: $(A \otimes B)(x \otimes y) = Ax \otimes By$

Idea 1: From Kronecker Product to iid Product

If $A = A_1 \otimes \cdots \otimes A_n$ then $x^\top A x = \prod_{i=1}^n x_i^\top A_i x_i$

The product of iid random variables is very poorly concentrated

Idea 2: Invoke Orthogonality Locally

If $A_i = uu^\top$ then $x_i^\top A_i x_i = \langle x_i, u_i \rangle^2$

Random vectors are typically orthogonal, so $\langle x_i, u_i \rangle^2 \approx 0$

Overwhelming Orthogonality

We have that $x^\top A x = \prod_{i=1}^n \langle x_i, u_i \rangle^2 = 0$ if at least one $\langle x_i, u_i \rangle = 0$

Zero Testing *(Is the matrix $A = 0$?)*

Sampling $x = x_1 \otimes \cdots \otimes x_k$ for $x_i \sim \mathcal{N}(0, I)$ works with 1 product

Theorem: Any method with $x_i \in \{\pm1\}^d$ needs $\Omega(2^k)$ matvecs

Let $u_i = [\pm1,\ \pm1,\ 0,\ 0,\ \cdots,\ 0]$

Then $\langle x_i, u_i \rangle = 0$ with probability $\frac{1}{2}$

Then $x^\top A x = 0$ with probability $1 - \frac{1}{2^n}$, even though $A \neq 0$!

Implication: Sub-Gaussianity does not matter

Planted Matrix Testing

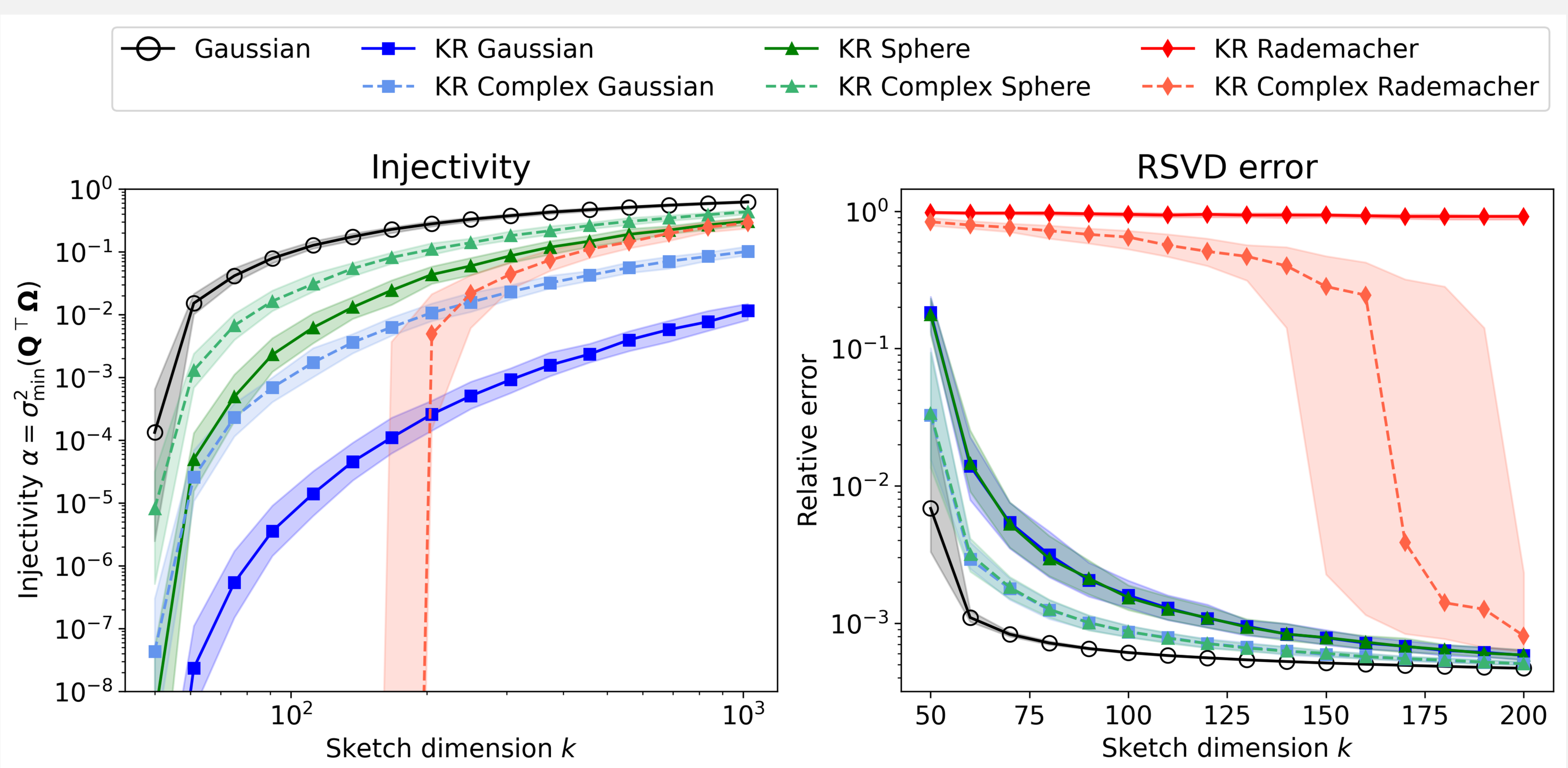
Determine if $A = W$ or $A = W + \lambda uu^T$ from matrix-vector products

Theorem: With mild assumption, any method needs $\Omega(c^k)$ matvecs

Let $u = u_1 \otimes \cdots \otimes u_n$ where $\|u_i\| = \sqrt{d}$ u.a.r.

For any $x = x_1 \otimes \cdots \otimes x_n$, we get $\langle x, u \rangle^2 \approx C^{-n}\|x\|$

So $Wx \approx (W + \lambda uu^\top)x$ for all x



Open Problem

Let $X = \left[x^{(1)}\ x^{(2)}\ \ldots\ x^{(\ell)}\right] \in \mathbb{R}^{d^n \times \ell}$ where $x^{(i)} = x_1^{(i)} \otimes \cdots \otimes x_n^{(i)}$

Let $u = u_1 \otimes \cdots \otimes u_n$ where $\|u_i\| = \sqrt{d}$ randomly

Then $\left\|\mathrm{proj}_{span(X)} u\right\| \le C^{-n}$ for $\ell \le C^{-n}$ for some $C > 1$