Norm Bounds for Sparse Random Tensors and Spectral Gap of Random Hypergraphs

Kevin Lucca[*] Lucas Pesenti[†]

July 8, 2026

Abstract

Friedman and Wigderson (1995) introduced a notion of second eigenvalue for hypergraphs that generalizes the second eigenvalue of the adjacency matrix of a graph. We show that r-uniform Erdős-Rényi hypergraphs on n vertices exhibit a spectral gap as soon as their expected number of hyperedges m satisfies $m \gg n^{r/2}$. Prior work identified this scale only up to logarithmic factors; removing these factors is the main technical challenge.

Our proof overcomes this obstacle through an explicit decomposition of an associated selector process, inspired by a generic decomposition theorem of Talagrand (2021). As a consequence of our techniques, we obtain improved norm bounds for sparse random tensors with independent entries. Finally, under a mild moment equivalence assumption, we extend to tensors a seminal result of Seginer (2000) for random matrices with i.i.d. entries.

1 Introduction

Can we develop a spectral theory for hypergraphs that matches the power of the well-developed spectral theory of graphs? This long-standing question motivates the development of new tools for studying matrices that extend to tensors.

In this line of work, Friedman and Wigderson [FW95] introduced a notion of second eigenvalue for hypergraphs that generalizes the second eigenvalue of the adjacency matrix of a graph. They also asked for the value of this second eigenvalue for *random* hypergraphs.

Definition 1.1 (Erdős-Rényi model)**.** Let $n, r \in \mathbb{N}$ and $p = p(r, n) \in [0, 1]$. Consider the distribution over r-uniform hypergraphs on the vertex set $[n] := \{1, \ldots, n\}$ obtained by independently including each subset of vertices $e \subseteq [n]$ of size $|e| = r$ as a hyperedge with probability p.

We represent a hypergraph by its adjacency tensor $T \in (\mathbb{R}^n)^{\otimes r}$, defined by $T_{i_1,\ldots,i_r} = 1$ if $\{i_1, \ldots, i_r\}$ is a hyperedge, and $T_{i_1,\ldots,i_r} = 0$ otherwise. Let $\mathcal{H}_r(n, p)$ denote the distribution of the adjacency tensors of Erdős-Rényi hypergraphs.

Definition 1.2 (Injective tensor norm)**.** The injective norm of $T \in (\mathbb{R}^n)^{\otimes r}$ is

$$\|T\|_{\text{inj}} := \max_{X \in \mathcal{X}} |\langle T, X \rangle| \ ,$$

where $\mathcal{X} := \{x_1 \otimes \ldots \otimes x_r : \|x_1\|_2 = \ldots = \|x_r\|_2 = 1\}$ is the set of unit rank-1 tensors.

[*]ETH Zürich. kevin.lucca@ifor.math.ethz.ch
[†]ETH Zürich. lpesenti@ethz.ch.

When T is invariant under permutations of its indices, the injective norm can equivalently be written as the coupled maximization problem over $\{x^{\otimes r} : \|x\|_2 = 1\}$ [Kel28].

Following the terminology of [FW95], when $T \sim \mathcal{H}_r(n,p)$, we call $\|T - \mathbb{E}\,T\|_{\text{inj}}$ the *second eigenvalue*[1] of T, and $\|T\|_{\text{inj}} - \|T - \mathbb{E}\,T\|_{\text{inj}}$ the *spectral gap* of T. More explicitly, the second eigenvalue of an Erdős-Rényi hypergraph is:

$$\|T - \mathbb{E}\,T\|_{\text{inj}} = \max_{\|x_1\|_2 = \ldots = \|x_r\|_2 = 1} \sum_{\substack{i_1,\ldots,i_r=1 \\ \text{distinct}}}^{n} (T_{i_1,\ldots,i_r} - p) x_{1,i_1} \ldots x_{r,i_r} \,.$$

In the graph case ($r = 2$), $\|T\|_{\text{inj}}$ and $\|T - \mathbb{E}\,T\|_{\text{inj}}$ are, respectively, the spectral norm of the uncentered and centered adjacency matrices of an Erdős-Rényi random graph. These quantities have been central in the development of spectral graph theory. In sparse random graphs, the emergence of a spectral gap coincides with the existence of efficient algorithms certifying (nearly) tight bounds on large cliques and independent sets; see Section 1.2. The emergence of a spectral gap also marks the region in which the expander mixing lemma holds; see Example 3.3.

1.1 Emergence of a spectral gap in Erdős-Rényi hypergraphs

Motivated by this analogy, we determine the sparsity scale at which the spectral gap of Erdős-Rényi hypergraphs diverges. The upper bound

$$\|T\|_{\text{inj}} - \|T - \mathbb{E}\,T\|_{\text{inj}} \leqslant \|\mathbb{E}\,T\|_{\text{inj}} \leqslant pn^{r/2}$$

shows that $p \gg n^{-r/2}$ is a necessary condition for the spectral gap to diverge. Our first main result shows that, for $r \geqslant 3$, this condition is also sufficient.

Theorem 1.3 (Spectral gap of Erdős-Rényi hypergraphs; consequence of Theorem 3.1)**.** *Let $r \geqslant 3$ be fixed, and let $p = p(n)$ be such that $pn^{r/2} \to \infty$ as $n \to \infty$. Then, with high probability over $T \sim \mathcal{H}_r(n,p)$,*

$$\|T\|_{\text{inj}} - \|T - \mathbb{E}\,T\|_{\text{inj}} = (1 - o(1))pn^{r/2} \qquad \text{as } n \to \infty \,.$$

In the graph case $r = 2$, the matrix Bernstein inequality shows that $p \gg \log n/n$ is sufficient for the spectral gap to diverge. This is also the regime of p studied in the original paper of Friedman and Wigderson on random tensors [FW95]. Theorem 1.3 shows that, unlike in graphs, this regime lies far beyond the point $p \gg n^{-r/2}$ at which a spectral gap first emerges.

Theorem 1.3 overcomes a barrier present in several prior works. Using tensor flattening, Zhou and Zhu [ZZ21, Theorem 2.3] obtain an upper bound on the second eigenvalue throughout the sparsity regime of Theorem 1.3. However, their bound always carries at least a polylogarithmic factor in n, and therefore cannot establish the phase transition. As we explain in Section 1.4, this loss is inherent to the flattening approach. Similar bounds arise in the analyses of algorithms for tensor completion [JO14, Theorem 2.1] and multilayer community detection [LCL20, Theorem 2], but they likewise incur extra $\operatorname{polylog} n$ factors. Section 3.1 discusses consequences of our techniques for these problems.

[1]This terminology is not entirely standard, as one usually subtracts the top eigenvector contribution from T to define the second eigenvalue.

1.2 Motivation: refutation of random constraint satisfaction problems

Theorem 1.3 yields a certificate that upper bounds the size of the maximum independent set in a random hypergraph. For a hypergraph, a subset of vertices S is an independent set if no hyperedge is fully contained in S. If S is an independent set of the hypergraph represented by $T \sim \mathcal{H}_r(n,p)$, the indicator vector $\mathbf{1}_S$ satisfies

$$\frac{\left|\langle T - \mathbb{E}\, T, \mathbf{1}_S^{\otimes r}\rangle\right|}{\|\mathbf{1}_S^{\otimes r}\|_2} = \frac{p|S|(|S|-1)\dots(|S|-r+1)}{|S|^{r/2}} \,.$$

Hence, when $p \gg n^{-r/2}$, Theorem 1.3 implies that such random hypergraphs typically do not contain linear-sized independent sets, and this is certified by the second eigenvalue of T.

Together with known reductions from the problem of refuting random constraint satisfaction problems to that of certifying the absence of linear-sized independent sets in random hypergraphs [FGK05], this suggests that the second eigenvalue could provide an efficient certificate of unsatisfiability for random r-SAT formulas with $O(n^{r/2})$ constraints. For even r, the existence of such certificates already follows from a flattening argument and Grothendieck's inequality [AOW15]. However, for odd r, such certificates were not known until the recent work [dT23], which uses a very different spectral certificate based on the non-backtracking walk matrix and has a significantly more involved analysis.

Our work is motivated by the question of whether the second eigenvalue of sparse random hypergraphs can be computed efficiently. While there is strong evidence that, for dense random tensors ($p = 1/2$), the second eigenvalue can only be approximated within an $\Omega(n^{1/4})$ factor [HSS15, PR22], it is unclear whether such lower bounds extend to the sparse setting. In fact, a simple flattening argument shows that the integrality gap of the basic SDP relaxation at the threshold $p = n^{-r/2}$ is at most $O(\log n)$.

Our proof technique decomposes test vectors into several components, with the aim of clarifying which components might be approximated efficiently. This decomposition is guided by the theory of selector processes, as we discuss further in Section 1.4. We leave the algorithmic aspects of this decomposition to future work. More broadly, it suggests a possible route towards rehabilitating the "Chernoff + union bound" approach for the design of efficient approximation algorithms.

1.3 Concentration of random tensors

On our way to prove Theorem 1.3, we show the following general result on the injective norm of sparse random tensors:

Theorem 1.4 (Concentration of sparse random tensors; see Corollary 4.3)**.** *Let $T \in (\mathbb{R}^n)^{\otimes r}$ be a random tensor with independent entries such that for some $\varepsilon \in (0,1)$,* $\Pr(T_i \neq 0) \leqslant n^{-1-\varepsilon}$ *for all $i \in [n]^r$. Then,*

$$\mathbb{E}\,\|T - \mathbb{E}\,T\|_{\mathrm{inj}} \lesssim \frac{r^5}{\varepsilon^2}\log^2\frac{r}{\varepsilon} \cdot \mathbb{E}\max_{i\in[n]^r}|T_i| \,.$$

Thus, in the sparse regime where each entry is nonzero with probability $\ll 1/n$, the injective norm is controlled, up to a polynomial factor in r, by the largest entry of the tensor. Previous bounds for sparse random tensors, such as [ZZ21], not only depend on n, but also depend exponentially on r. Improving this dependence was explicitly posed there as a direction for future work. We do not expect the particular polynomial dependence on r and ε in Theorem 1.4 to be optimal. We note that even for tensors with i.i.d. Gaussian entries, the precise asymptotic constant in the injective norm was determined only recently [DM24, BS25].

Theorem 1.3 is a special case of Theorem 1.4, where T has i.i.d. entries up to the symmetry. A landmark result on the norm of general random matrices with i.i.d. entries is Seginer's theorem:

Theorem 1.5 ([Seg00, Corollary 2.2])**.** *Let $A \in \mathbb{R}^{n\times m}$ be a random matrix with i.i.d. mean-zero entries. Then,*

$$\mathbb{E}\,\|A\|_{\mathrm{inj}} \asymp \mathbb{E}\max_{i\in[n]}\left(\sum_{j=1}^{m} A_{ij}^2\right)^{1/2} + \mathbb{E}\max_{j\in[m]}\left(\sum_{i=1}^{n} A_{ij}^2\right)^{1/2}.$$

For example, when applied to the centered adjacency matrix of Erdős-Rényi random graphs, Theorem 1.5 implies that the expected second eigenvalue is within a constant factor of the expected square root of the maximum degree.

Remarkably, Theorem 1.5 does not require any assumption on the distribution of the entries. Recently, [LS25] established a generalization of Theorem 1.5 to the $\ell_p \to \ell_q$ operator norm of matrices, under a moment equivalence assumption on the distribution of the entries. Building on Theorem 1.4, we derive a generalization of Theorem 1.5 for the injective norm of tensors, under the following moment equivalence assumption:

Assumption 1.6. Let Z be a mean-zero random variable such that, for some $\varepsilon \in (0,1)$ and $K \geqslant 1$,

$$\mathbb{E}\,|Z|^{2+\varepsilon} \leqslant K\sigma^{2+\varepsilon}, \qquad \text{where } \sigma^2 := \mathbb{E}\,Z^2\,.$$

When the entries of A are i.i.d. copies of a random variable satisfying Assumption 1.6, the maximum row- and column-norm parameter on the right-hand side of Theorem 1.5 simplifies to[2]

$$\mathbb{E}\max_{i\in[n]}\left(\sum_{j=1}^{m} A_{ij}^2\right)^{\frac12} + \mathbb{E}\max_{j\in[m]}\left(\sum_{i=1}^{n} A_{ij}^2\right)^{\frac12} \asymp_{\varepsilon,K} \sigma\sqrt{\max(n,m)} + \mathbb{E}\max_{\substack{i\in[n]\\ j\in[m]}} |A_{ij}|\,.$$

(See the $r = 2$ case of Proposition B.1 for a self-contained proof of this equivalence.)

Our next result is a tensor analogue of this consequence of Seginer's theorem.

Theorem 1.7 (Seginer-type theorem for tensors; see Theorem 5.1)**.** *Let $T \in \mathbb{R}^{n_1\times\cdots\times n_r}$ be a random tensor whose entries are i.i.d. copies of a random variable satisfying Assumption 1.6. Then,*

$$\mathbb{E}\,\|T\|_{\mathrm{inj}} \asymp_{\varepsilon,K,r} \sigma\cdot\max_{i\in[r]}\sqrt{n_i} + \mathbb{E}\max_{\substack{i_1\in[n_1]\\ \cdots\\ i_r\in[n_r]}} |T_{i_1,\dots,i_r}|\,.$$

The dependence on ε, K, r in the upper bound is polynomial, whereas the dependence in the lower bound can be removed entirely by replacing σ with a slightly different quantity; see Theorem 5.1 for a precise statement.

[2] Throughout, we use the notation $a \asymp_{\varepsilon,K} b$ to mean that $cb \leqslant a \leqslant Cb$ for some constants $c = c(\varepsilon,K)$ and $C = C(\varepsilon,K) > 0$ depending only on ε and K. This dependence is necessary in general (see Remark 5.3).

Our work leaves open the question of a full analogue of Theorem 1.5 for tensors whose entries do not satisfy Assumption 1.6:

Conjecture 1.8. *Let $T \in \mathbb{R}^{n_1 \times \dots \times n_r}$ be a random tensor with i.i.d. mean-zero entries. Then,*

$$\mathbb{E}\,\|T\|_{\mathrm{inj}} \asymp_r \max_{k \in [r]} \mathbb{E} \max_{\substack{i_1 \in [n_1] \\ \vdots \\ i_r \in [n_r]}} \left(\sum_{j=1}^{n_k} T^2_{i_1,\dots,i_{k-1},j,i_{k+1},\dots,i_r} \right)^{1/2}.$$

We prove that Conjecture 1.8 is equivalent to Theorem 5.1 under Assumption 1.6 (for constant K, ε independent of n) in Proposition B.1. We note that even the special case of Conjecture 1.8 where $n_1 = \dots = n_r = n$ and the entries have a centered Bernoulli distribution with parameter $1/n$ remains open and would already be of independent interest.

In a related direction, Boedihardjo [Boe24] recently gave a tensor generalization of a different theorem of [Seg00], which holds in the setting of independent (but not necessarily identically distributed) entries; see Theorem 4.1 below. This remarkable result often yields sharp bounds on the norm of random tensors with dense entries, but it is not suited to the sparse setting because its error term involves polylog n factors. Such logarithmic factors are prohibitive for applications such as Theorem 1.3, but they are unavoidable for the general independent entry model even in the matrix case, as shown in [Seg00]. Nevertheless, our proof of Theorem 1.7 crucially relies on the combination of Theorem 1.3 and the results in [Boe24], where Boedihardjo's bound covers the small entries of the tensor and our result controls the outliers.

1.4 Proof idea: an explicit selector process decomposition

The trace method is the classical approach to estimating the spectral norm of a random matrix, but it has no known analogue in the tensor setting. One way to reduce the problem to the matrix case is to flatten the tensor. For instance, given $T \in (\mathbb{R}^n)^{\otimes 4}$, one may bound its injective norm by that of the $n^2 \times n^2$ matrix whose entry in position $((i,j),(k,\ell))$ is $T_{ijk\ell}$. However, this reduction is not sufficient for Theorem 1.3, since the resulting matrix contains a row with $\omega(1)$ nonzero entries when $p = \Omega(n^{-2})$. Moreover, other common tools that have been used recently to study the injective norm of random tensors [Ade25, DM24, DM26, Boe24, BGJ+25, ACS25] are tailored to the Gaussian and sub-Gaussian setting. The sparsity regime we consider lies outside of the Gaussian universality class, so these approaches cannot be applied without major modifications.

Among the standard approaches to the matrix problem, only the union bound argument of Kahn and Szemerédi [FKS89] appears to generalize. The method discretizes the unit sphere by restricting to vectors with signed dyadic entries, which changes the supremum by at most a universal constant factor. The main step is then a technical case analysis establishing the desired union bound for this discretization. Although the approach uses only first principles, its implementation typically requires a complex and lengthy case analysis. This is already apparent in the matrix setting and becomes particularly cumbersome for tensors [FKS89, FW95, FO05, JO14, HKP21, ZZ21]. This motivates the search for a more systematic and conceptual way to design such arguments.

We instead use the theory of selector processes [Tal21, Section 11.11] as a guide for constructing union bound arguments. Given $\mathcal{X} \subseteq \mathbb{R}^m$, let

$$\delta(\mathcal{X}) := \mathbb{E} \sup_{X \in \mathcal{X}} \langle T - \mathbb{E}\, T, X \rangle \,, \tag{1}$$

where $(T_i)_{i\in[m]}$ are i.i.d. Bernoulli random variables. Stochastic processes such as $(\langle T - \mathbb{E}\, T, X \rangle)_{X\in\mathcal{X}}$ are called *selector processes*. Talagrand shows [Tal21, Theorem 11.12.1] that the following strategy gives an upper bound on $\delta(\mathcal{X})$ that is optimal up to universal constants:

1. Find a decomposition $\mathcal{X} \subseteq \mathcal{X}_B + \mathcal{X}_P$, where we refer to $\mathcal{X}_B$ as the "Bernstein part" and $\mathcal{X}_P$ as the "positive part".

2. Bound $\delta(\mathcal{X}_B)$ by a chaining argument based only on Bernstein's inequality.

3. Bound the positive part $\delta(\mathcal{X}_P)$ after discarding sign cancellations:

   $$\delta(\mathcal{X}_P) \leqslant \mathbb{E} \sup_{X\in\mathcal{X}_P} \langle T, X \rangle \,.$$ [3]

   Moreover, [Tal21, Research Problem 13.1.1] further suggests the existence of a finite witness set $\mathcal{W} \subseteq \mathbb{R}^m_{\geqslant 0}$ such that

   $$\forall X \in \mathcal{X}_P \,,\ \exists Y \in \mathrm{Conv}(\mathcal{W}) \text{ such that } X_i \leqslant Y_i \text{ for all } i \in [m] \,,$$

   (where Conv denotes the convex hull), and for which a direct union bound over $\mathcal{W}$ yields a tight estimate:

   $$\delta(\mathcal{X}_P) \asymp \inf \left\{ t > 0 : \int_t^\infty \sum_{W\in\mathcal{W}} \Pr\left(\langle T, W \rangle > s\right) \mathrm{d}s \leqslant t \right\} .$$

Such a decomposition immediately gives an upper bound on $\delta(\mathcal{X})$. The nontrivial fact is that there always exists a decomposition providing a bound on $\delta(\mathcal{X})$ that is tight up to constant factors. We emphasize that, to the best of our knowledge, the additional assertion concerning the existence of a witness set remains a (far-reaching) conjecture (see [PP24] for recent progress on a related question).

Unfortunately, Talagrand's proof of existence of such a decomposition does not explain how to choose the decomposition in practice. This limitation is shared by much of the chaining literature. For instance, it is known abstractly that the supremum of a Gaussian process can always be bounded optimally using a generic chaining construction. Nevertheless, for most Gaussian processes arising in random matrix theory applications it remains open to exhibit an explicit construction with this property [Tal21, vH18, Lee23]. A partial exception is the recent work [Lat24], which relies on the decomposition theorem for Bernoulli processes proved in [BL14] as a guide for the analysis of the norm of Rademacher matrices.

In this paper, we implement this strategy explicitly for our selector process (1). We prove Theorems 1.3 and 1.4 by constructing a decomposition together with a witness set for the positive part. Our goal is to bound the expected supremum of the selector process over

$$\mathcal{X} := \{x_1 \otimes \ldots \otimes x_r : \|x_1\|_2 = \ldots = \|x_r\|_2 = 1\} \,. \tag{2}$$

[3] Assuming that $\mathcal{X}_P = -\mathcal{X}_P$, as will be the case in our application.

To define the Bernstein part, we clip every entry of a test tensor $X \in \mathcal{X}$ at the level $1/(n \log n)$. The remaining contribution forms the positive part; see (4) for the formal decomposition. This decomposition plays the role of the small-pair/large-pair decomposition in the traditional Kahn-Szemerédi argument.

The witness set consists of suitably normalized pure tensors whose factors are $\{0, 1\}$-valued vectors; see Definition 3.8. The general idea of transferring bounds from $\{0, 1\}$-valued test vectors to arbitrary test vectors also appears in the ad hoc discretization arguments [FKS89, FO05]. Such a lifting generally requires a carefully chosen normalization of the indicator vectors, because a naive reduction incurs a loss depending on the largest ℓ_1-norm of a tensor slice; see [BL06, Section 4.2] for the matrix case. Identifying the appropriate normalization is the main substantive step of our proof; once this is in place, the rest of the proof follows directly from the above strategy.

1.5 Organization of the paper

After the preliminaries in Section 2, we prove the general sparse tensor result underlying Theorem 1.3 (Theorem 3.1) in Section 3. In particular, we discuss some applications of this result in Section 3.1. We derive some extensions, including Theorem 1.4, in Section 4. In Section 5, we prove our generalization of Seginer's theorem (Theorem 1.7). Appendix A contains omitted lemmas and their proofs. Appendix B shows the equivalence between Conjecture 1.8 and Theorem 1.7 under our moment equivalence assumption.

1.6 Acknowledgments

LP thanks Luca Trevisan for suggesting, several years ago, the problem of determining the spectral gap of Erdős-Rényi hypergraphs. LP's work on this project was supported by the Swiss National Science Foundation, grant no. 10004947.

The authors used AI tools for literature research and to scan the manuscript for typos. All content was written and verified by the authors.

2 Preliminaries

Asymptotic notation: We write $a \lesssim b$ (resp. $a \gtrsim b$) to denote $a \leqslant Cb$ (resp. $a \geqslant Cb$) for some universal constant $C > 0$. We write $a \asymp b$ as a shortcut for $a \lesssim b$ and $a \gtrsim b$. We write $a \lesssim_s b$ when the hidden constant C is allowed to depend on s.

Tensor notation: For tensors $T, T' \in \mathbb{R}^{n_1 \times \ldots n_r}$, write $T \leqslant T'$ when $T_i \leqslant T'_i$ for every index $i \in [n_1] \times \ldots \times [n_r]$. Denote by $|T|$ the tensor obtained from T by applying the absolute value entrywise. For $p \in [1, \infty]$, write $\|T\|_p$ for the ℓ_p-norm of the vectorization of T, viewed as an element of $\mathbb{R}^N$ for $N = \prod_{i=1}^r n_i$.

Definition 2.1 (Clip)**.** Given $\tau \in \mathbb{R}_{\geqslant 0}$, let $\mathrm{clip}_\tau : \mathbb{R} \to \mathbb{R}$ be defined by

$$\mathrm{clip}_\tau(x) = \begin{cases} x & \text{if } |x| \leqslant \tau, \\ \tau & \text{if } x > \tau, \\ -\tau & \text{if } x < -\tau. \end{cases}$$

For a tensor $T \in \mathbb{R}^{n_1 \times \cdots n_r}$, we extend the notation by letting $\mathrm{clip}_\tau(T)$ be the tensor obtained from T by applying clip_τ entrywise.

Concentration inequalities: We use only two elementary concentration inequalities: Bernstein's inequality and the multiplicative Chernoff bound.

Lemma 2.2 (Bernstein's inequality; see, e.g., [Tal21, Lemma 4.5.6])**.** *Let $X_1, \ldots, X_N$ be independent mean-zero random variables such that $|X_i| \leqslant M$ almost surely. Then for any $t \geqslant 0$,*

$$\Pr\left(\sum_{i=1}^{N} X_i > t\right) \leqslant \exp\left(-\frac{1}{4}\min\left(\frac{t^2}{\sum_{i=1}^{N} \mathbb{E}\, X_i^2}, \frac{t}{M}\right)\right) .$$

Let $\mathrm{Ber}(p)$ denote the Bernoulli distribution with parameter $p \in (0,1)$.

Lemma 2.3 (Chernoff bound; see, e.g., [Ver26, Theorem 2.3.1])**.** *Let $X_1, \ldots, X_N$ be i.i.d. $\mathrm{Ber}(p)$ random variables. Then, for any $t \geqslant e^2 \cdot Np$,*

$$\Pr\left(\sum_{i=1}^{N} X_i > t\right) \leqslant \exp\left(-\frac{t}{2}\log\left(\frac{t}{Np}\right)\right) .$$

Solid convex hull: Given $\mathcal{X} \subseteq (\mathbb{R}^n)^{\otimes r}$, we denote by $\mathrm{Conv}(\mathcal{X})$ the convex hull of $\mathcal{X}$ and by

$$\mathrm{solid}(\mathcal{X}) := \{X \in (\mathbb{R}^n)^{\otimes r} : \exists Y \in \mathrm{Conv}(\mathcal{X}) \text{ s.t. } X \leqslant Y\}$$

its solid convex hull [Tal21].

3 Tensors with sparse independent entries

In this section, we prove a general bound on the norm of sparse random tensors. Section 3.1 presents applications of the theorem to the settings introduced above. We then prove the theorem in Sections 3.2 through 3.5.

Theorem 3.1 (Bound on the injective norm of sparse random tensors)**.** *Let $n \geqslant 5$, $r \geqslant 2$ and $\varepsilon \in (0,1)$ such that*

$$r \leqslant \frac{\varepsilon \log n}{5 \log\log n} . \tag{3}$$

Let $T \in (\mathbb{R}^n)^{\otimes r}$ be a random tensor with independent entries taking values in $[-1,1]$, and suppose that for some $L \geqslant 1$, we have $\Pr(T_i \neq 0) \leqslant L n^{-1-\varepsilon}$ for all $i \in [n]^r$. Then,

$$\mathbb{E}\, \|T - \mathbb{E}\, T\|_{\mathrm{inj}} \lesssim \frac{L r^4}{\varepsilon} .$$

Moreover, for any $t > 0$,

$$\Pr\left(\|T - \mathbb{E}\, T\|_{\mathrm{inj}} \gtrsim \frac{L r^3}{\varepsilon}(r+t)\right) \leqslant n^{-t} .$$

3.1 Applications of Theorem 3.1

Before proving Theorem 3.1, we give several applications and comparisons with prior work. First, if each entry of T is distributed as $\mathrm{Ber}(n^{-1-\varepsilon})$, Theorem 3.1 recovers Theorem 1.3 as a corollary.

Proof of Theorem 1.3 from Theorem 3.1. Let $T \sim \mathcal{H}_r(n,p)$. The upper bound

$$\|T\|_{\mathrm{inj}} - \|T - \mathbb{E}\,T\|_{\mathrm{inj}} \leqslant \|\mathbb{E}\,T\|_{\mathrm{inj}} \leqslant pn^{r/2}$$

is immediate. For the lower bound, first, applying the all-ones test vector shows that

$$\|T\|_{\mathrm{inj}} \geqslant \frac{1}{n^{r/2}} \sum_{i \in [n]^r} T_i \,.$$

By Lemma 2.2, the right-hand side is at least $(1-o(1))pn^{r/2}$ with high probability. It remains to apply Theorem 3.1 to $\|T - \mathbb{E}\,T\|_{\mathrm{inj}}$.

Although the entries of an Erdős-Rényi adjacency tensor are not independent (because of the symmetry), we can partition its entries according to the ordering of their indices, writing the tensor as a sum of $r!$ tensors with independent entries. Since r is fixed, (3) holds for sufficiently large n. Therefore, setting $L = \max(1, pn^{1.01})$, we can apply Theorem 3.1 and we obtain, with high probability,

$$\|T - \mathbb{E}\,T\|_{\mathrm{inj}} \lesssim_r L = o(pn^{r/2})\,,$$

whenever $r \geqslant 3$ and $pn^{r/2} \to \infty$. This concludes the proof. $\square$

Remark 3.2. The high probability bound also shows that the conclusion of Theorem 1.3 still holds for the variant of the Erdős-Rényi model in which m hyperedges are selected uniformly at random. This follows by conditioning the model in Definition 1.1 on having exactly $m = p\binom{n}{r}$ hyperedges, since this event has probability at least $\Omega(m^{-1/2})$.

Example 3.3 (Expander mixing lemma)**.** Applying the injective norm bound to $\{0,1\}$-valued test vectors directly yields an improved expander mixing lemma for sparse Erdős-Rényi hypergraphs. More precisely, for every $\varepsilon \in (0,1)$, with high probability over $T \sim \mathcal{H}_r(n,p)$, the following holds simultaneously for every collection of pairwise disjoint sets $U_1, \ldots, U_r \subseteq [n]$:

$$\left| e_T(U_1, \ldots, U_r) - p \prod_{i=1}^{r} |U_i| \right| \lesssim_{r,\varepsilon} (1 + pn^{1+\varepsilon}) \prod_{i=1}^{r} \sqrt{|U_i|}\,,$$

where $e_T(U_1, \ldots, U_r)$ denotes the number of hyperedges containing exactly one vertex from each of $U_1, \ldots, U_r$.

Example 3.4 (Tensor completion)**.** Given a deterministic tensor $T \in (\mathbb{R}^n)^{\otimes 3}$, we form a random set $\Omega \subseteq [n]^3$ of entries by including each entry independently with probability $p := \alpha n^{-3/2}$ for some $\alpha \geqslant 1$. Then, we reveal the entries of T in Ω. Denote by $P_\Omega(T)$ the observed tensor (i.e., $P_\Omega(T)_i = T_i$ if $i \in \Omega$, and $P_\Omega(T)_i = 0$ otherwise). Note that $\frac{1}{p} P_\Omega(T)$ is an unbiased estimator of T. The question is how accurately this estimator approximates T in injective norm.

This question arises, for example, in the analysis of algorithms for tensor completion, where the goal is to reconstruct T from $P_\Omega(T)$. In their analysis of the alternating minimization algorithm for this task, Jain and Oh [JO14, Appendix A] prove the estimate

$$\frac{1}{\|T\|_\infty n^{3/2}} \mathbb{E}\left\|\frac{1}{p}P_\Omega(T) - T\right\|_{\text{inj}} \lesssim \frac{\log^2 n}{\sqrt{\alpha}}\,.$$

In this setting, Theorem 3.1 implies that if $p \leqslant n^{-1.01}$, then

$$\frac{1}{\|T\|_\infty n^{3/2}} \mathbb{E}\left\|\frac{1}{p}P_\Omega(T) - T\right\|_{\text{inj}} \lesssim \frac{1}{\alpha}\,.$$

Example 3.5 (Multilayer community detection)**.** In the stochastic block model, one considers n individuals and an unknown community assignment $\sigma : [n] \to [k]$ of the individuals to k communities. In the multilayer version of the problem, we observe L undirected graphs (layers) $G_1, \ldots, G_L$ on the same vertex set $[n]$. In G_ℓ, an edge between u and v appears with probability $p_\ell(\sigma(u), \sigma(v))$, independently over all $\ell \in [L]$ and all $1 \leqslant u < v \leqslant n$. Given $G_1, \ldots, G_L$, the goal is to recover σ (up to a permutation of the community labels).

The observations can be collected in an order-3 tensor $T \in \mathbb{R}^{L\times n\times n}$, where $T_{\ell uv} = 1$ if $\{u, v\}$ is an edge in G_ℓ, and $T_{\ell uv} = 0$ otherwise. Lei, Chen, and Lynch [LCL20] analyze a least-squares estimator for this problem. A key ingredient in their analysis is an estimate for the concentration of T around its expectation. When $L \leqslant n$, they show [LCL20, Theorem 2] that

$$\|T - \mathbb{E}\,T\|_{\text{inj}} \lesssim \log^{3/2} n + \sqrt{n\rho}\log n\,, \qquad \rho := \max_{\ell\in[L],a,b\in[k]} p_\ell(a, b)\,.$$

In contrast, if $\rho \leqslant n^{-1.01}$, then Theorem 3.1 gives the stronger bound $\|T - \mathbb{E}\,T\|_{\text{inj}} \leqslant O(1)$ with high probability.

3.2 Proof of Theorem 3.1

Recall that $\mathcal{X}$ denotes the set of unit pure tensors in (2). We decompose $\mathcal{X}$ into

$$\mathcal{X}_B := \{\text{clip}_\tau(X) : X \in \mathcal{X}\} \text{ and } \mathcal{X}_P := \{X - \text{clip}_\tau(X) : X \in \mathcal{X}\}\,, \tag{4}$$

for the threshold $\tau := \frac{1}{n \log n}$. Clearly, we have $\mathcal{X} \subseteq \mathcal{X}_P + \mathcal{X}_B$ by construction. As in Section 1.4, we refer to $\mathcal{X}_B$ as the "Bernstein part" of the process and to $\mathcal{X}_P$ as its "positive part". The following lemmas bound the two suprema separately.

Lemma 3.6 (Bounding the Bernstein part)**.** *Let r, n, ε, L, and $T \in (\mathbb{R}^n)^{\otimes r}$ satisfy the assumptions of Theorem 3.1. Then, for any $t > 0$,*

$$\Pr\left(\sup_{X\in\mathcal{X}_B} \langle T - \mathbb{E}\,T, X\rangle \gtrsim \sqrt{L}(r^2 + t)\right) \leqslant e^{-tn\log n}\,.$$

Lemma 3.7 (Bounding the positive part)**.** *Let $T \in (\mathbb{R}^n)^{\otimes r}$ be a random tensor with i.i.d.* $\text{Ber}(Ln^{-1-\varepsilon})$ *entries, where r, n, ε, L satisfy the assumptions of Theorem 3.1. Then, for any $t > 0$,*

$$\Pr\left(\sup_{X\in\mathcal{X}_P} \langle T, X\rangle \gtrsim \frac{Lr^3}{\varepsilon}(r + t)\right) \leqslant n^{-t}\,.$$

Proof of Theorem 3.1 from Lemmas 3.6 and 3.7. Lemma 3.6 directly bounds the supremum of the Bernstein part of the original process. For the positive part, note that every $X \in \mathcal{X}_P$ has at most $n^2 \log^2 n$ nonzero entries. Therefore

$$|\langle \mathbb{E}\, T, X\rangle| \leqslant \|\mathbb{E}\, T\|_\infty \|X\|_1 \leqslant Ln^{-1-\varepsilon} n \log n = Ln^{-\varepsilon}\log n \overset{(3)}{\leqslant} \frac{Lr^4}{\varepsilon}\,,$$

so it suffices to bound the uncentered process $\sup_{X\in\mathcal{X}_P}\langle T, X\rangle$. Since $\mathcal{X}_P = -\mathcal{X}_P$, we may furthermore assume, without loss of generality, that the entries of T are supported on $[0,1]$. Among all distributions on $[0,1]$ satisfying the assumptions of Theorem 3.1, the probability $\Pr\left(\sup_{X\in\mathcal{X}_P}\langle T, X\rangle > t\right)$ is maximized when the entries of T are distributed as $\mathrm{Ber}(Ln^{-1-\varepsilon})$ (for every t). Hence, the tail bound in Theorem 3.1 follows from Lemmas 3.6 and 3.7, and integrating this tail bound yields the expectation bound. □

The rest of the section is devoted to the proof of Lemmas 3.6 and 3.7.

3.3 Bernstein part (proof of Lemma 3.6)

First, for every fixed $X \in \mathcal{X}_B$, every $C \geqslant 1$, and every $t > 0$, Bernstein's inequality (Lemma 2.2) gives

$$\begin{aligned}\Pr\left(\langle T - \mathbb{E}\, T, X\rangle > C\sqrt{L}(r^2+t)\right) &\leqslant \exp\left(-\frac{1}{4}\min\left\{\frac{C^2(r^2+t)^2}{n^{-1-\varepsilon}}, \frac{C\sqrt{L}(r^2+t)}{2\tau}\right\}\right)\\ &\overset{(3)}{\leqslant} e^{-\frac{C}{8}(r^2+t)n\log n}\,.\end{aligned}$$

Next, we show that a net of $2^{O(r^2 n\log n)}$ test vectors suffices to control the supremum of the process over $\mathcal{X}_B$. Indeed, for any unit vectors $x_1,\dots,x_r$ and $y_1,\dots,y_r$, letting $X := \bigotimes_{i\in[r]} x_i$ and $Y := \bigotimes_{i\in[r]} y_i$,

$$\begin{aligned}|\langle T-\mathbb{E}\, T, \mathrm{clip}_\tau(X) - \mathrm{clip}_\tau(Y)\rangle| &\leqslant \|T-\mathbb{E}\, T\|_2\|\mathrm{clip}_\tau(X)-\mathrm{clip}_\tau(Y)\|_2 && \text{(Cauchy-Schwarz)}\\ &\leqslant \|T-\mathbb{E}\, T\|_2\|X-Y\|_2 && (\mathrm{clip}_\tau \text{ is 1-Lipschitz})\\ &\leqslant 2n^{\frac{r}{2}}\|X-Y\|_2 && (|T_i|\leqslant 1)\\ &\leqslant 2rn^{\frac{r}{2}}\max_{i\in[r]}\|x_i-y_i\|_2\,. && \text{(triangle inequality and } x_i, y_i \text{ unit)}\end{aligned}$$

Let $\mathcal{N}$ be a $\frac{1}{2rn^{r/2}}$-net of the unit sphere of $\mathbb{R}^n$; we may choose it so that $|\mathcal{N}| \leqslant O(rn^{r/2})^n$ using standard constructions. Define $\mathcal{S} := \{\mathrm{clip}_\tau(y_1\otimes\ldots\otimes y_r) : y_1,\dots,y_r\in\mathcal{N}\}$. By the above inequality, for any $X\in\mathcal{X}_B$, there exists $Y\in\mathcal{S}$ such that

$$|\langle T-\mathbb{E}\, T, X-Y\rangle| \leqslant 1 \leqslant r^2\sqrt{L}\,.$$

Moreover, since $|\mathcal{S}| \leqslant O(rn^{r/2})^{nr} = 2^{O(r^2 n\log n)}$, the above tail bound holds simultaneously for all vectors in the net, provided that C is sufficiently large. Lemma 3.6 then follows.

3.4 Positive part: witness set construction

We begin by constructing a witness set for $\mathcal{X}_P$ (see Section 1.4 for background). We use the following normalization for the indicator $\mathbf{1}_S$ of a non-empty set $S \subseteq [n]$:

$$\phi(S) := \frac{\sqrt{|S|}}{\max(4r^2, \log^2 |S|)} \,.$$

We will repeatedly use the following elementary consequence of (3) and $n \geqslant 5$:

$$\phi(S) \geqslant \frac{\sqrt{|S|}}{\log^2 n} \,. \tag{5}$$

Definition 3.8 (Witness set for $\mathcal{X}_P$)**.** Define the witness set $\mathcal{W} = \mathcal{W}_0 \cup \ldots \cup \mathcal{W}_r$ as follows. First, let $\mathcal{W}_0$ be the set of balanced indicator witnesses:

$$\mathcal{W}_0 := \left\{ 2 \bigotimes_{i \in [r]} \frac{\mathbf{1}_{A_i}}{\phi(A_i)} : A_1, \ldots, A_r \text{ s.t. } \frac{1}{r} \sum_{i=1}^{r} |A_i| \leqslant \prod_{i=1}^{r} \phi(A_i) \text{ and } \prod_{i=1}^{r} |A_i| \leqslant r^{2r} n^2 \log^2 n \right\} ,$$

where $A_1, \ldots, A_r$ range over all non-empty subsets of $[n]$ satisfying the conditions. Second, for each $j \in [r]$, let $\mathcal{W}_j$ be the set of unbalanced indicator witnesses whose j-th coordinate is unconstrained:

$$\mathcal{W}_j := \left\{ \frac{\bigotimes_{i=1}^{j-1} \mathbf{1}_{A_i} \otimes \mathbf{1}_{[n]} \otimes \bigotimes_{i=j+1}^{r} \mathbf{1}_{A_i}}{\frac{1}{24r^4} \sum_{\substack{i=1 \\ i \neq j}}^{r} |A_i| + \frac{1}{4r \log^4 n} \prod_{\substack{i=1 \\ i \neq j}}^{r} \phi(A_i)^2} : A_1, \ldots, A_{j-1}, A_{j+1}, \ldots, A_r \subseteq [n] \text{ non-empty} \right\} .$$

We show that $\mathcal{W}$ is a valid witness set for $\mathcal{X}_P$.

Lemma 3.9. *We have*

$$\mathcal{X}_P \subseteq \text{solid} \left(\bigcup_{j=0}^{r} \mathcal{W}_j \right) .$$

Proof. It suffices to consider vectors with non-negative entries. Let $x_1, \ldots, x_r \in \mathbb{R}^n_{\geqslant 0}$ be unit vectors. Given $k = (k_1, \ldots, k_r) \in \mathbb{N}^r$ and $i \in [r]$, write $A_{k,i} := \{j \in [n] : (x_i)_j \in [r^{-k_i - 1}, r^{-k_i}]\}$. Since each x_i is a unit vector,

$$\sum_{k \geqslant 0} |A_{k,i}| r^{-2k_i - 2} \leqslant 1 \implies |A_{k,i}| \leqslant r^{2k_i + 2} \,. \tag{6}$$

Taking tensor products of this level set decomposition gives

$$\bigotimes_{i \in [r]} x_i - \text{clip}_\tau \left(\bigotimes_{i \in [r]} x_i \right) \leqslant \sum_{\substack{k_1, \ldots, k_r \geqslant 0 \\ \prod_{i \in [r]} r^{-k_i} > \tau}} \bigotimes_{i \in [r]} \frac{\mathbf{1}_{A_{k,i}}}{r^{k_i}} \,. \tag{7}$$

We cover the indices of the sum on the right-hand side of (7) by

$$\mathcal{K}_0 = \left\{ (k_1, \dots, k_r) \in \mathbb{N}^r : \frac{1}{r} \sum_{i=1}^{r} |A_{k,i}| \leqslant \prod_{i=1}^{r} \phi(A_{k,i}) \text{ and } \prod_{i=1}^{r} |A_{k,i}| \leqslant r^{2r} n^2 \log^2 n \right\},$$

$$\mathcal{K}_j = \left\{ (k_1, \dots, k_r) \in \mathbb{N}^r : \frac{1}{r} \sum_{i=1}^{r} |A_{k,i}| > \prod_{i=1}^{r} \phi(A_{k,i}) \text{ and } |A_{k,j}| = \max_{i \in [r]} |A_{k,i}| \right\} \quad \forall j \in [r] .$$

First, we claim that the sets $\mathcal{K}_0, \dots, \mathcal{K}_r$ cover every index appearing in the sum on the right-hand side of (7). Indeed, by (6), we have

$$\prod_{i \in [r]} r^{-k_i} > \tau \implies \prod_{i=1}^{r} |A_{k,i}| \leqslant r^{2r} \prod_{i=1}^{r} r^{2k_i} \leqslant \frac{r^{2r}}{\tau^2} = r^{2r} n^2 \log^2 n ,$$

so the second condition in the definition of $\mathcal{K}_0$ is automatically satisfied.

Second, we show that, for each $j \in \{0, \dots, r\}$, the sum over $\mathcal{K}_j$ lies in the solid convex hull of the corresponding $\mathcal{W}_j$.

Lemma 3.10. *We have*

$$\sum_{k \in \mathcal{K}_0} \bigotimes_{i \in [r]} \frac{\mathbf{1}_{A_{k,i}}}{r^{k_i}} \in \frac{1}{2} \operatorname{solid}(\mathcal{W}_0) .$$

Lemma 3.11. *For any $j \in [r]$, we have*

$$\sum_{k \in \mathcal{K}_j} \bigotimes_{i \in [r]} \frac{\mathbf{1}_{A_{k,i}}}{r^{k_i}} \in \frac{1}{2r} \operatorname{solid}(\mathcal{W}_j) .$$

Lemmas 3.10 and 3.11 complete the proof of Lemma 3.9. □

Proof of Lemma 3.10. After an appropriate rescaling, each tensor $\bigotimes_i \mathbf{1}_{A_{k,i}}$ appearing in the sum is an element of $\mathcal{W}_0$. Thus, we can expand the sum as a linear combination of elements of $\mathcal{W}_0$, with total coefficient

$$\frac{1}{2} \sum_{k \in \mathcal{K}_0} \prod_{i=1}^{r} r^{-k_i} \phi(A_{k,i}) \leqslant \frac{1}{2} \sum_{k_1, \dots, k_r \geqslant 0} \prod_{i=1}^{r} \frac{r}{\max(4r^2, (2k_i + 2)^2 \log^2 r)} ,$$

where we used (6) and the fact that $x \mapsto \frac{\sqrt{x}}{\max(4r^2, \log^2 x)}$ is increasing on $[1, \infty)$. The sum factorizes, yielding

$$\leqslant \frac{1}{2} \left(\sum_{k \geqslant 1} \min \left(\frac{1}{4r}, \frac{r}{4k^2 \log^2 r} \right) \right)^r .$$

Let $k^* := \frac{r}{\log r}$ denote the crossover value at which the second term achieves the minimum. By a sum-integral comparison,

$$\sum_{k \geqslant 1} \min \left(\frac{1}{4r}, \frac{r}{4 \log^2 r} \cdot \frac{1}{k^2} \right) \leqslant \frac{k^*}{4r} + \frac{r}{4 \log^2 r} \int_{k^*}^{\infty} \frac{1}{x^2} \, \mathrm{d}x = \frac{1}{2 \log r} < 1 ,$$

since $r \geqslant 2$. Hence, the sum of the coefficients is at most $1/2$, concluding the argument. □

Proof of Lemma 3.11. Without loss of generality, assume $j = 1$. As in Lemma 3.10, we bound the total coefficient in the natural expansion of the left-hand side in terms of vectors in $\mathcal{W}_1$, which is

$$\sum_{k \in \mathcal{K}_1} \prod_{i=1}^{r} r^{-k_i} \left(\frac{1}{24r^4} \sum_{i=2}^{r} |A_{k,i}| + \frac{1}{4r \log^4 n} \prod_{i=2}^{r} \phi(A_{k,i})^2 \right).$$

For the second term, since $\log^4 n \geqslant \max(4r^2, \log^2 |A_{k,1}|)^2$ by (3), we have

$$\frac{1}{\log^4 n} \prod_{i=2}^{r} \phi(A_{k,i})^2 \leqslant \frac{\prod_{i=1}^{r} \phi(A_{k,i})^2}{|A_{k,1}|} \leqslant \frac{\prod_{i=1}^{r} \phi(A_{k,i}) \cdot \frac{1}{r} \sum_{i=1}^{r} |A_{k,i}|}{|A_{k,1}|} \leqslant \prod_{i=1}^{r} \phi(A_{k,i}).$$

Therefore, using the same argument as in Lemma 3.10,

$$\frac{1}{4r \log^4 n} \sum_{k \in \mathcal{K}_1} \prod_{i=1}^{r} r^{-k_i} \prod_{i=2}^{r} \phi(A_{k,i})^2 \leqslant \frac{1}{4r} \sum_{k_1, \dots k_r \geqslant 0} \prod_{i=1}^{r} r^{-k_i} \phi(A_{k,i}) < \frac{1}{4r}.$$

It remains to bound the first term,

$$\begin{aligned} \sum_{k \in \mathcal{K}_1} \prod_{i=1}^{r} r^{-k_i} \sum_{i=2}^{r} |A_{k,i}| &\leqslant \sum_{k_1 \geqslant 0} r^{-k_1} \sum_{i=2}^{r} \sum_{\substack{k_i \geqslant 0 \\ |A_{k,i}| \leqslant |A_{k,1}|}} |A_{k,i}| r^{-k_i} \prod_{\substack{\ell=2 \\ \ell \neq i}}^{r} \sum_{k_\ell \geqslant 0} r^{-k_\ell} \\ &\leqslant r \left(\frac{1}{1 - \frac{1}{r}} \right)^{r-2} \max_{i=2,\dots,r} \sum_{\substack{k_1, k_i \geqslant 0 \\ |A_{k,i}| \leqslant |A_{k,1}|}} r^{-k_1 - k_i} |A_{k,i}|. \end{aligned}$$

Finally, for each $i \in \{2, \dots, r\}$, split the sum according to whether $k_i \geqslant k_1$ or $k_1 \geqslant k_i$:

$$\begin{aligned} \sum_{k_1, k_i \geqslant 0} r^{-k_1 - k_i} \min(|A_{k,1}|, |A_{k,i}|) &\leqslant \sum_{k_1 \geqslant 0} r^{-2k_1} |A_{k,1}| \sum_{k_i \geqslant k_1} r^{-(k_i - k_1)} + \sum_{k_i \geqslant 0} r^{-2k_i} |A_{k,i}| \sum_{k_1 \geqslant k_i} r^{-(k_1 - k_i)} \\ &\leqslant 2 \cdot r^2 \cdot \frac{1}{1 - \frac{1}{r}}. \end{aligned}$$

Using $\left(\frac{r}{r-1} \right)^{r-1} \leqslant 3$, we get that the sum of the coefficients is at most $\frac{1}{2r}$, as desired. □

3.5 Positive part: union bound (proof of Lemma 3.7)

Having constructed a witness set for $\mathcal{X}_P$, it remains to apply a union bound over the witnesses.

Lemma 3.12. *Let T be as in Lemma 3.7. For all $t \geqslant 1$,*

$$\Pr\left(\max_{W \in \mathcal{W}_0} \langle T, W \rangle \gtrsim \frac{tLr}{\varepsilon} \right) \leqslant n^{-t}.$$

Proof. Let $A_1, \dots, A_r \subseteq [n]$ satisfy the conditions in the definition of $\mathcal{W}_0$. We apply the Chernoff bound (Lemma 2.3): for all sufficiently large t,

$$\Pr\left(\left\langle T, \bigotimes_{i \in [r]} \frac{\mathbf{1}_{A_i}}{\phi(A_i)} \right\rangle > \frac{tLr}{\varepsilon} \right) \leqslant \exp\left(-\frac{tr \prod_{i=1}^{r} \phi(A_i)}{2\varepsilon} \log\left(\frac{tLr \prod_{i=1}^{r} \phi(A_i)}{\varepsilon n^{-1-\varepsilon} \prod_{i=1}^{r} |A_i|} \right) \right).$$

Note that the assumption of Lemma 2.3 is verified, as

$$\frac{tLr\prod_{i=1}^{r}\phi(A_i)}{\varepsilon n^{-1-\varepsilon}\prod_{i=1}^{r}|A_i|} \overset{(5)}{\geqslant} \frac{trn^{1+\varepsilon}}{\varepsilon}\cdot\frac{1}{\log^{2r} n\prod_{i=1}^{r}\sqrt{|A_i|}} \geqslant tn^{\varepsilon}\cdot\frac{1}{r^r\log^{2r+1} n} \overset{(3)}{\geqslant} tn^{\frac{3\varepsilon}{10}}\,.$$

Therefore, the assumption holds provided that t is sufficiently large. Next, using the assumptions on A_i, we can simplify the upper bound to

$$\begin{aligned}\Pr\left(\left\langle T, \bigotimes_{i\in[r]}\frac{\mathbf{1}_{A_i}}{\phi(A_i)}\right\rangle > \frac{tLr}{\varepsilon}\right) &\leqslant \exp\left(-\frac{t\sum_{i=1}^{r}|A_i|}{2\varepsilon}\log\left(tn^{\frac{3\varepsilon}{10}}\right)\right)\\ &\leqslant \exp\left(-\frac{3t}{20}\sum_{i=1}^{r}|A_i|\log n\right)\,,\end{aligned}$$

using the same estimate as above. Finally, the number of witnesses in $\mathcal{W}_0$ corresponding to a prescribed r-tuple of set sizes $|A_1|,\ldots,|A_r|$ is at most $\prod_{i=1}^{r}\binom{n}{|A_i|}$, so we can take a union bound and use Lemma A.1 provided that t is large enough. Increasing the implicit universal constant extends the result to all $t\geqslant 1$. □

Lemma 3.13. *Let $j\in[r]$ and T be as in Lemma 3.7. Then, for any $t>0$,*

$$\Pr\left(\max_{W\in\mathcal{W}_j}\langle T, W\rangle \gtrsim \frac{Lr^3}{\varepsilon}(r+t)\right) \leqslant n^{-t}\,.$$

Proof. Without loss of generality, assume $j=1$. Let $A_2,\ldots,A_r\subseteq[n]$ and denote by

$$\Delta := \frac{1}{24r^4}\sum_{i=2}^{r}|A_i| + \frac{1}{4r\log^4 n}\prod_{i=2}^{r}\phi(A_i)^2\,.$$

Then, by the Chernoff bound (Lemma 2.3), we have, for any $t>0$ and for a sufficiently large universal constant C,

$$\Pr\left(\left\langle T, \frac{\mathbf{1}_{[n]}\otimes\bigotimes_{i=2}^{r}\mathbf{1}_{A_i}}{\Delta}\right\rangle \geqslant C\frac{Lr^3}{\varepsilon}(r+t)\right) \leqslant \exp\left(-\frac{Cr^3(r+t)\Delta}{2\varepsilon}\log\left(\frac{C\Delta Lr^3(r+t)}{\varepsilon n^{-\varepsilon}\prod_{i=2}^{r}|A_i|}\right)\right)\,.$$

Note that the assumption of Lemma 2.3 is verified, as

$$\frac{C\Delta Lr^3(r+t)}{\varepsilon n^{-\varepsilon}\prod_{i=2}^{r}|A_i|} \overset{(5)}{\geqslant} \frac{Cr^2(r+t)n^{\varepsilon}}{4\varepsilon\log^{4r} n} \overset{(3)}{\geqslant} \frac{Cn^{\frac{\varepsilon}{5}}}{4}\,.$$

For C sufficiently large, this lower bound exceeds both a universal constant and $n^{\frac{\varepsilon}{5}}$. Therefore, the above tail bound implies

$$\Pr\left(\left\langle T, \frac{\mathbf{1}_{[n]}\otimes\bigotimes_{i=2}^{r}\mathbf{1}_{A_i}}{\Delta}\right\rangle \geqslant C\frac{Lr^3}{\varepsilon}(r+t)\right) \leqslant n^{-\frac{C}{10}r^3(r+t)\Delta} \leqslant n^{-\frac{C}{240}\sum_{i=2}^{r}|A_i|}\cdot n^{-\frac{Ct}{480}}\,.$$

Finally, choosing C sufficiently large, we can take a union bound over $\mathcal{W}_j$, and use Lemma A.1 as in the previous lemma to obtain the desired statement. □

Proof of Lemma 3.7. By Lemma 3.9, we have

$$\sup_{X\in\mathcal{X}_P}\langle T,X\rangle \leqslant \max_{j\in\{0,\ldots,r\}}\max_{W\in\mathcal{W}_j}\langle T,W\rangle\,.$$

Taking a union bound over all $j\in\{0,\ldots,r\}$ and using Lemmas 3.12 and 3.13, we obtain, for every $t>0$,

$$\Pr\left(\sup_{X\in\mathcal{X}_P}\langle T,X\rangle \gtrsim \frac{Lr^3}{\varepsilon}(r+t)\right)\leqslant (r+1)n^{-t}\,.$$

Finally, using (3), the factor $r+1$ can be absorbed into the implicit constant after shifting t. □

4 Extensions to arbitrary order and unbounded entries

In this section, we give extensions of Theorem 3.1.

First, if the entries of the random tensor take values in $[-M,M]$ rather than $[-1,1]$, rescaling yields $\mathbb{E}\,\|T-\mathbb{E}\,T\|_{\mathrm{inj}}\lesssim MLr^4/\varepsilon$ under otherwise identical assumptions. Moreover, by padding the dimensions, the theorem also applies to tensors $T\in\mathbb{R}^{n_1\times\cdots\times n_r}$, by setting $n=\max(n_1,\ldots,n_r)$.

Next, we combine Theorem 3.1 with the following recent result of Boedihardjo [Boe24] to obtain a bound for arbitrary r, without assuming (3), at the cost of a slightly worse dependence on r and ε.

Theorem 4.1 ([Boe24, Corollary 1.4])**.** *Let $T\in(\mathbb{R}^n)^{\otimes r}$ be a random tensor with independent, mean-zero entries taking values in $[-K,K]$. Then,*

$$\mathbb{E}\,\|T\|_{\mathrm{inj}}\lesssim\sqrt{r}\sum_{k=1}^{r}\max_{i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_r\in[n]}\left(\sum_{i_k=1}^{n}\mathbb{E}\,T_{i_1,\ldots,i_r}^2\right)^{\frac{1}{2}}+Kr^3\log^2 n\,.$$

Moreover, for all $t\geqslant 1$,

$$\Pr\left(\left|\|T\|_{\mathrm{inj}}-\mathbb{E}\,\|T\|_{\mathrm{inj}}\right|\gtrsim Kt\right)\leqslant e^{-t^2}\,.$$

In the setting of Theorem 3.1, Theorem 4.1 does not give a dimension-free bound. However, it applies to arbitrary r, independent of n and ε.

Corollary 4.2 (Extension of Theorem 3.1 to arbitrary r)**.** *Let $n,r\geqslant 2$ be integers, and let $\varepsilon\in(0,1)$. Let $T\in(\mathbb{R}^n)^{\otimes r}$ be a random tensor with independent entries in $[-1,1]$ and such that for some $L\geqslant 1$, we have $\Pr(T_i\neq 0)\leqslant Ln^{-1-\varepsilon}$ for all $i\in[n]^r$. Then,*

$$\mathbb{E}\,\|T-\mathbb{E}\,T\|_{\mathrm{inj}}\lesssim\frac{Lr^5}{\varepsilon^2}\log^2\frac{r}{\varepsilon}\,.$$

Moreover, for any $t>0$,

$$\Pr\left(\|T-\mathbb{E}\,T\|_{\mathrm{inj}}\gtrsim\frac{Lr^3}{\varepsilon}\left(\frac{r^2}{\varepsilon}\log^2\frac{r}{\varepsilon}+t\right)\right)\leqslant n^{-t}\,.$$

Proof. Up to rescaling L, we assume that $n \geqslant 5$ without loss of generality. If r satisfies (3), the result follows from Theorem 3.1. Suppose instead that $r > \frac{\varepsilon \log n}{5 \log\log n}$. By assumption, for any $k \in [r]$ and $i_1, \ldots, i_{k-1}, i_{k+1}, \ldots, i_r \in [n]$, we have

$$\left(\sum_{i_k=1}^{n} \operatorname{Var} T_{i_1,\ldots,i_r}\right)^{\frac{1}{2}} \leqslant \sqrt{L} n^{-\frac{\varepsilon}{2}} .$$

Applying Theorem 4.1 gives

$$\mathbb{E} \|T - \mathbb{E}\, T\|_{\mathrm{inj}} \lesssim \sqrt{L} r^{\frac{3}{2}} n^{-\frac{\varepsilon}{2}} + r^3 \log^2 n .$$

The first term is bounded by the target bound. For the second term, the assumption on r gives $n \leqslant (r/\varepsilon)^{O(r/\varepsilon)}$, so $r^3 \log^2 n \lesssim \frac{r^5}{\varepsilon^2} \log^2 \frac{r}{\varepsilon}$. For the tail bound, the expectation bound with an additive term $O(\sqrt{t \log n}) \lesssim (t+1) r/\varepsilon$ holds with probability at least $1 - n^{-t}$. ☐

The almost sure boundedness assumption on the entries can also be removed via a truncation argument.

Corollary 4.3 (Extension of Theorem 3.1 to unbounded entries)**.** *Consider the setting of Corollary 4.2, except that the entries of T may be unbounded. Let*

$$C(r, \varepsilon) := \begin{cases} \frac{r^4}{\varepsilon} & \text{if (3) is satisfied,} \\ \frac{r^5}{\varepsilon^2} \log^2 \frac{r}{\varepsilon} & \text{otherwise.} \end{cases} \tag{8}$$

Then,

$$\mathbb{E} \|T - \mathbb{E}\, T\|_{\mathrm{inj}} \lesssim C(r, \varepsilon) L\, \mathbb{E} \|T\|_\infty .$$

Moreover, for any $t > 0$ and $M \geqslant 2\, \mathbb{E} \|T\|_\infty$,

$$\Pr\left(\|T - \mathbb{E}\, T\|_{\mathrm{inj}} \gtrsim LM \left(C(r, \varepsilon) + \frac{r^3 t}{\varepsilon} \right) \right) \leqslant n^{-t} + \Pr\left(\|T\|_\infty > M\right) .$$

Proof. Let $M \geqslant 2\, \mathbb{E} \|T\|_\infty$ and decompose $T = T^{>M} + T^{\leqslant M}$ where $T^{>M}$ contains the entries of magnitude greater than M and $T^{\leqslant M}$ contains the remaining entries.

$$\|T - \mathbb{E}\, T\|_{\mathrm{inj}} \leqslant \left\|T^{>M} - \mathbb{E}\, T^{>M}\right\|_{\mathrm{inj}} + \left\|T^{\leqslant M} - \mathbb{E}\, T^{\leqslant M}\right\|_{\mathrm{inj}} .$$

For the expectation bound, we apply Corollary 4.2 (or Theorem 3.1 if (3) is satisfied) to the second term. For the first term, we use the crude bound

$$\mathbb{E} \left\|T^{>M} - \mathbb{E}\, T^{>M}\right\|_{\mathrm{inj}} \lesssim \mathbb{E} \left\|T^{>M}\right\|_{\mathrm{inj}} \leqslant \mathbb{E} \left\|T^{>M}\right\|_1 \lesssim M ,$$

where the last inequality follows from Lemma A.2 (in the special case $M = 2\, \mathbb{E} \|T\|_\infty$). The tail bound follows similarly, using $\Pr\left(T^{>M} \neq 0\right) = \Pr\left(\|T\|_\infty > M\right)$ and $\left\|\mathbb{E}\, T^{>M}\right\|_{\mathrm{inj}} \lesssim M$. ☐

5 Tensors with i.i.d. entries

In this section, we extend Seginer's theorem (Theorem 1.5) to tensors under a $L^{2+\varepsilon}$-L^2 moment equivalence assumption (Assumption 1.6). We do not state a tail bound, although similar arguments to the ones in Section 4 yield one.

Theorem 5.1 (Seginer-type theorem for tensors)**.** *Let $r \geqslant 2$ and $n_1, \ldots, n_r \geqslant 2$ be integers. Let $T \in \mathbb{R}^{n_1 \times \cdots \times n_r}$ be a tensor whose entries are i.i.d. copies of a random variable Z satisfying Assumption 1.6 with parameters $\varepsilon \in (0,1)$ and $K \geqslant 1$. Set $M := 2\,\mathbb{E}\,\|T\|_\infty$ and $n := \max_{i\in[r]} n_i$. Then,*

$$\sqrt{n\,\mathbb{E}\,[\mathrm{clip}_M(Z)^2]} + M \lesssim \mathbb{E}\,\|T\|_{\mathrm{inj}} \lesssim r^3 \sqrt{n\,\mathbb{E}\,[\mathrm{clip}_M(Z)^2]} + \frac{r^5 \log^2 \frac{r}{\varepsilon}}{\varepsilon^6} KM\,.$$

Both the upper and lower bound in Theorem 1.7 can be deduced from Theorem 5.1 (see the proof below). We note that the dependence on $1/\varepsilon$ of the lower bound in Theorem 1.7 is necessarily exponential, as the following example shows.

Example 5.2. For any fixed $K \geqslant 1$ and $\varepsilon \in (0,1)$, let

$$Z = \begin{cases} K^{\frac{1}{\varepsilon}} & \text{with probability } \frac{K^{-\frac{2}{\varepsilon}}}{2}, \\ -K^{\frac{1}{\varepsilon}} & \text{with probability } \frac{K^{-\frac{2}{\varepsilon}}}{2}, \\ 0 & \text{with probability } 1 - K^{-\frac{2}{\varepsilon}}. \end{cases}$$

Then $\mathbb{E}\,Z = 0$, $\mathbb{E}\,Z^2 = 1$, and $\mathbb{E}\,|Z|^{2+\varepsilon} = K$ so Assumption 1.6 is verified, but $\mathbb{E}\,|Z| = K^{-\frac{1}{\varepsilon}}$. Thus, $\mathbb{E}\,\|T\|_{\mathrm{inj}} \leqslant \mathbb{E}\,\|T\|_1 = K^{-\frac{1}{\varepsilon}} \prod_{i=1}^r n_i$ and hence for $K \gg (\prod_{i=1}^r n_i)^\varepsilon$ one needs to divide by exponential factors in $1/\varepsilon$ to make a lower bound in terms of the second moment valid.

Proofs of Theorems 5.1 and 1.7. We start with the lower bound. Clearly, $\mathbb{E}\,\|T\|_{\mathrm{inj}} \geqslant \mathbb{E}\,\|T\|_\infty = M/2$. To obtain the other term in the lower bound, we use the fact that the injective norm is at least the ℓ_2-norm of any row. Without loss of generality, assume $n_r = n$. Then

$$\begin{aligned} \mathbb{E}\,\|T\|_{\mathrm{inj}} &\geqslant \mathbb{E}\,\|\mathrm{clip}_M(T)\|_{\mathrm{inj}} - \mathbb{E}\,\|T - \mathrm{clip}_M(T)\|_{\mathrm{inj}} \\ &\geqslant \mathbb{E}\,\|\mathrm{clip}_M(T)_{1,\ldots,1,\bullet}\|_2 - \mathbb{E}\,\|T - \mathrm{clip}_M(T)\|_1\,. \end{aligned} \tag{9}$$

By Lemma A.2, the second term satisfies $\mathbb{E}\,\|T - \mathrm{clip}_M(T)\|_1 \lesssim M$. For the first term, we apply Lemma A.3. Since $\mathbb{E}\,\|\mathrm{clip}_M(T)_{1,\ldots,1,\bullet}\|_2^2 = n\,\mathbb{E}\left[\mathrm{clip}_M(Z)^2\right]$, we obtain

$$\mathbb{E}\,\|T\|_{\mathrm{inj}} \geqslant C\sqrt{n\,\mathbb{E}\,[\mathrm{clip}_M(Z)^2]} - C'M\,,$$

for some universal constants $C, C' > 0$, which concludes the proof of the lower bound of Theorem 5.1. To deduce the lower bound of Theorem 1.7, note that by Lemma A.2,

$$\sqrt{n\,\mathbb{E}\,[\mathrm{clip}_M(Z)^2]} + M \gtrsim \sqrt{n\,\mathbb{E}\left[Z^2 \mathbf{1}_{|Z|\leqslant M}\right]} + n\,\mathbb{E}\left[|Z|\,\mathbf{1}_{|Z|>M}\right] \gtrsim \sqrt{n}\,\mathbb{E}\,|Z|\,,$$

and the right-hand side is at least $\sqrt{n}\sigma K^{-\frac{1}{\varepsilon}}$ by Lemma A.5.

For the upper bound, set

$$B := \min\left(\frac{\sqrt{n\,\mathbb{E}\left[\mathrm{clip}_M(Z)^2\right]}}{\log^2 n}, M\right)$$

as the clipping threshold. We use the triangle inequality and bound separately the norms of the clipped and unclipped tensors. For the clipped part, Theorem 4.1 implies

$$\mathbb{E}\left\|\mathrm{clip}_B(T) - \mathbb{E}\left[\mathrm{clip}_B(T)\right]\right\|_{\mathrm{inj}} \lesssim r^3\left(\sqrt{n\,\mathbb{E}\left[\mathrm{clip}_B(Z)^2\right]} + B\log^2 n\right) \lesssim r^3\sqrt{n\,\mathbb{E}\left[\mathrm{clip}_M(Z)^2\right]}\,.$$

For the unclipped part, we consider two cases: if $B = M$, then Lemma A.2 directly gives

$$\mathbb{E}\left\|T - \mathrm{clip}_B(T) - \mathbb{E}\left[T - \mathrm{clip}_B(T)\right]\right\|_{\mathrm{inj}} \leqslant \mathbb{E}\left\|T - \mathrm{clip}_B(T) - \mathbb{E}\left[T - \mathrm{clip}_B(T)\right]\right\|_1 \lesssim M\,.$$

If $B < M$, we apply Corollary 4.3. To apply it, we first verify its sparsity assumption. Note that

$$\Pr\left(Z \neq \mathrm{clip}_B(Z)\right) = \Pr\left(|Z| > B\right) = \Pr\left(|\mathrm{clip}_M(Z)| > B\right) \leqslant \frac{\mathbb{E}\left|\mathrm{clip}_M(Z)\right|^{2+\varepsilon}}{B^{2+\varepsilon}}\,.$$

Here, the second equality uses $B < M$, and the inequality follows from Markov's inequality. Using Lemma A.4 and Assumption 1.6, we have

$$\mathbb{E}\left|\mathrm{clip}_M(Z)\right|^{2+\varepsilon} \leqslant \left(\mathbb{E}\,\mathrm{clip}_M(Z)^2\right)^{1+\frac{\varepsilon}{2}} \frac{\mathbb{E}\,Z^{2+\varepsilon}}{\left(\mathbb{E}\,Z^2\right)^{1+\frac{\varepsilon}{2}}} \leqslant K\left(\mathbb{E}\,\mathrm{clip}_M(Z)^2\right)^{1+\frac{\varepsilon}{2}}\,.$$

Combining these estimates with the definition of B gives

$$\Pr\left(Z - \mathrm{clip}_B(Z) \neq 0\right) \leqslant \frac{K\log^{4+2\varepsilon} n}{n^{1+\frac{\varepsilon}{2}}} \lesssim \frac{K}{n^{1+\frac{\varepsilon}{4}}\varepsilon^4}\,.$$

Thus, we can set $L := C\frac{K}{\varepsilon^4}$ for some constant $C > 0$ and use Corollary 4.3 to obtain

$$\mathbb{E}\left\|T - \mathrm{clip}_B(T) - \mathbb{E}\left[T - \mathrm{clip}_B(T)\right]\right\|_{\mathrm{inj}} \lesssim \frac{r^5\log^2\frac{4r}{\varepsilon}}{\varepsilon^6}KM\,.$$

This bounds the unclipped part in both cases and completes the proof. □

Remark 5.3 (Necessity of the dependence on ε). As in Section 4, the dependence on ε and r can be improved in certain parameter regimes; for example, when r satisfies (3) or when $n^{\frac{\varepsilon}{4}} \geqslant \log^{4+2\varepsilon} n$. However, the divergence as $\varepsilon \to 0$ is unavoidable. For example, let A be an $n \times n$ random matrix such that

$$A_{ij} \overset{\text{i.i.d.}}{\sim} \frac{1}{2n}(\delta_{\sqrt{n}} + \delta_{-\sqrt{n}}) + \frac{1}{2}\left(1 - \frac{1}{n}\right)(\delta_1 + \delta_{-1})\,.$$

Then the maximum column norm of A is of order $\sqrt{\frac{n\log n}{\log\log n}}$. On the other hand, if we denote by σ the standard deviation of the entries, then $\sigma\sqrt{n} + \mathbb{E}\left\|A\right\|_\infty = \Theta(\sqrt{n})$.

References

[ACS25] Omar Al-Ghattas, Jiaheng Chen, and Daniel Sanz-Alonso. Sharp concentration of simple random tensors. *Information and Inference: A Journal of the IMA*, 14(4):iaaf029, 2025. 5

[Ade25] Ishaq Aden-Ali. On the Injective Norm of Sums of Random Tensors and the Moments of Gaussian Chaoses. *Preprint arXiv:2503.10580*, 2025. 5

[AOW15] Sarah R. Allen, Ryan O'Donnell, and David Witmer. How to refute a random CSP. In *Symposium on Foundations of Computer Science (FOCS)*, pages 689–708, 2015. 3

[BGJ+25] Afonso S. Bandeira, Sivakanth Gopi, Haotian Jiang, Kevin Lucca, and Thomas Rothvoss. Tensor Concentration Inequalities: A Geometric Approach. In *Symposium on Theory of Computing (STOC)*, pages 822–832, 2025. 5

[BL06] Yonatan Bilu and Nathan Linial. Lifts, Discrepancy and Nearly Optimal Spectral Gap. *Combinatorica*, 26:495–519, 2006. 7

[BL14] Witold Bednorz and Rafał Latała. On the boundedness of Bernoulli processes. *Annals of Mathematics*, 180(3):1167–1203, 2014. 6

[Boe24] March T. Boedihardjo. Injective norm of random tensors with independent entries. *Preprint arXiv:2412.21193*, 2024. 5, 16

[BS25] Erik Bates and Youngtak Sohn. Balanced multi-species spin glasses. *Preprint arXiv:2507.06522*, 2025. 4

[DM24] Stephane Dartois and Benjamin McKenna. Injective norm of real and complex random tensors I: From spin glasses to geometric entanglement. *Preprint arXiv:2404.03627*, 2024. 4, 5

[DM26] Stephane Dartois and Benjamin McKenna. A moment-based approach to the injective norm of random tensors. *Preprint arXiv:2603.01342*, 2026. 5

[dT23] Tommaso d'Orsi and Luca Trevisan. A Ihara-Bass Formula for Non-Boolean Matrices and Strong Refutations of Random CSPs. In *Computational Complexity Conference (CCC)*, volume 264, pages 27:1–27:16, 2023. 3

[FGK05] Joel Friedman, Andreas Goerdt, and Michael Krivilevich. Recognizing More Unsatisfiable Random k-SAT Instances Efficiently. *SIAM Journal on Computing*, 35, 2005. 3

[FKS89] Joel Friedman, Jeff Kahn, and Endre Szemerédi. On the Second Eigenvalue in Random Regular Graphs. In *Symposium on Theory of Computing (STOC)*, pages 587–598, 1989. 5, 7

[FO05] Uriel Feige and Eran Ofek. Spectral techniques applied to sparse random graphs. *Random Structures & Algorithms*, 27(2):251–275, 2005. 5, 7

[FW95] Joel Friedman and Avi Wigderson. On the second eigenvalue of hypergraphs. *Combinatorica*, 15(1):43–65, 1995. 1, 2, 5

[HKP21] Christopher Hoffmann, Matthew Kahle, and Elliot Paquette. Spectral Gap of Random Graphs and Applications. *International Mathematics Research Notices*, 2021:8353–8404, 2021. 5

[HSS15] Samuel B. Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In *Conference on Learning Theory (COLT)*, volume 40, pages 956–1006, 2015. 3

[JO14] Prateek Jain and Sewoong Oh. Provable Tensor Factorization with Missing Data. In *Advances in Neural Information Processing Systems (NIPS)*, volume 27, 2014. 2, 5, 10

[Kel28] Oliver D. Kellogg. On bounded polynomials in several variables. *Mathematische Zeitschrift*, 27:55–64, 1928. 2

[Lat24] Rafał Latała. On the spectral norm of Rademacher matrices. *Preprint arXiv:2405.13656*, 2024. 6

[LCL20] Jing Lei, Kehui Chen, and Brian Lynch. Consistent community detection in multi-layer network data. *Biometrika*, 107:61–73, 2020. 2, 10

[Lee23] James R. Lee. Spectral Hypergraph Sparsification via Chaining. In *Symposium on Theory of Computing (STOC)*, pages 207–218, 2023. 6

[LS25] Rafał Latała and Marta Strzelecka. Operator $\ell_p \to \ell_q$ norms of random matrices with iid entries. *Journal of Functional Analysis*, 288(3):110720, 2025. 4

[PP24] Jinyoung Park and Huy Tuan Pham. On a conjecture of Talagrand on selector processes and a consequence on positive empirical processes. *Annals of Mathematics*, 199(3):1293–1321, 2024. 6

[PR22] Aaron Potechin and Goutham Rajendran. Sub-exponential time Sum-of-Squares lower bounds for Principal Components Analysis. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 35724–35740, 2022. 3

[Seg00] Yoav Seginer. The Expected Norm of Random Matrices. *Combinatorics, Probability and Computing*, 9:149–166, 2000. 4, 5

[Tal21] Michel Talagrand. *Upper and Lower Bounds for Stochastic Processes: Decomposition Theorems*, volume 60. Springer, 2021. 6, 8

[Ver26] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2026. 8

[vH18] Ramon van Handel. Chaining, interpolation, and convexity. *Journal of the European Mathematical Society*, 20(10):2413–2435, 2018. 6

[ZZ21] Zhixin Zhou and Yizhe Zhu. Sparse random tensors: Concentration, regularization and applications. *Electronic Journal of Statistics*, 15:2483–2516, 2021. 2, 3, 5

A Auxiliary lemmas

Lemma A.1. *Let $r \in \mathbb{N}$, $n \geqslant 2$ and $\lambda > 2$. Then,*

$$\sum_{k_1,\ldots,k_r=1}^{n} \prod_{i=1}^{r} \binom{n}{k_i} n^{-\lambda k_i} \leqslant n^{-r(\lambda-2)}.$$

Proof. Using $\binom{n}{k} \leqslant \frac{n^k}{k!}$, we get

$$\sum_{k_1,\ldots,k_r=1}^{n} \binom{n}{k_1} \cdots \binom{n}{k_r} n^{-\lambda(k_1+\ldots+k_r)} \leqslant \sum_{k_1,\ldots,k_r=1}^{\infty} \frac{n^{-(\lambda-1)(k_1+\ldots+k_r)}}{k_1! \cdots k_r!} = \left(\exp(n^{-(\lambda-1)}) - 1\right)^r.$$

For $x \in [0, \frac{1}{2}]$, we have $e^x - 1 \leqslant 2x$, and since $n \geqslant 2$, this gives

$$\left(\exp(n^{-(\lambda-1)}) - 1\right)^r \leqslant \left(2n^{-(\lambda-1)}\right)^r \leqslant n^{-r(\lambda-2)}.$$

□

Lemma A.2. *Let $X_1, \ldots, X_N \geqslant 0$ be independent random variables and $M := 2\,\mathbb{E}\max_{i\in[N]} X_i$. Then,*

$$\sum_{i=1}^{N} \mathbb{E}\left[X_i \mathbf{1}_{X_i > M}\right] \lesssim M.$$

Proof. We start with the tail integration formula

$$\begin{aligned}\sum_{i=1}^{N} \mathbb{E}\left[X_i \mathbf{1}_{X_i > M}\right] &= \sum_{i=1}^{N} \int_0^\infty \Pr\left(X_i \mathbf{1}_{X_i > M} > t\right) \mathrm{d}t \\ &= M \sum_{i=1}^{N} \Pr\left(X_i > M\right) + \int_M^\infty \sum_{i=1}^{N} \Pr\left(X_i > t\right) \mathrm{d}t. \qquad (10)\end{aligned}$$

By independence,

$$\sum_{i=1}^{N} \Pr\left(X_i > t\right) \leqslant -\log \Pr\left(\max_{i\in[N]} X_i \leqslant t\right) = -\log\left(1 - \Pr\left(\max_{i\in[N]} X_i > t\right)\right).$$

For $t \geqslant M$, Markov's inequality gives $\Pr\left(\max_{i\in[N]} X_i > t\right) \leqslant \frac{1}{2}$. We use the fact that $-\log(1-x) \leqslant 2x$ holds for $0 \leqslant x \leqslant \frac{1}{2}$ and obtain

$$-\log\left(1 - \Pr\left(\max_{i\in[N]} X_i > t\right)\right) \leqslant 2\Pr\left(\max_{i\in[N]} X_i > t\right).$$

Apply this estimate to both terms in (10). For the first term, Markov's inequality gives a bound of M. For the integral term, the tail integration formula gives the same bound. □

Lemma A.3. *Let X be a random vector with independent entries satisfying $\|X\|_\infty \leqslant M$ almost surely for some $M > 0$. Then,*

$$\left(\mathbb{E}\,\|X\|_2^2\right)^{1/2} \leqslant \sqrt{2}\,\mathbb{E}\,\|X\|_2 + M.$$

Proof. Applying $\sqrt{x} - \sqrt{y} \leqslant \sqrt{|x-y|}$ and then Jensen's inequality, we have

$$\left(\mathbb{E}\,\|X\|_2^2\right)^{\frac{1}{2}} \leqslant \mathbb{E}\,\|X\|_2 + \mathbb{E}\left|\|X\|_2^2 - \mathbb{E}\,\|X\|_2^2\right|^{\frac{1}{2}} \leqslant \mathbb{E}\,\|X\|_2 + \left(\operatorname{Var}\|X\|_2^2\right)^{\frac{1}{4}} .$$

By independence and boundedness of the entries,

$$\operatorname{Var}\|X\|_2^2 = \sum_i \operatorname{Var} X_i^2 \leqslant \sum_i \mathbb{E}\,X_i^4 \leqslant M^2 \sum_i \mathbb{E}\,X_i^2 = M^2\,\mathbb{E}\,\|X\|_2^2 .$$

Using $\sqrt{xy} \leqslant \frac{x+y}{2}$ and combining these estimates,

$$\sqrt{\mathbb{E}\,\|X\|_2^2} \leqslant \mathbb{E}\,\|X\|_2 + \frac{M + \sqrt{\mathbb{E}\,\|X\|_2^2}}{2} ,$$

and the claim follows by rearranging the inequality. □

Lemma A.4. *Let $0 < q < p$, and let X be a random variable such that $\mathbb{E}\,|X|^q > 0$ and $\mathbb{E}\,|X|^p < \infty$. Then for any $M > 0$,*

$$\frac{(\mathbb{E}\,|\mathrm{clip}_M(X)|^p)^{1/p}}{(\mathbb{E}\,|\mathrm{clip}_M(X)|^q)^{1/q}} \leqslant \frac{(\mathbb{E}\,|X|^p)^{1/p}}{(\mathbb{E}\,|X|^q)^{1/q}} .$$

Proof. Let

$$m_p(M) \coloneqq \mathbb{E}\,|\mathrm{clip}_M(X)|^p = p\int_0^M t^{p-1} \Pr\left(|X| > t\right) \mathrm{d}t ,$$

where the equality follows from the tail integration formula and a change of variables. Differentiating gives

$$m_p'(M) = pM^{p-1} \Pr\left(|X| > M\right) .$$

The logarithmic derivative of the quotient is therefore

$$\left(\log \frac{m_p^{1/p}}{m_q^{1/q}}\right)'(M) = \frac{1}{p}\frac{m_p'(M)}{m_p(M)} - \frac{1}{q}\frac{m_q'(M)}{m_q(M)} = \Pr\left(|X| > M\right)\left(\frac{M^{p-1}}{m_p(M)} - \frac{M^{q-1}}{m_q(M)}\right) \geqslant 0 ,$$

since $m_p(M) \leqslant M^{p-q} m_q(M)$. Hence, the function $M \mapsto \frac{m_p^{1/p}(M)}{m_q^{1/q}(M)}$ is non-decreasing. Taking the limit $M \to \infty$ proves the desired inequality. □

Lemma A.5. *Let Z be a random variable satisfying Assumption 1.6 for some $\varepsilon \in (0,1)$ and $K \geqslant 1$. Then,*

$$\frac{\sqrt{\mathbb{E}\,Z^2}}{\mathbb{E}\,|Z|} \leqslant K^{\frac{1}{\varepsilon}} .$$

Proof. By Hölder's inequality,

$$\mathbb{E}\,Z^2 \leqslant (\mathbb{E}\,|Z|)^{\frac{\varepsilon}{1+\varepsilon}} (\mathbb{E}\,|Z|^{2+\varepsilon})^{\frac{1}{1+\varepsilon}} \leqslant K^{\frac{1}{1+\varepsilon}} (\mathbb{E}\,|Z|)^{\frac{\varepsilon}{1+\varepsilon}} (\mathbb{E}\,Z^2)^{\frac{1+\frac{\varepsilon}{2}}{1+\varepsilon}} .$$

Rearranging the inequality gives the claim. □

B Seginer-type theorems under moment equivalence

Proposition B.1 (Equivalence between Theorem 1.7 and Conjecture 1.8 under Assumption 1.6)**.** *Let $r \geqslant 2$ and $n_1, \ldots, n_r \geqslant 2$ be integers. Let $T \in \mathbb{R}^{n_1 \times \ldots n_r}$ be a random tensor whose entries are i.i.d. copies of a random variable satisfying Assumption 1.6 for some $\varepsilon \in (0,1)$ and $K \geqslant 1$, and let σ denote the standard deviation of that random variable. Then, letting $n := \max(n_1, \ldots, n_r)$,*

$$\max_{k \in [r]} \mathbb{E} \left[\max_{i_1 \in [n_1], \ldots, i_{k-1} \in [n_{k-1}], i_{k+1} \in [n_{k+1}], \ldots, i_r \in [n_r]} \left(\sum_{i_k=1}^{n_k} T_{i_1,\ldots,i_r}^2 \right)^{1/2} \right] \underset{r,K,\varepsilon}{\asymp} \sigma\sqrt{n} + \mathbb{E}\,\|T\|_\infty \,.$$

Proof. To simplify notations, we assume without loss of generality that $n_r = n$. For the lower bound, it suffices to lower bound the term for $k = r$. Note first that the inequality

$$\max_{i_1,\ldots,i_{r-1}} \left(\sum_{i_r=1}^{n} T_{i_1,\ldots,i_r}^2 \right)^{1/2} \geqslant \|T\|_\infty \tag{11}$$

holds pointwise. For the other term, using $\|x\|_1 \leqslant \sqrt{n} \cdot \|x\|_2$ and Lemma A.5,

$$\mathbb{E} \max_{i_1,\ldots,i_{r-1}} \left\|T_{i_1,\ldots,i_{r-1},\bullet}\right\|_2 \geqslant \mathbb{E}\,\|T_{1,\ldots,1,\bullet}\|_2 \geqslant \frac{\mathbb{E}\,\|T_{1,\ldots,1,\bullet}\|_1}{\sqrt{n}} \geqslant K^{-\frac{1}{\varepsilon}} \sigma\sqrt{n} \,. \tag{12}$$

Combining (11) and (12) completes the proof of the lower bound.

For the upper bound, fix a threshold $B > 0$ and decompose $T = T^{\leqslant B} + T^{>B}$, where $T^{\leqslant B}$ contains the entries of T of magnitude at most B, and $T^{>B}$ contains the remaining entries. Similarly, we use the notation $Z^{\leqslant B} := Z\mathbf{1}_{|Z| \leqslant B}$ and $Z^{>B} := Z\mathbf{1}_{|Z|>B}$, where Z is a random variable distributed as the entries of T.

To simplify notation, we treat the case $k = r$; the same argument applies to arbitrary $k \in [r]$. Moreover, by the triangle inequality, it suffices to bound $\mathbb{E}\max_{i_1,\ldots,i_{r-1}} \|T^{\leqslant B}_{i_1,\ldots,i_{r-1},\bullet}\|_2$ and $\mathbb{E}\max_{i_1,\ldots,i_{r-1}} \|T^{>B}_{i_1,\ldots,i_{r-1},\bullet}\|_2$.

We start with $T^{\leqslant B}$. Bernstein's inequality (Lemma 2.2), followed by a union bound over the rows $(i_1, \ldots, i_{r-1})$ and integrating the resulting tail bound, implies

$$\mathbb{E} \max_{i_1,\ldots,i_{r-1}} \|T^{\leqslant B}_{i_1,\ldots,i_{r-1},\bullet}\|_2^2 \lesssim n\,\mathbb{E}\left[\left(Z^{\leqslant B}\right)^2\right] + \left(nr \log n \operatorname{Var}\left[\left(Z^{\leqslant B}\right)^2\right]\right)^{1/2} + B^2 r \log n \,.$$

Since

$$\mathbb{E}\left[\left(Z^{\leqslant B}\right)^2\right] \leqslant \sigma^2 \quad \text{and} \quad \operatorname{Var}\left[\left(Z^{\leqslant B}\right)^2\right] \leqslant B^2\sigma^2 \,,$$

we choose $B := \sigma\sqrt{n/\log n}$. Then, by Jensen's inequality, we obtain

$$\mathbb{E} \max_{i_1,\ldots,i_{r-1}} \|T^{\leqslant B}_{i_1,\ldots,i_{r-1},\bullet}\|_2 \leqslant \left(\mathbb{E} \max_{i_1,\ldots,i_{r-1}} \|T^{\leqslant B}_{i_1,\ldots,i_{r-1},\bullet}\|_2^2\right)^{\frac{1}{2}} \lesssim_r \sigma\sqrt{n} \,, \tag{13}$$

as desired. It remains to handle $T^{>B}$. By Markov's inequality and Assumption 1.6,

$$\Pr\left(Z^{>B} \neq 0\right) \leqslant \frac{K\sigma^{2+\varepsilon}}{B^{2+\varepsilon}} \leqslant \frac{K \log^{1+\frac{\varepsilon}{2}} n}{n^{1+\frac{\varepsilon}{2}}} \,.$$

Applying the Chernoff bound (Lemma 2.3) to bound the number of nonzero entries in one row, followed by a union bound over all rows, yields the following: there exists $C = C(\varepsilon, K, r) \geqslant 1$ such that $T^{>B}$ has at most C nonzero entries in each row $(i_1, \ldots, i_{r-1})$, with probability at least $1 - n^{-r}$. Let Ω denote this event. Then,

$$\begin{aligned}\mathbb{E} \max_{i_1,\ldots,i_{r-1}} \|T^{>B}_{i_1,\ldots,i_{r-1},\bullet}\|_2 &= \mathbb{E}\left[\max_{i_1,\ldots,i_{r-1}} \|T^{>B}_{i_1,\ldots,i_{r-1},\bullet}\|_2 \mathbf{1}_\Omega\right] + \mathbb{E}\left[\max_{i_1,\ldots,i_{r-1}} \|T^{>B}_{i_1,\ldots,i_{r-1},\bullet}\|_2 \mathbf{1}_{\Omega^c}\right] \\ &\leqslant \sqrt{C}\,\mathbb{E}\,\|T\|_\infty + \Pr(\Omega^c)^{\frac{1}{2}} \left(\mathbb{E}\,\|T\|_2^2\right)^{\frac{1}{2}} \\ &\lesssim_{\varepsilon,K,r} \mathbb{E}\,\|T\|_\infty + \sigma\sqrt{n}\,. \end{aligned} \tag{14}$$

Combining (13) and (14) completes the proof. □