Probabilistic results for random tensors

Daniel Kressner

EPFL

Why tensor-structured randomness?

High-dimensional tensor problems live in an ambient space of dimension

$$N = n^d.$$

An unstructured Gaussian vector in $\mathbb{R}^{n^d}$ needs n^d random numbers.

A rank-one tensor probe

$$\omega = \omega^{(d)} \otimes \cdots \otimes \omega^{(1)}, \qquad \omega^{(\mu)} \in \mathbb{R}^n,$$

needs only dn random numbers and often interacts with tensor data without ever forming a length-n^d vector.

Price: strong dependencies among the n^d entries.

Goal of this lecture: Understand this tradeoff more deeply.

Roadmap

1. Rank-one random tensors: Isotropy, small-ball probabilities, fourth moments, trace estimation.
2. Khatri–Rao test matrices: Subspace injection/embedding and why large d is difficult.
3. Random tensor trains: Adding just enough internal randomness to remove exponential dependence on d.

A recurring message:

It's (almost) all about moments.

Rank-one probes

Rank-one random vectors

Let ν be a distribution on $\mathbb{R}^n$ and draw independent

$$\omega^{(1)}, \ldots, \omega^{(d)} \sim \nu.$$

Define

$$\boxed{\omega = \omega^{(d)} \otimes \cdots \otimes \omega^{(1)} \in \mathbb{R}^{n^d}.}$$

Equivalently, ω is the vectorization of the outer product $\omega^{(1)} \circ \cdots \circ \omega^{(d)}$.

Names in different communities:

- rank-one / simple / pure / decomposable tensor,
- product state in quantum physics,
- simple random tensor in probability.

Storage and random-number count: dn instead of n^d.

Good news: Isotropy tensorizes perfectly

Assume base vector is isotropic:

$$\mathbb{E}\left[\omega^{(\mu)}(\omega^{(\mu)})^T\right] = I_n.$$

Independence gives

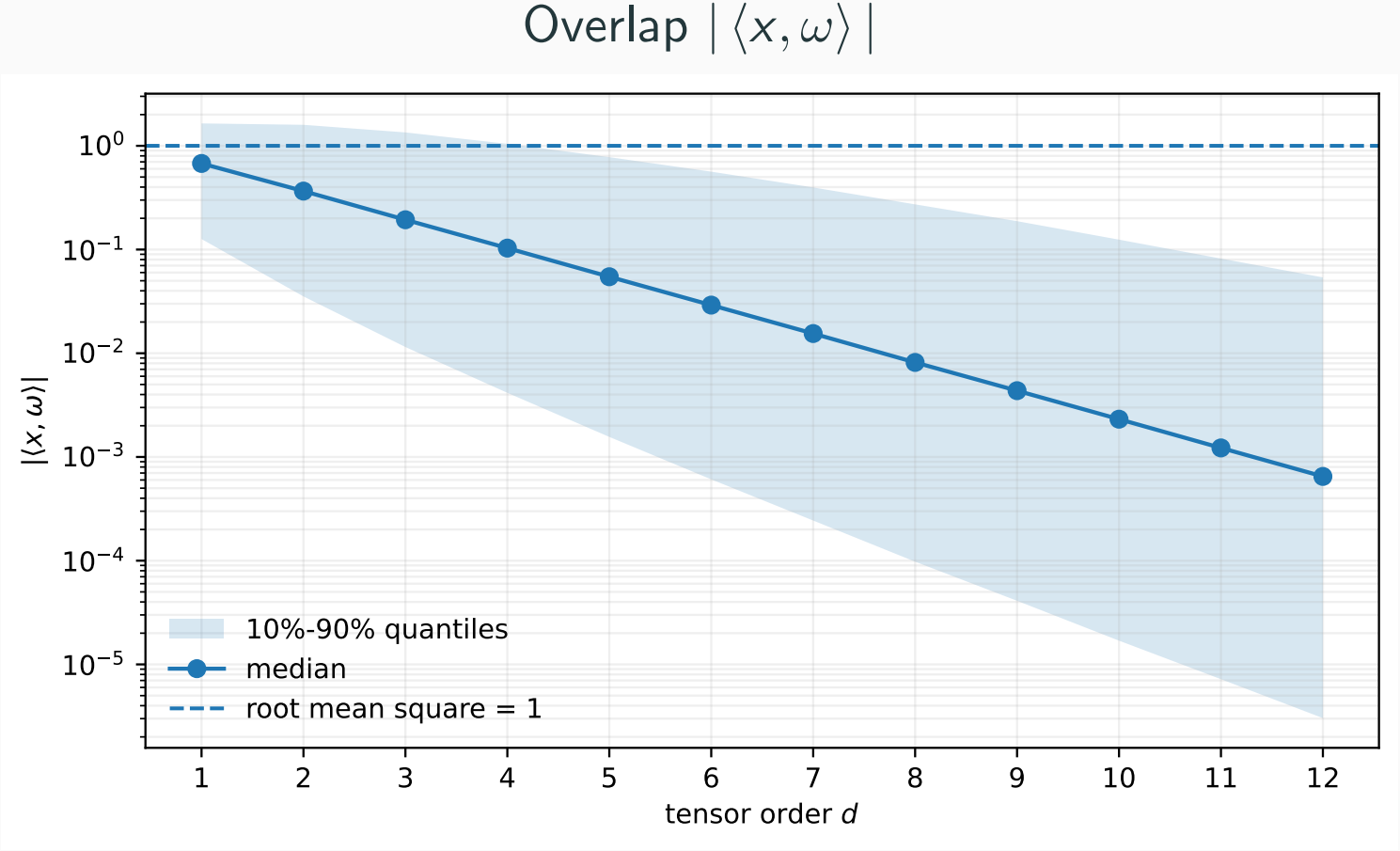
$$\begin{aligned}\mathbb{E}\left[\omega\omega^T\right] &= \mathbb{E}\left[\omega^{(d)}(\omega^{(d)})^T \otimes \cdots \otimes \omega^{(1)}(\omega^{(1)})^T\right] \\ &= I_n \otimes \cdots \otimes I_n = I_{n^d}.\end{aligned}$$

Hence, for deterministic vector $x \in \mathbb{R}^{n^d}$ and $n^d \times n^d$ matrix A:

$$\mathbb{E}\left[|\langle x, \omega\rangle|^2\right] = \|x\|^2, \qquad \mathbb{E}[\omega^T A \omega] = \mathrm{tr}(A).$$

Unbiasedness preserved by tensorization.

Overlaps typically collapse with d



x is rank one; ω is a Gaussian rank-one probe. 300,000 samples for each d.

Second moment $\mathbb{E}\left[|\langle x, \omega \rangle|^2\right]$ maintained by increasingly rare, very large values.

Explanation of overlap collapse

Take rank-one deterministic vector

$$x = x^{(d)} \otimes \cdots \otimes x^{(1)}, \qquad \left\| x^{(\mu)} \right\| = 1,$$

and rank-one Gaussian ω. Then

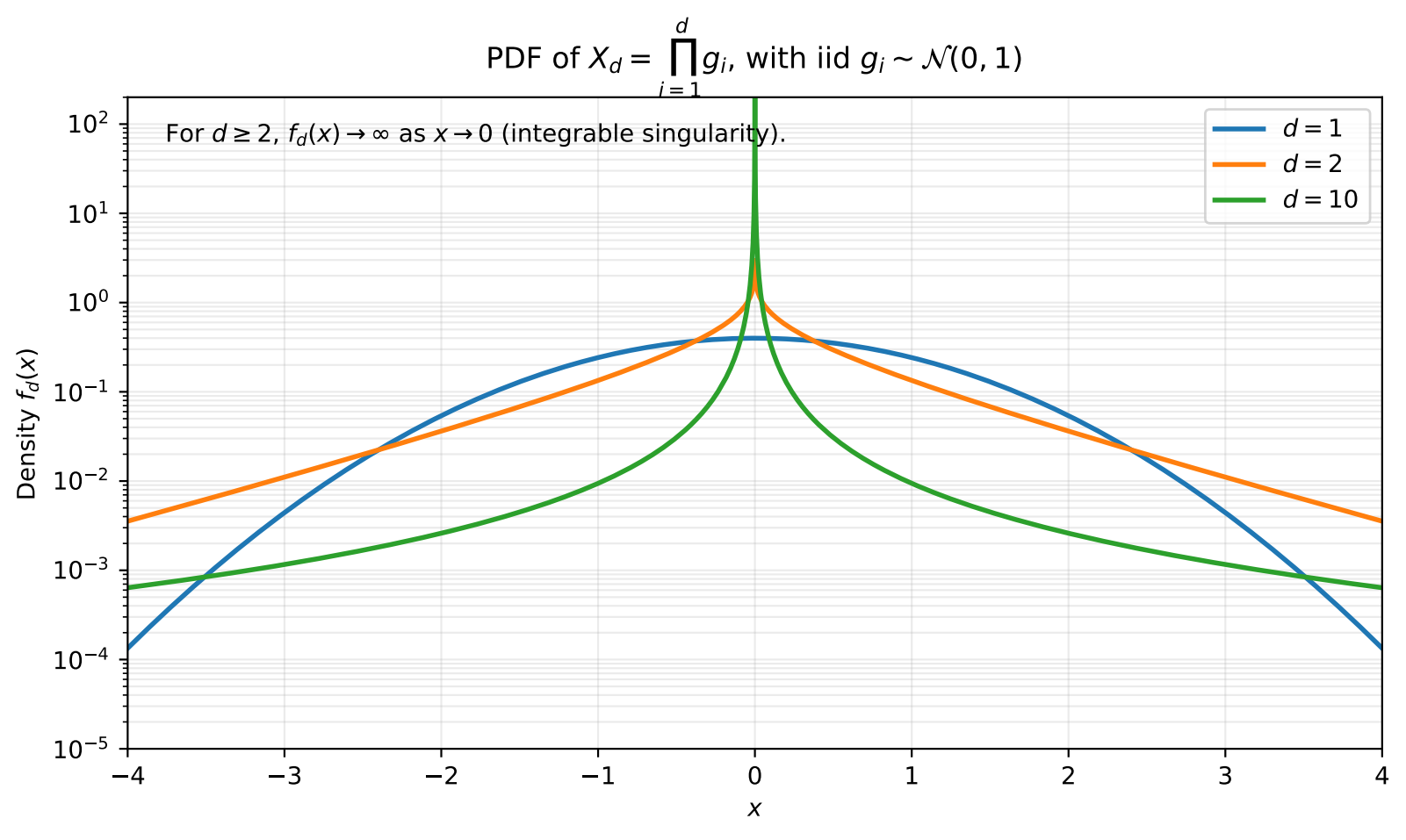
$$\langle x, \omega \rangle = g_1 g_2 \cdots g_d, \qquad g_\mu \overset{\text{iid}}{\sim} N(0,1).$$

While

$$\mathbb{E} |\langle x, \omega \rangle|^2 = 1,$$

the distribution becomes highly concentrated (singular) at 0 and heavy-tailed.

Explanation of overlap collapse



Explanation of overlap collapse

$$\langle x, \omega \rangle = g_1 g_2 \cdots g_d, \qquad g_\mu \overset{\text{iid}}{\sim} N(0,1).$$

Let $g \sim N(0,1)$. Since $g^2 \sim \chi^2_1$,

$$\mathbb{E} \log(g^2) = \frac{\Gamma'(1/2)}{\Gamma(1/2)} + \log 2 \approx -1.97 + \log 2.$$

Therefore $\mathbb{E} \log |g| = -\frac{\mathbb{E} \log(g^2)}{2} \approx -0.64$ and, hence,

$$\mathbb{E} \log |\langle x, \omega \rangle| = \sum_{\mu=1}^{d} \mathbb{E} \log |g_\mu| \approx -0.64 \cdot d.$$

By the law of large numbers,

$$|\langle x, \omega \rangle|^{1/d} \to e^{-(\gamma + \log 2)/2} \approx 0.530.$$

Thus a typical overlap is of size roughly 0.53^d.

Small-ball probability: $d = 2$

For $d = 2$, write $x = \mathrm{vec}(X)$ and let $\omega^{(1)}, \omega^{(2)} \sim N(0, I_n)$ independently. Then

$$\langle x, \omega \rangle = \left(\omega^{(1)}\right)^T X \omega^{(2)}, \qquad \|x\| = \|X\|_F .$$

By [Bujanović/Kressner'21], for $\epsilon > 0$,

$$\boxed{\Pr\left\{ \left|\left(\omega^{(1)}\right)^T X \omega^{(2)}\right| \le \epsilon \|X\|_F \right\} \le \frac{2}{\pi} \epsilon \left(2 + \log(1 + 2\epsilon^{-1})\right).}$$

For comparison, for unstructured Gaussian ω results holds with $O(\epsilon)$.

So, for $d = 2$ no severe underestimation caused by rank-one structure (overestimation easier to rule out).

Application to building surrogates of matrix-valued function $\lambda \mapsto A(\lambda)$ for rational approximation / nonlinear eigenvalue problems [Güttel/Kressner/Vandereycken'2024].

Results for $d > 2$ in [Hu/Paouris'26; Vershynin'20; Bamberger/Krahmer/Ward'22].

Directional fourth moments

For isotropic random vector $\xi \in \mathbb{R}^N$, define

$$\boxed{h_4(\xi) := \sup_{\|u\|=1} \mathbb{E}|\langle u, \xi \rangle|^4.}$$

For unstructured standard Gaussian vector, $h_4 = 3$, independent of N.

Tensorization of the fourth moment: the case $d = 2$

Let

$$\omega = \omega^{(2)} \otimes \omega^{(1)}, \qquad h := h_4(\omega^{(1)}) = \sup_{\|x\|=1} \mathbb{E}|\langle x, \omega^{(1)}\rangle|^4,$$

where $\omega^{(1)}, \omega^{(2)} \in \mathbb{R}^n$ are iid isotropic.

For $u \in \mathbb{R}^{n^2}$, write $u = \mathrm{vec}(U)$ with $U \in \mathbb{R}^{n\times n}$. Then

$$\langle u, \omega\rangle = (\omega^{(2)})^T U \omega^{(1)}.$$

Conditioning on $\omega^{(1)}$ gives

$$\mathbb{E}_{\omega^{(2)}} \left|(\omega^{(2)})^T U \omega^{(1)}\right|^4 \le h \|U\omega^{(1)}\|^4.$$

Now let $r_1^T, \dots, r_n^T$ denote the rows of U. Then

$$\|U\omega^{(1)}\|^4 = \Big(\sum_{j=1}^n |\langle r_j, \omega^{(1)}\rangle|^2\Big)^2 = \sum_{j,k=1}^n |\langle r_j, \omega^{(1)}\rangle|^2 |\langle r_k, \omega^{(1)}\rangle|^2.$$

Tensorization of the fourth moment: the case $d = 2$

By Cauchy–Schwarz,

$$\mathbb{E}\left[|\langle r_j, \omega^{(1)}\rangle|^2 |\langle r_k, \omega^{(1)}\rangle|^2\right] \le \sqrt{\mathbb{E}|\langle r_j, \omega^{(1)}\rangle|^4 \, \mathbb{E}|\langle r_k, \omega^{(1)}\rangle|^4} \le h \, \|r_j\|^2 \|r_k\|^2.$$

Therefore

$$\mathbb{E}\|U\omega^{(1)}\|^4 \le h \sum_{j,k=1}^{n} \|r_j\|^2 \|r_k\|^2 = h\Big(\sum_{j=1}^{n} \|r_j\|^2\Big)^2 = h\,\|U\|_F^4.$$

Combining both estimates,

$$\boxed{\mathbb{E}|\langle u, \omega\rangle|^4 \le h^2 \|U\|_F^4 = h^2 \|u\|^4.}$$

Hence

$$h_4(\omega) \le h^2.$$

Equality is attained by choosing

$$u = u_2 \otimes u_1,$$

where u_1, u_2 are worst directions for the base distribution.

Fourth moments tensorize: The curse of dimensionality!

Let $h = h_4(\omega^{(1)})$ for iid isotropic factors and set $\omega = \omega^{(d)} \otimes \cdots \otimes \omega^{(1)}$. Then

$$\boxed{h_4(\omega) = h^d.}$$

Proof: Condition successively on the factors.

Examples in $\mathbb{R}^n$:

base distribution	$h_4(\omega^{(1)})$	$h_4(\omega)$
$N(0, I_n)$	3	3^d
Rademacher	$3 - 2/n$	$(3 - 2/n)^d$
uniform on $\sqrt{n}S^{n-1}$	$3n/(n+2)$	$(3n/(n+2))^d$

[Meyer–Avron'26]

Paley–Zygmund: a lower-tail inequality from two moments

Let $Z \geq 0$ with $\mathbb{E}Z^2 < \infty$. For every $0 < \theta < 1$,

$$\boxed{\mathbb{P}\{Z \geq \theta\,\mathbb{E}Z\} \;\geq\; (1-\theta)^2 \frac{(\mathbb{E}Z)^2}{\mathbb{E}Z^2}.}$$

[Paley–Zygmund'32]

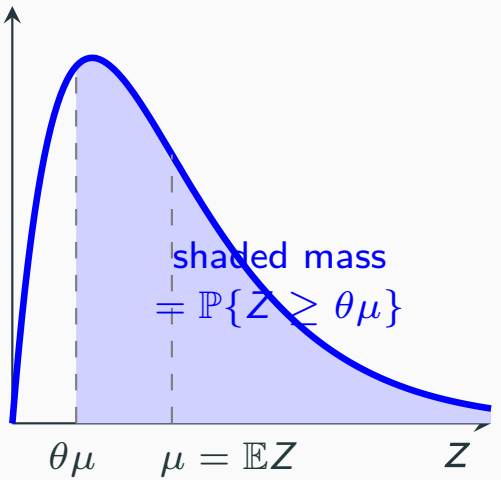
One-line proof. Put $E = \{Z \geq \theta \mathbb{E}Z\}$.
Then

$$\begin{aligned}\mathbb{E}Z &= \mathbb{E}[Z\mathbf{1}_{E^c}] + \mathbb{E}[Z\mathbf{1}_E] \\ &\leq \theta\mathbb{E}Z + \sqrt{\mathbb{E}Z^2\,\mathbb{P}(E)}.\end{aligned}$$

Hence

$$(1-\theta)\mathbb{E}Z \leq \sqrt{\mathbb{E}Z^2\,\mathbb{P}(E)},$$

which gives the inequality after squaring.



Directional fourth moments + Paley–Zygmund

For isotropic $\xi \in \mathbb{R}^N$, define

$$h_4(\xi) := \sup_{\|u\|=1} \mathbb{E}\,|\langle u, \xi\rangle|^4.$$

Apply Paley–Zygmund. Fix $\|u\| = 1$ and set

$$X = \langle u, \xi\rangle, \qquad Z = X^2.$$

Then

$$\mathbb{E}Z = 1, \qquad \mathbb{E}Z^2 = \mathbb{E}X^4 \le h_4(\xi).$$

With $\theta = t^2$,

$$\begin{aligned}\mathbb{P}\{|X| \ge t\} &= \mathbb{P}\{Z \ge t^2\mathbb{E}Z\} \\ &\ge (1-t^2)^2 \frac{(\mathbb{E}Z)^2}{\mathbb{E}Z^2} \\ &\ge \boxed{\frac{(1-t^2)^2}{h_4(\xi)}}.\end{aligned}$$

This holds uniformly for every unit direction u.

Gaussian rank-one tensor.

$$\omega = g^{(d)} \otimes \cdots \otimes g^{(1)}, \qquad g^{(\mu)} \sim N(0, I_n).$$

Its worst directional fourth moment is

$$\boxed{h_4(\omega) = 3^d.}$$

Choosing $t = 1/2$ gives

$$\boxed{\mathbb{P}\{|\langle u, \omega\rangle| \ge \tfrac{1}{2}\} \ge \frac{9}{16 \cdot 3^d}.}$$

Anti-concentration degrades like 3^{-d}.

Trace estimation: fourth moments become sample complexity

Let A be SPSD with eigenvectors $u_1, \dots, u_N$ and eigenvalues $\lambda_1, \dots, \lambda_n \geq 0$.
Let ξ be isotropic. By Cauchy–Schwarz,

$$\mathbb{E}(\xi^T A \xi)^2 = \sum_{i,j} \lambda_i \lambda_j \mathbb{E}[(u_i^T \xi)^2 (u_j^T \xi)^2] \leq h_4(\xi) \, (\operatorname{tr} A)^2.$$

Therefore

$$\boxed{\operatorname{Var}(\xi^T A \xi) \leq (h_4(\xi) - 1)(\operatorname{tr} A)^2.}$$

For ℓ independent probes,

$$\operatorname{Var}(H_\ell(A)) \leq \frac{h_4(\xi) - 1}{\ell} (\operatorname{tr} A)^2.$$

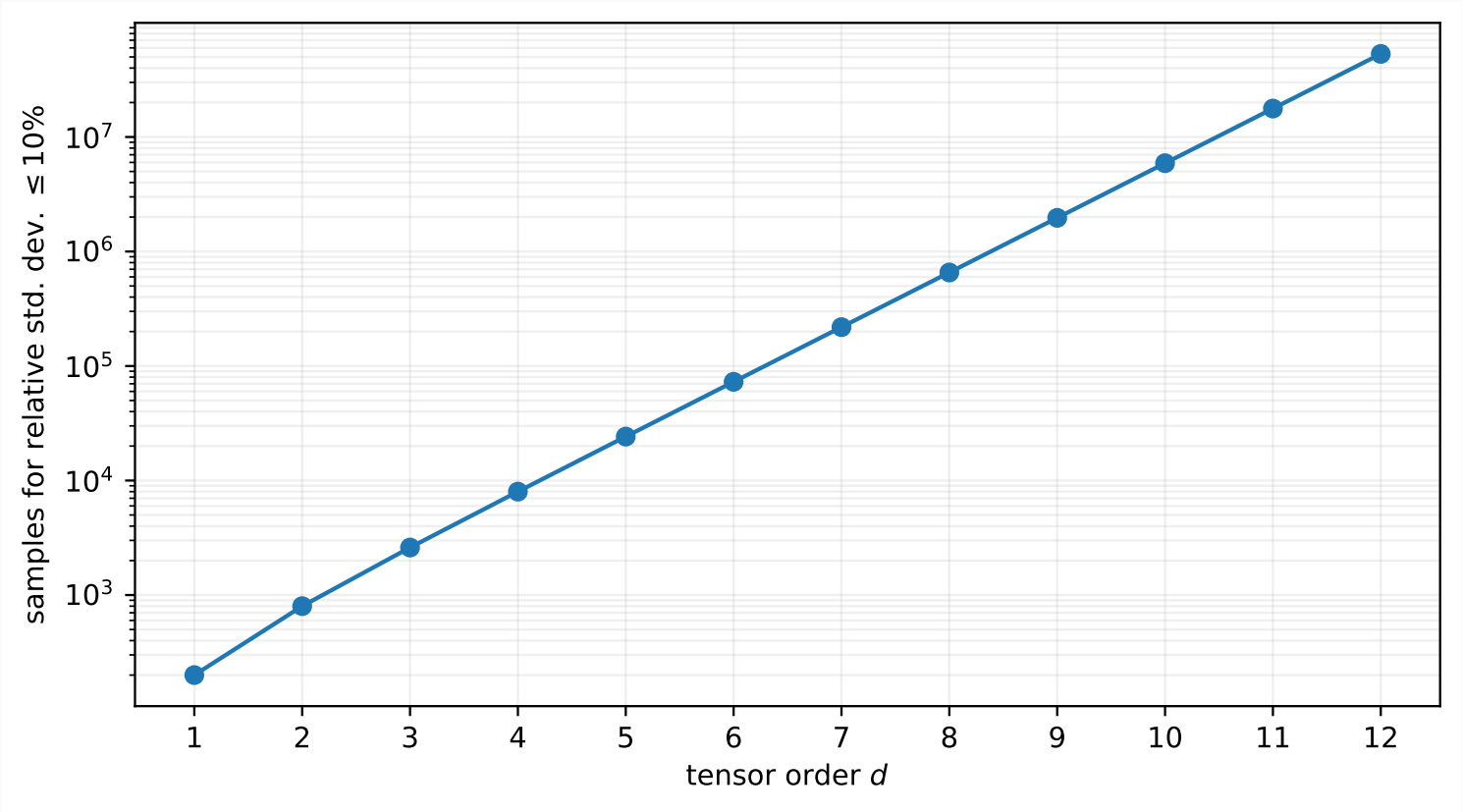
This inequality is *tight*: take $A = uu^T$ in a worst fourth-moment direction.

For Gaussian rank-one probes, worst-case relative standard deviation is

$$\sqrt{\frac{3^d - 1}{\ell}}.$$

[Meyer–Avron'26]

Experiment: Trace estimation deteriorates exponentially



Gaussian rank-one probes and $A = uu^T$:

$(3^d - 1)/0.1^2$ samples for 10% relative standard deviation.

This is not merely a weakness of Hutchinson's estimator

Meyer–Swartworth–Woodruff study access model

$$x = x_1 \otimes \cdots \otimes x_d \quad \longmapsto \quad Ax.$$

They prove exponential query lower bounds for several tasks, including

- trace estimation,
- top-eigenvalue estimation,
- even basic testing problems for restricted query distributions.

Their lower bounds allow adaptive algorithms (under a mild conditioning assumption on the queries).

For rank-one queries, exponential dependence on d is often unavoidable.

[Meyer–Swartworth–Woodruff, ICML'25]

Khatri–Rao

From one probe to a test matrix

Let

$$\Omega^{(\mu)} = [\omega_1^{(\mu)}, \dots, \omega_s^{(\mu)}] \in \mathbb{R}^{n \times s}, \qquad \mu = 1, \dots, d,$$

have independent isotropic columns. The Khatri–Rao product is the columnwise Kronecker product:

$$\boxed{\Omega = \frac{1}{\sqrt{s}} \left(\Omega^{(d)} \odot \cdots \odot \Omega^{(1)} \right) = \frac{1}{\sqrt{s}} [\omega_1, \dots, \omega_s] \in \mathbb{R}^{n^d \times s}.}$$

Each column ω_j is a rank-one random vector.

Construction/storage cost: dns random numbers rather than $n^d s$.

For tensor-structured inputs, $\Omega^T X$ can often be formed without ever materializing Ω.

Recall: OSE versus OSI

Fix an r-dimensional subspace $V \subset \mathbb{R}^N$.

Oblivious subspace embedding (OSE)

With probability $\geq 1 - \delta$,

$$(1 - \varepsilon) \|x\|^2 \leq \left\|\Omega^T x\right\|^2 \leq (1 + \varepsilon) \|x\|^2 \qquad \forall x \in V.$$

Oblivious subspace injection (OSI)

Only lower bound (no vector in V is nearly annihilated) + isotropy.

Distinction matters a lot for heavy-tailed sketches: Lower tails are much easier than two-sided concentration.

Fourth-moments $\Rightarrow$ OSI

Theorem (Oliveira)

Let $\omega \in \mathbb{R}^N$ be isotropic with finite $h_4(\omega)$, and let

$$\Omega = s^{-1/2}[\omega_1, \ldots, \omega_s].$$

There is a universal C such that an r-dimensional OSI with injectivity parameter ϵ and failure probability δ holds provided

$$\boxed{s \geq C\, h_4(\omega)\, \epsilon^{-2}\big(r + \log(1/\delta)\big).}$$

Proof idea: Lower-tail bound for minimum eigenvalue of the empirical covariance, using only isotropy and fourth moments.

[Oliveira'16; see also Tropp'26]

Immediate consequence for Khatri–Rao sketches

For Gaussian rank-one columns,

$$h_4(\omega) = 3^d.$$

Therefore a constant-quality OSI follows from

$$\boxed{s \gtrsim 3^d \big(r + \log(1/\delta)\big).}$$

For Rademacher factors, replace 3^d by $(3 - 2/n)^d$.

Excellent in subspace dimension r (linear, just as for an unstructured Gaussian sketch), but exponential in tensor order!

[Camaño–Epperly–Meyer–Tropp'25]

Stress test for a Khatri–Rao sketch

To see the deterioration directly, choose the coherent tensor subspace

$$V = \mathrm{span}\{e_1^{\otimes d}, \dots, e_r^{\otimes d}\}, \qquad n \geq r.$$

Let Q contain this orthonormal basis and let Ω be a Gaussian Khatri–Rao sketch.

Then

$$(Q^T \Omega)_{ij} = \frac{1}{\sqrt{s}} \prod_{\mu=1}^{d} g_{\mu ij},$$

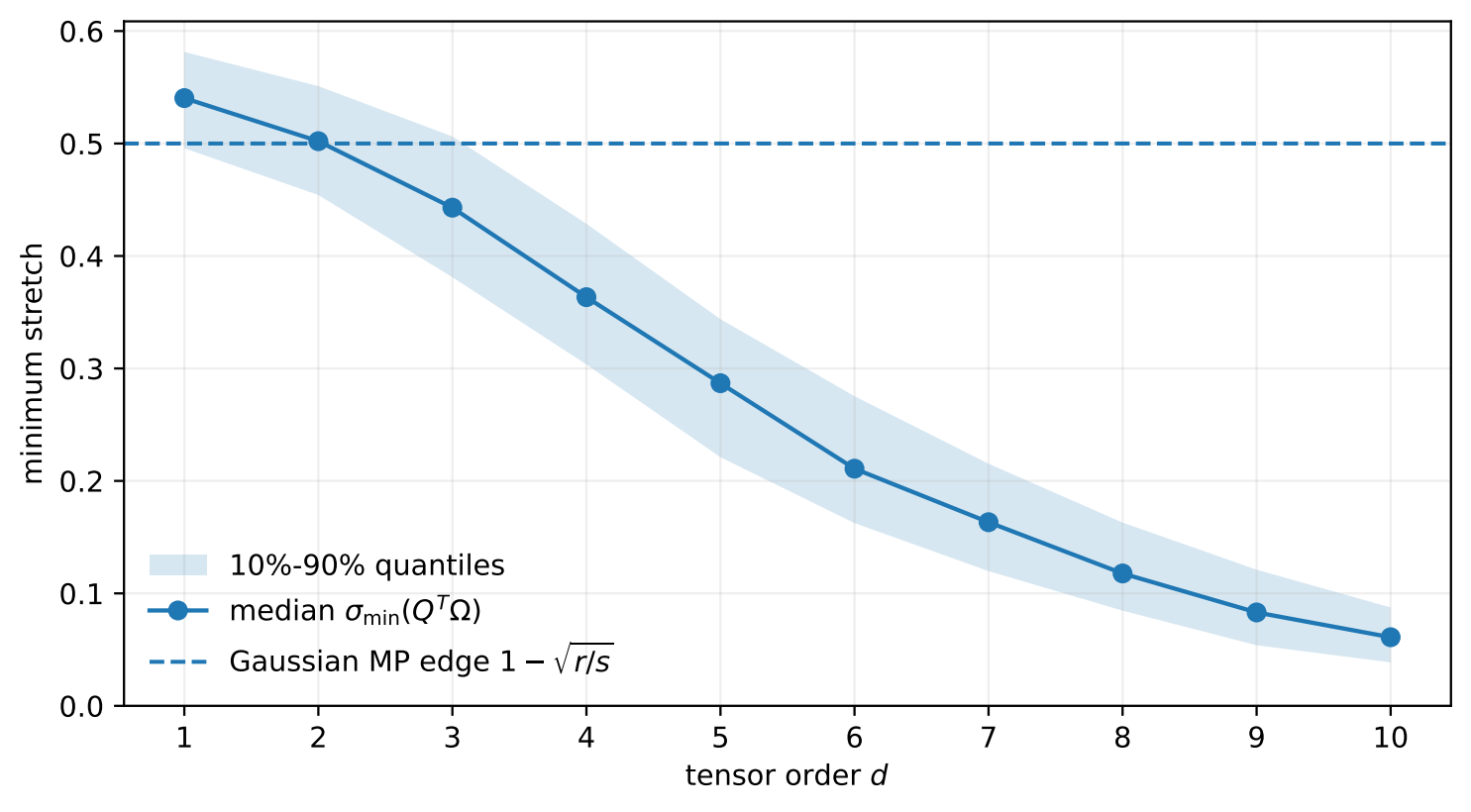
so the $r \times s$ compressed matrix has iid entries equal to products of d Gaussians.

We monitor

$$\sigma_{\min}(Q^T \Omega),$$

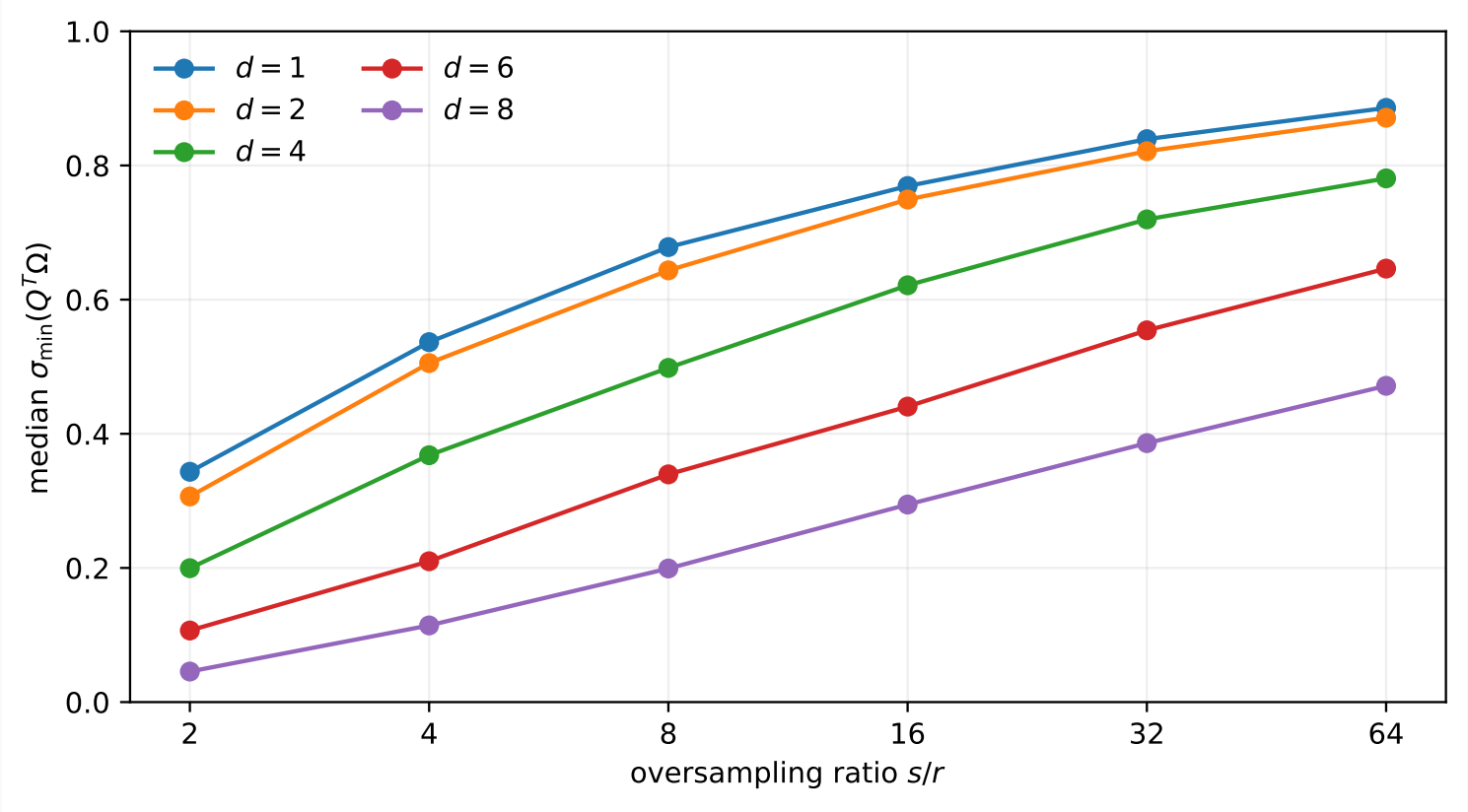
the minimum stretch on V.

Experiment: fixed oversampling no longer suffices



$r = 20$, $s = 4r$, 350 independent trials. $d = 1$ is the ordinary Gaussian case.

Experiment: increasing d demands increasingly aggressive oversampling



Same coherent tensor subspace, $r = 20$. Curves show median over 130 trials.

OSE via Johnson–Lindenstrauss moments: the classical route

Ahle et al. proved JL moment property for Khatri–Rao sketches with isotropic sub-Gaussian factors. For one fixed vector, bound on number of samples s takes the form

$$s = O_d\left(\frac{\log(1/\delta)}{\varepsilon^2} + \frac{\log(1/\delta)^d}{\varepsilon}\right).$$

ε-net of an r-dimensional sphere contains $\sim 3^r$ points for $\varepsilon = 1/2$. Replacing δ by $\delta/3^r$ yields

$$s = O_d\left(\frac{r + \log(1/\delta)}{\varepsilon^2} + \frac{(r + \log(1/\delta))^d}{\varepsilon}\right).$$

ε-net argument turns
bad $\log(1/\delta)^d$ dependence into bad r^d dependence.

[Ahle–Kapralov–Knudsen–Pagh–Velingker–Woodruff–Zandieh, SODA'20]

Hot off the press

Beretta–Musco (August 2026) improve dependence on r.

Theorem

Let Khatri–Rao factors have independent isotropic sub-Gaussian columns. Then

$$\boxed{s = O_d\left(\varepsilon^{-2}\, r\, \log^{d+1}\left(\frac{r}{\varepsilon\delta}\right)\right)}$$

suffices for an (ε, δ, r) OSE.

Here $O_d(\cdot)$ hides $2^{O(d)}$ factors.

Thus, for each fixed d, the dependence is essentially the optimal r/ε^2, up to logarithms.

Nearly optimal for low orders.

[Beretta–Musco, arXiv:2608.28094]

Random TT

Can we inject more randomness without losing tensor structure?

A natural idea is a normalized sum of R independent rank-one tensors. This helps, but fourth-moment bounds predict that R must itself grow exponentially with d to obtain dimension-independent behavior.

Tensor networks provide a more effective interpolation between

$$\text{rank one} \longleftrightarrow \text{Gaussian-like randomness.}$$

Most important example: Tensor train (TT) format, also known as a matrix product state.

[Rakhshan–Rabusseau'20; Sun–Guo–Tropp–Udell'21; Oseledets'11; Schollwöck'11]

Tensor trains in one slide

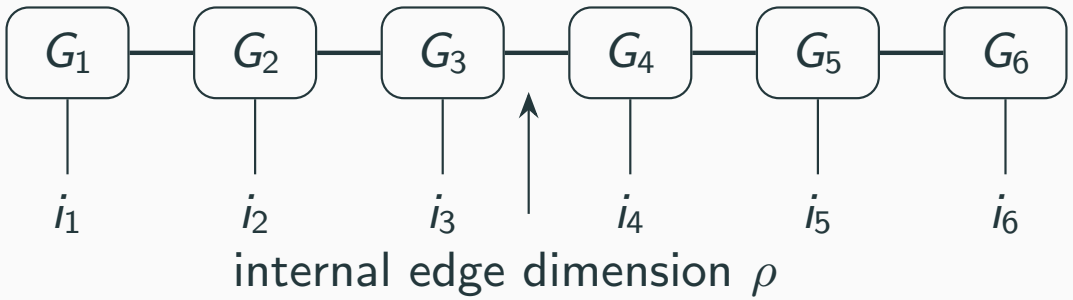
A tensor $X \in \mathbb{R}^{n \times \cdots \times n}$ has a TT representation

$$X(i_1, \ldots, i_d) = X_1(i_1) X_2(i_2) \cdots X_d(i_d),$$

where

$$X_1(i_1) \in \mathbb{R}^{1 \times \rho_1}, \quad X_\mu(i_\mu) \in \mathbb{R}^{\rho_{\mu-1} \times \rho_\mu}, \quad X_d(i_d) \in \mathbb{R}^{\rho_{d-1} \times 1}.$$

For uniform TT rank ρ, storage is $O(dn\rho^2)$.



Gaussian random tensor trains

Fill the TT cores independently with standard Gaussian entries and normalize:

$$W(i_1, \ldots, i_d) = \rho^{-(d-1)/2} G_1(i_1) \cdots G_d(i_d).$$

Let $\omega = \mathrm{vec}(W) \in \mathbb{R}^{n^d}$. Then

$$\mathbb{E}\omega = 0, \qquad \mathbb{E}\omega\omega^T = I_{n^d}.$$

So Gaussian random TT vectors are isotropic.

At $\rho = 1$, this is exactly the Gaussian rank-one construction.
Choosing modest TT rank ρ increases internal randomness while preserving low construction/storage requirements.

[Huber–Schneider–Wolf'17; Rakhshan–Rabusseau'20; Camaño et al.'26]

Key moment bounds for random TT vectors

For a Gaussian random TT vector of rank ρ,

$$\boxed{h_4(\omega) \leq 3\left(1+\frac{2}{\rho}\right)^{d-1}.}$$

More generally, for every unit vector u and integer $p \geq 1$,

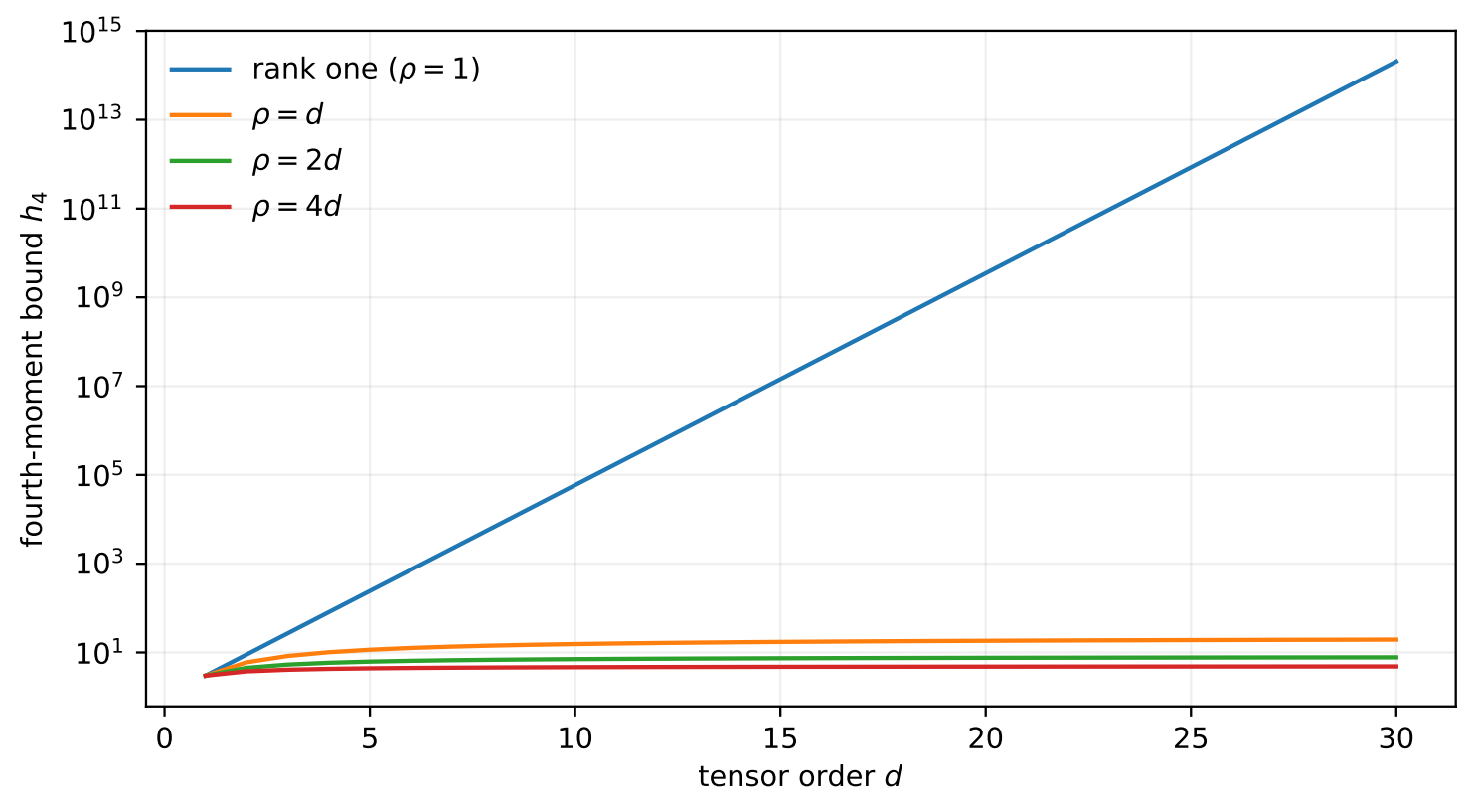
$$\boxed{\mathbb{E}(u^T\omega)^{2p} \leq (2p-1)!!\,\exp\left(\frac{p^2(d-1)}{\rho}\right).}$$

If $\rho \gtrsim d$, every fixed-order directional moment is bounded uniformly in d.

TT rank $\rho \sim d$ sufficent to remove exponential fourth-moment blow-up!

[Rakhshan–Rabusseau'20; Bujanović–Kressner–Olić'26]

Fourth moments: Random rank one versus random TT



Bound $3(1 + 2/\rho)^{d-1}$. For $\rho = d$, it tends to $3e^2$ rather than growing like 3^d.

Random TT test matrices: OSI without exponential d

Form

$$\Omega = \frac{1}{\sqrt{s}}[\omega_1, \ldots, \omega_s]$$

from independent Gaussian random TT vectors of rank ρ.

Combine

$$h_4(\omega) \leq 3\left(1 + \frac{2}{\rho}\right)^{d-1}$$

with Oliveira's fourth-moment OSI theorem.

Consequence

Taking

$$\boxed{\rho = O(d), \qquad s = O\big(\epsilon^{-2}(r + \log(1/\delta))\big)}$$

gives an OSI for an r-dimensional subspace.

Number of independent Gaussian random variables is only $O(nd^3r)$ at constant accuracy, versus $O(nd\,3^d r)$ for the corresponding Khatri–Rao fourth-moment guarantee.

[Camaño et al.'26; Bujanović–Kressner–Olić'26]

Two-sided embedding: TTStack

Cazeaux–Dupuy–Figueroa Justiniano introduce a block TT construction with two parameters:

- TT rank ρ controls internal randomness,
- One extra leg of size $\alpha = \rho \rightsquigarrow$ block TT vector.
- P independent blocks $\rightsquigarrow$ embedding dimension $s = P\rho$.

Includes Khatri–Rao ($\rho = 1$), a Gaussian TT map ($\alpha = 1$), and construction by Ma/Solomonik'2022.

TTStack OSE

Choose $\rho \gtrsim d$. Then an r-dimensional OSE with distortion ε and failure probability δ is obtained with

$$\boxed{\rho = O\big(d(r + \log(1/\delta))\big), \qquad P = O(\varepsilon^{-2}).}$$

Therefore: $\quad s = P\rho = O\big(d\varepsilon^{-2}(r + \log(1/\delta))\big).$

Linear dependence on both d and r at the level of embedding dimension.

Trace estimation with Gaussian random TT probes

For SPSD A and ℓ independent Gaussian random TT probes of rank ρ,

$$\boxed{\operatorname{Var}(H_\ell(A)) \le \frac{1}{\ell}\left[3\left(1+\frac{2}{\rho}\right)^{d-1}-1\right](\operatorname{tr} A)^2.}$$

Thus $\rho \gtrsim d$ gives order-independent worst-case variance.

Two improvements/refinments:

- [Bujanović–Kressner–Olić, arXiv:2606.15679] also derives higher-moment/tail bounds. With median-of-means amplification, $\rho \ge d-1$ and $\ell \gtrsim \varepsilon^{-2}\log(1/\delta)$ recover the classical dependence on accuracy and failure probability.
- [Camaño–Epperly–Meyer–Tropp, arXiv:2606.15350] computes full fourth-moment tensor to get bound of the form that allows one to conclude $O(\epsilon^{-1})$ sample complexity Nyström++ using Gaussian random TT probes.

Summary and applications

Summary of results

Random vector	Storage / randomness	Fourth moment	Guarantee
Unstructured Gaussian	n^d per vector	3	$s \sim r/\varepsilon^2$ OSE
Rank-one Gaussian / KRP	dn per vector	3^d	OSI $s \sim 3^d r$; OSE nearly linear in r for fixed d
Gaussian TT, $\rho \sim d$	$dn\rho^2$ per vector	$O(1)$	OSI $s \sim r$
TTStack	structured	controlled via ρ, P	OSE $s \sim dr/\varepsilon^2$

Tensor networks trade a modest increase in representation complexity for a dramatic improvement in probability.

Note: Extension to tensor tree networks by Daniil Homza.

Where these sketches are used

- TT rounding / recompression. Khatri–Rao sketches support fixed-rank and adaptive randomized TT rounding; the adaptive method uses a KRP Frobenius-norm estimator to monitor the residual.
- Tucker approximation and streaming. KRP measurements accelerate range finding on tensor unfoldings and can be applied without forming the exponentially large test matrix.
- Trace estimation. Random TT probes give matrix-free estimators compatible with tensor-network operators and can be combined with Nyström++.
- Exponential-scale linear algebra / quantum many-body problems. Tensor-network dimension reduction gives provable algorithms at ambient dimensions such as 2^{200}.

[Al Daas et al.'25; Saibaba–Verma–Ballard'25; Haselby et al.'26; Camaño et al.'26; Kressner–Vandereycken–Vorhaar'23]

Take-home messages

1. Isotropy survives rank-one tensorization, but small-ball probabilities and fourth moments deteriorate strongly with d.
2. The fourth moment h_4 is a useful unifying quantity: it explains both trace-estimation variance and Khatri–Rao subspace-injection bounds.
3. New Khatri–Rao OSE theory is nearly optimal in the subspace dimension r for fixed d, but large tensor order remains difficult.
4. Random tensor trains with rank $\rho \sim d$ inject enough randomness to remove exponential dependence on d from key moment, OSI, OSE, and trace-estimation guarantees.

References

Selected references I

BK'21 Z. Bujanović and D. Kressner. *Norm and trace estimation with random rank-one vectors*. SIAM J. Matrix Anal. Appl. 42 (2021), 202–223.

HP'26 X. Hu and G. Paouris. *Small ball probabilities for simple random tensors*. J. Funct. Anal. 290 (2026), 111309.

V'20 R. Vershynin. *Concentration inequalities for random tensors*. Bernoulli 26 (2020), 3139–3162.

BKW'22 S. Bamberger, F. Krahmer, R. Ward. *The Hanson–Wright inequality for random tensors*. Sampling Theory, Signal Processing, and Data Analysis 20 (2022), 14.

MA'26 R. A. Meyer and H. Avron. *Hutchinson's estimator is bad at Kronecker-Trace-Estimation*. SIAM J. Matrix Anal. Appl. 47 (2026), 353–387.

MSW'25 R. A. Meyer, W. J. Swartworth, D. P. Woodruff. *Understanding the Kronecker matrix-vector complexity of linear algebra*. ICML/PMLR 267 (2025), 43909–43933.

Selected references II

O'16 R. I. Oliveira. *The lower tail of random quadratic forms with applications to ordinary least squares*. Probab. Theory Relat. Fields 166 (2016), 1175–1194.

A+20 T. D. Ahle et al. *Oblivious sketching of high-degree polynomial kernels*. SODA 2020, 141–160.

BGKL'26 Z. Bujanović, L. Grubišić, D. Kressner, H. Y. Lam. *Subspace embedding with random Khatri–Rao products and its application to eigensolvers*. IMA J. Numer. Anal. 46 (2026), 1328–1355.

BM'26 L. Beretta and C. Musco. *A tight analysis of Khatri-Rao oblivious subspace embeddings*. arXiv:2608.28094 (2026).

SVB'25 A. K. Saibaba, B. D. Verma, G. Ballard. *Improved analysis of Khatri–Rao random projections and applications*. arXiv:2507.23207v2.

Selected references III

RR'20 B. Rakhshan and G. Rabusseau. *Tensorized random projections*. AISTATS/PMLR 108 (2020), 3306–3316.

BKO'26 Z. Bujanović, D. Kressner, H. Olić. *Stochastic trace estimation with tensor train random vectors*. arXiv:2606.15679 (2026).

CEMT'26 C. Camaño, E. N. Epperly, R. A. Meyer, J. A. Tropp. *Linear algebra at exponential scale via tensor network dimension reduction*. arXiv:2606.15350 (2026).

CDF'26 P. Cazeaux, M.-S. Dupuy, R. Figueroa Justiniano. *Linear-scaling tensor train sketching*. arXiv:2603.11009v3 (2026).

AD+25 H. Al Daas et al. *Adaptive randomized tensor train rounding using Khatri–Rao products*. arXiv:2511.03598 (2025).

OK'11 I. V. Oseledets. *Tensor-train decomposition*. SIAM J. Sci. Comput. 33 (2011), 2295–2317.