CONCENTRATION OF BOUNDED SPARSE CHAOSES AND SPARSE KHATRI–RAO EMBEDDINGS

GUOZHENG DAI, YIYUN HE, KE WANG, AND YIZHE ZHU

Abstract. We prove moment and mixed-tail inequalities for decoupled homogeneous chaoses of arbitrary fixed order generated by independent, centered, bounded random variables with coordinate-dependent second-moment bounds. The estimates hold for arbitrary real rectangular coefficient tensors and are expressed through weighted slice and partition norms. A Bennett-type logarithmic correction sharpens the contribution of the largest coefficient to these bounds. The proof combines a comparison with sums of sparse exponential variables with an elementary replication argument. We also extend the results to diagonal-free non-decoupled chaoses and sparse sub-exponential inputs. As an application, we establish conditions under which sparse Khatri–Rao random projections approximately preserve norms and inner products for finite collections of tensors. For a sparse random relation containing a planted dense block, we prove matching upper and lower bounds on the required projection dimension, demonstrating the sharpness of the logarithmic correction.

1. Introduction and main results

Concentration inequalities quantify how far a random quantity can deviate from its mean or another typical value. They provide tools for controlling estimation error, studying random matrices, and choosing sample sizes in randomized algorithms [37, 10, 47]. For nonlinear statistics, the useful bounds must reflect both the input distributions and the structure of the function being studied. Polynomials of independent variables arise in subgraph counts in random graphs and in many estimators built from random vectors. For concentration and large deviations of random graph polynomials, see, for example, [30, 12, 13, 38].

Our main object is a polynomial of the form

$$Z \coloneqq \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} a_{i_1,\dots,i_d} X_{i_1}^{(1)} \cdots X_{i_d}^{(d)}, \tag{1}$$

where the coefficients $a_{i_1,\dots,i_d}$ are arbitrary real numbers and the input variables $(X_i^{(s)})$ are independent, centered, and bounded in absolute value by one. Each term contains exactly d factors, one from each of d independent groups of variables. Such a polynomial is called a *decoupled homogeneous chaos of order d*. For $d = 1$ this is a sum; for $d = 2$ it is a bilinear form in two independent random vectors. The coefficients form a multidimensional array, or tensor. We seek bounds on the moments of Z and on the probability that $|Z|$ is large.

We focus on sparse inputs, which are usually zero. For example, an input taking values ± 1 with probability $p/2$ each and zero otherwise has second moment p. Our bounds allow general bounded inputs with individual second-moment bounds, capturing information that bounds based only on their maximum size can miss.

2020 *Mathematics Subject Classification.* Primary 60E15; Secondary 60G50, 65F55.

Key words and phrases. Polynomial chaos, Bennett inequality, sparse random variable, Khatri–Rao embedding, tensor sketching.

Two distinct gains arise for a Bernoulli sum. Let $S \coloneqq \sum_{i=1}^{n}(\delta_i - p)$, where the δ_i are independent Bernoulli variables with success probability $0 < p < 1$, and write $v \coloneqq np(1-p)$ for its variance. Hoeffding's inequality [28] and Bennett's inequality [7] give, respectively,

$$\mathbb{P}(S \geq t) \leq \exp(-2t^2/n), \qquad \mathbb{P}(S \geq t) \leq \exp(-vh(t/v)), \qquad t > 0,$$

where $h(u) \coloneqq (1+u)\log(1+u) - u$. For small p and t comparable to np, the two exponents have orders np^2 and np. This *sparsity gain* comes from using the variance and is also captured by Bernstein's inequality. For larger deviations, $t \gg v$, Bennett's exponent has order $t\log(t/v)$, whereas Bernstein's has order t. This additional logarithmic factor is the *Bennett-type improvement* we seek for higher-degree polynomials.

We first place this question in the existing literature. Much of the theory of polynomial concentration proceeds through moment estimates, which control averages of powers of the random quantity and yield tail bounds through Markov's inequality. For sums of independent variables, Montgomery-Smith [39] obtained tail estimates in the Rademacher case. Gluskin and Kwapień [22] and Hitczenko, Montgomery-Smith, and Oleszkiewicz [27] obtained sharp estimates for symmetric summands with log-concave and log-convex tails, respectively. More generally, Latała [34] established sharp moment estimates for sums of independent nonnegative or symmetric random variables. For quadratic forms in independent centered sub-Gaussian variables, the Hanson–Wright inequality [24, 43] expresses concentration in terms of the Frobenius and operator norms of the coefficient matrix. Further results on the concentration of quadratic forms can be found in [35, 49, 1, 31, 15].

Beyond quadratic forms, Latała [36] obtained sharp moment estimates for Gaussian chaoses of arbitrary order, using norms that measure the coefficient tensor along different groupings of its indices. For symmetric inputs with log-concave tails, Adamczak and Latała [2] obtained sharp bounds for chaoses of order at most three and for exponential chaoses of arbitrary order. Kolesko and Latała [33] developed the corresponding moment theory for log-convex tails. For general polynomials, Adamczak and Wolff [4] developed bounds in terms of arrays of partial derivatives, and Götze, Sambale, and Sinulis [23] obtained multilevel concentration inequalities for independent α-sub-exponential inputs. Related moment inequalities for U-statistics and general functions of independent variables can be found in [21, 9, 45]. Concentration inequalities for simple random tensors were established by Vershynin [48] and Bamberger, Krahmer, and Ward [6]. Adamczak, Latała, and Meller [3] obtained moment and tail estimates for Banach-valued Gaussian chaoses.

For sparse polynomials, Dai and Wang [17] obtained moment and tail bounds that explicitly account for the probabilities that the inputs are nonzero. Their results allow random amplitudes with α-sub-exponential tails, $0 < \alpha \leq 1$, and provide the starting point for the present work. We use the additional assumption of bounded inputs to obtain a Bennett-type logarithmic improvement.

For quadratic forms, Zhou [50] initiated the study of the Hanson–Wright inequality for sparse sub-Gaussian random vectors. He, Wang, and Zhu [26] established sparse Hanson–Wright inequalities and obtained a Bennett-type logarithmic gain through a combinatorial moment argument. Further results on sparse quadratic and bilinear forms and related missing-data models can be found in [51, 40, 16].

We establish moment and tail bounds for the bounded decoupled chaoses described above, of any fixed order (Theorem 1.1 and Corollary 1.2). Our bounds sharpen the contribution of the largest coefficient by a Bennett-type logarithmic factor. For sparse inputs given by Bernoulli variables times independent bounded amplitudes, the other terms match those of Dai and Wang [17], up to constants depending only on the order. We also give simpler formulations and explicit quadratic and cubic cases (Corollaries 1.3 to 1.5). The proof uses a comparison

argument and an elementary replication construction; we explain the main idea at the end of Section 1.

We apply our concentration bounds to sparse Khatri–Rao random projections, which compress tensor data while approximately preserving norms and inner products. To explain the construction, consider an order-d tensor with n entries along each dimension, hence n^d entries in total. An unstructured projection into $\mathbb{R}^M$ uses an $M \times n^d$ matrix, which can be expensive to store and apply. A Khatri–Rao map instead builds each row as the tensor product of d vectors of length n. Each row is stored implicitly through these factors, requiring dn numbers rather than n^d. Each projected coordinate is a polynomial of the form studied above, with the input tensor supplying its coefficients.

This structure also speeds up computation when the input is itself a tensor product of d vectors, called a rank-one tensor. Given these input factors, each projected coordinate can be computed as a product of d inner products of length-n vectors, without forming the full tensor. Computing all M coordinates therefore costs $O(Mdn)$ operations. Making the projection factors sparse reduces both storage and computation further: if each factor entry is nonzero with probability p, the expected computation cost becomes $O(Md(np+1))$. We study how this sparsity and the structure of the input affect the projection dimension M needed to approximately preserve norms and pairwise inner products.

Sparse Khatri–Rao sketches have been studied previously. Sun, Guo, Tropp, and Udell [46] introduce tensor random projections and discuss sparse sign factors. Chen and Jin [14] analyze normalized products of Bernoulli and sub-Gaussian variables, including the two-factor case of the sparse-sign maps considered in this paper. They prove finite-set, subspace, and constrained least-squares guarantees. Rakhshan and Rabusseau [42] study Gaussian projections in rank-one, sums-of-rank-one, and tensor-train forms. Ahle and Knudsen [5] establish strong Johnson–Lindenstrauss moment guarantees for tensor-product sketches. For subspace embeddings, Bujanović, Grubišić, Kressner, and Lam [11] study maps whose rows are tensor products of two independent Gaussian vectors. Beretta and Musco [8] consider tensor products of any fixed number of independent isotropic random vectors with uniformly bounded sub-Gaussian norm, and obtain a required projection dimension that grows nearly linearly with the dimension of the subspace.

Our application identifies conditions on the arrangement of nonzero tensor entries and the sparsity of the projection under which the Bennett-type logarithmic correction reduces the projection dimension required by our bounds. Under these conditions, a single random map approximately preserves the norms of a finite collection of tensors and the inner products between their normalized versions, simultaneously with high probability. For tensors consisting of a block of ones surrounded by a sparse random background of zeros and ones, we prove matching upper and lower bounds on the projection dimension required for norm preservation by our map. These bounds show that the logarithmic improvement is essential for this family.

Beyond randomized numerical linear algebra, we expect our inequalities to be useful for studying subgraph counts in sparse random graphs and hypergraphs, a classical setting for polynomial concentration and upper-tail problems [29, 19, 25]. We leave these developments to future work.

The rest of the paper is organized as follows. In the remainder of Section 1, we state the main concentration bounds, give their quadratic and cubic forms, and explain the proof idea. Section 2 states the application results, Section 3 proves the main concentration inequalities, and Section 4 gives the application proofs. Extensions beyond the bounded decoupled setting are given in Appendix A, and general embedding bounds in Appendix B.

Notation. For an integer $m \geq 1$, write $[m] \coloneqq \{1, \dots, m\}$, and adopt the convention that empty products equal one. For real numbers a, b, write $a \vee b \coloneqq \max\{a, b\}$. Unless otherwise specified, the letters $C, c > 0$ denote constants whose values may change from line to line; subscripts indicate their permitted parameter dependence. We write $f \lesssim_\alpha g$ if $f \leq C_\alpha g$, $f \asymp_\alpha g$ if both comparison inequalities hold, and use $O_\alpha(\cdot)$ with the same convention; omitted subscripts indicate universal constants. The notation $f \ll g$ means that f/g is sufficiently small in the stated regime. For a finite tensor $T = (t_\alpha)$, set $\|T\|_{\mathrm{F}} \coloneqq \big(\sum_\alpha |t_\alpha|^2\big)^{1/2}$ and $\|T\|_\infty \coloneqq \max_\alpha |t_\alpha|$. For a matrix $B = (b_{ij})$, let $\|B\|_{\mathrm{op}}$ denote its Euclidean operator norm and set $\|B\|_{2,\infty} \coloneqq \max_i \big(\sum_j |b_{ij}|^2\big)^{1/2}$. For a random variable Y and $r \geq 1$, write $\|Y\|_{L_r} \coloneqq (\mathbb{E}\,|Y|^r)^{1/r}$. Throughout the paper, $\ell(x) \coloneqq 1 + \log x$ for $x \geq 1$. Throughout, we assume that the coefficient tensor A is nonzero.

1.1. **Bennett-type inequalities for chaoses.** We first specify the input assumptions and the tensor norms needed for the bounds. We then state the general results, explain their low-order cases, and outline the proof.

Fix an order $d \geq 1$ and dimensions $n_1, \dots, n_d$, and write

$$\boldsymbol{i} \coloneqq (i_1, \dots, i_d), \qquad a_{\boldsymbol{i}} \coloneqq a_{i_1, \dots, i_d}.$$

Recall the decoupled homogeneous polynomial chaos of order d defined in (1), which we write as

$$Z_A(X) \coloneqq \sum_{\boldsymbol{i} \in [n_1] \times \cdots \times [n_d]} a_{\boldsymbol{i}} \prod_{s=1}^{d} X_{i_s}^{(s)}.$$

Here $A = (a_{\boldsymbol{i}})$ is a deterministic rectangular d-tensor, and $X = (X_i^{(s)})_{s \in [d],\, i \in [n_s]}$ is a mutually independent family satisfying

$$\mathbb{E} X_i^{(s)} = 0, \qquad \left|X_i^{(s)}\right| \leq 1 \text{ a.s.}, \qquad \mathbb{E}\big(X_i^{(s)}\big)^2 \leq p_i^{(s)}, \qquad 0 < p_i^{(s)} \leq 1. \tag{2}$$

As described above, *homogeneous* refers to the common degree d, and *decoupled* to the use of independent input vectors. For example, $\sum_{i,j} a_{ij} X_i^{(1)} X_j^{(2)}$ is a decoupled quadratic form. The numbers $p_i^{(s)}$ are upper bounds on second moments. They may vary from one coordinate to another, allowing the bounds to reflect different degrees of sparsity. A basic example is $X_i^{(s)} = \delta_i^{(s)} \sigma_i^{(s)}$, where $\delta_i^{(s)} \sim \mathrm{Bernoulli}(p_i^{(s)})$ are Bernoulli random variables and $\sigma_i^{(s)}$ are Rademacher random variables.

The bounds depend on both the sizes of the coefficients and their arrangement within the tensor. We describe these features using norms of slices obtained by fixing some indices, following [17, 23]. For $I \subseteq [d]$, write $I^c \coloneqq [d] \setminus I$ and $\boldsymbol{i}_I \coloneqq (i_s)_{s \in I}$. For a given d-tensor A, we use

$$A_{\boldsymbol{i}_{I^c}} \coloneqq (a_{\boldsymbol{i}})_{\boldsymbol{i}_{I^c}}$$

for the *slice* obtained by fixing the coordinates in I and allowing those in I^c to vary. Thus the subscript records the free coordinates; the fixed coordinates are understood from context. For a three-dimensional tensor, fixing the first index leaves a matrix indexed by the other two; fixing the first two leaves a vector indexed by the third. When $I = \varnothing$, the slice is the full tensor A, and the maximum over fixed indices is omitted. For any finite set S, let $\Pi(S)$ be the set of partitions of S, with the convention that $\Pi(\varnothing) \coloneqq \{\varnothing\}$.

Let $U \subseteq [d]$ and $\mathcal{J} = \{J_1, \dots, J_m\} \in \Pi(U)$, and write $|\mathcal{J}| \coloneqq m$ for its number of blocks. For an array $T = (t_{\boldsymbol{i}_U})$ indexed by the coordinates in U, the *partition norm associated with* $\mathcal{J}$ is

$$\|T\|_{\mathcal{J}} \coloneqq \sup\left\{ \sum_{\boldsymbol{i}_U} t_{\boldsymbol{i}_U} \prod_{q=1}^{m} x^{(q)}_{\boldsymbol{i}_{J_q}} : \left\|x^{(q)}\right\|_2 \leq 1,\ q \in [m] \right\}.$$

For $U = \varnothing$, this norm is the absolute value of the scalar.

The one-block and singleton partitions give the Frobenius and injective multilinear norms, respectively. For a matrix, these are precisely

$$\|A\|_{\{\{1,2\}\}} = \|A\|_{\mathrm{F}}, \qquad \|A\|_{\{\{1\},\{2\}\}} = \|A\|_{\mathrm{op}}.$$

More generally, a two-block partition is an operator norm after reshaping the tensor as a matrix. Partition norms thus extend familiar matrix norms by specifying which indices are grouped together.

We now incorporate the second-moment bounds into these norms. The weighted slice norm measures the size of a slice after weighting its entries; the moment scale also includes the powers of the moment order needed in the bound. For $I \subseteq [d]$ and $\mathcal{J} \in \Pi(I^c)$, the *weighted slice norm* $\mathcal{S}_{I,\mathcal{J}}(A)$ and, for $r \geq 2$, the associated *moment scale* $\Delta_I(A;r)$ are

$$\begin{aligned} \mathcal{S}_{I,\mathcal{J}}(A) &\coloneqq \max_{\boldsymbol{i}_I} \left\| \left(a_{\boldsymbol{i}} \prod_{s \in I^c} \sqrt{p_{i_s}^{(s)}} \right)_{\boldsymbol{i}_{I^c}} \right\|_{\mathcal{J}}, \\ \Delta_I(A;r) &\coloneqq \max_{\mathcal{J} \in \Pi(I^c)} r^{\,\kappa_{I,\mathcal{J}}}\, \mathcal{S}_{I,\mathcal{J}}(A), \qquad \kappa_{I,\mathcal{J}} \coloneqq |I| + |\mathcal{J}|/2. \end{aligned} \tag{3}$$

We omit the dependence on the variance proxies $p_i^{(s)}$ from the notation. In particular, $\mathcal{S}_{[d],\varnothing}(A) = \|A\|_\infty$, and $\Delta_{[d]}(A;r) = r^d \|A\|_\infty$ has the highest order in r.

To compute the weighted slice norm, fix the coordinates in I, weight each free coordinate by its standard-deviation bound $\sqrt{p_i^{(s)}}$, and take the partition norm of the resulting slice. Finally, maximize over the choices of fixed coordinates. Each fixed coordinate contributes a factor r to the moment estimate; each block of free coordinates contributes $\sqrt{r}$. For a matrix and a common variance bound p for all input coordinates, the four resulting scales are

$$\sqrt{r}\, p\, \|A\|_{\mathrm{F}}, \qquad rp\, \|A\|_{\mathrm{op}}, \qquad r^{3/2}\sqrt{p} \max\left\{ \|A\|_{2,\infty}, \left\|A^{\mathsf{T}}\right\|_{2,\infty} \right\}, \qquad r^2 \|A\|_\infty.$$

Only the last scale receives the logarithmic correction in the theorem.

Write $m_A \coloneqq \|A\|_\infty$. To state the theorem compactly, we group the moment scales by the number of free coordinates. For $0 \leq k \leq d$, let $B_k(A;r)$ be the largest scale with k free coordinates, and let $\lambda_A(r)$ compare these grouped scales with the largest-entry contribution:

$$B_k(A;r) \coloneqq \max_{\substack{I \subseteq [d] \\ |I^c| = k}} \Delta_I(A;r), \qquad \lambda_A(r) \coloneqq 1 \vee \min_{1 \leq k \leq d} \left(\frac{r^d m_A}{B_k(A;r)} \right)^{2/k}. \tag{4}$$

Then $B_0(A;r) = \Delta_{[d]}(A;r) = r^d m_A$. The quantity $\lambda_A(r)$ measures how far the largest-entry contribution $r^d m_A$ exceeds the other moment scales. Recall that $\ell(x) \coloneqq 1 + \log x$ for $x \geq 1$.

Theorem 1.1 (Bennett-type moment bound)**.** *For $d \geq 1$, let A be a real d-tensor, and let the mutually independent family $X = (X_i^{(s)})_{s \in [d],\, i \in [n_s]}$ satisfy* (2)*. For every $r \geq 2$,*

$$\|Z_A(X)\|_{L_r} \lesssim_d \max_{1 \leq k \leq d} B_k(A;r) + \frac{r^d m_A}{\ell(\lambda_A(r))^d}. \tag{5}$$

The first term of (5) collects the usual weighted slice and partition scales. When $\lambda_A(r) > 1$, the largest-entry contribution dominates, and the bound simplifies to

$$\|Z_A(X)\|_{L_r} \lesssim_d \frac{r^d m_A}{\ell(\lambda_A(r))^d}. \tag{6}$$

Indeed, $\lambda_A(r) > 1$ implies $B_k(A;r) \le r^d m_A \lambda_A(r)^{-k/2}$, and these terms are absorbed into the right-hand side up to a constant depending only on d. The gain over $r^d m_A$ grows with $\lambda_A(r)$. For Bernoulli random variables times independent bounded amplitudes, the terms $B_k(A;r)$, $1 \le k \le d$, match the corresponding scales in the moment estimates of Dai and Wang [17], up to constants depending only on d. The new feature of Theorem 1.1 is the logarithmic improvement of the largest-entry contribution, replacing $r^d m_A$ by $r^d m_A/\ell(\lambda_A(r))^d$, without changing the other scales.

The moment estimates also give the following tail bound. Here $\eta_A(t)$ plays the role of $\lambda_A(r)$ at a specified deviation level. It controls the logarithmic improvement in the tail exponent.

Corollary 1.2 (Bennett-type tail bound)**.** *Under the assumptions of Theorem 1.1, for $t > 0$ set*

$$\eta_A(t) \coloneqq 1 \vee \min_{\substack{I \subsetneq [d] \\ \mathcal{J} \in \Pi(I^c)}} \left[\frac{m_A}{\mathcal{S}_{I,\mathcal{J}}(A)} \left(\frac{t}{m_A} \right)^{1-\kappa_{I,\mathcal{J}}/d} \right]^{2/|I^c|}. \tag{7}$$

Then

$$\mathbb{P}\left(|Z_A(X)| \ge t\right) \le 2 \exp\left[-c_d \min\left\{ \min_{\substack{I \subsetneq [d] \\ \mathcal{J} \in \Pi(I^c)}} \left(\frac{t}{\mathcal{S}_{I,\mathcal{J}}(A)} \right)^{1/\kappa_{I,\mathcal{J}}}, \left(\frac{t}{m_A} \right)^{1/d} \ell(\eta_A(t)) \right\} \right]. \tag{8}$$

The logarithm in (8) improves the largest-entry rate $(t/m_A)^{1/d}$ when $\eta_A(t)$ is large.

Simplified bounds via a moment threshold. The parameters $\lambda_A(r)$ and $\eta_A(t)$ depend on the moment order r and deviation level t, respectively. A simpler form of the bounds uses a single threshold determined by the coefficient tensor and second-moment bounds. Recall $\mathcal{S}_{I,\mathcal{J}}(A)$ and $\kappa_{I,\mathcal{J}}$ from (3). Define

$$\gamma_d(A) \coloneqq \max_{\substack{I \subsetneq [d] \\ \mathcal{J} \in \Pi(I^c)}} \left(\frac{\mathcal{S}_{I,\mathcal{J}}(A)}{m_A} \right)^{1/(d-\kappa_{I,\mathcal{J}})}. \tag{9}$$

This threshold is well defined and positive since $d - \kappa_{I,\mathcal{J}} = |I^c| - |\mathcal{J}|/2 \ge |I^c|/2 > 0$ for $I \subsetneq [d]$, and all $p_i^{(s)} > 0$. The term $r^d m_A$ dominates all the other moment scales exactly when $r \ge \gamma_d(A)$. The following corollary expresses the logarithmic gain in terms of this threshold.

Corollary 1.3 (Bennett bounds with a fixed threshold)**.** *Under the assumptions of Theorem 1.1, for every $r \ge \max\{2, e\,\gamma_d(A)\}$,*

$$\|Z_A(X)\|_{L_r} \lesssim_d \frac{r^d m_A}{[1 + \log(r/\gamma_d(A))]^d}. \tag{10}$$

Moreover, for every $t > 0$,

$$\begin{aligned} \mathbb{P}\left(|Z_A(X)| \ge t\right) \le 2 \exp\Bigg[- c_d \min\Bigg\{ & \min_{\substack{I \subsetneq [d] \\ \mathcal{J} \in \Pi(I^c)}} \left(\frac{t}{\mathcal{S}_{I,\mathcal{J}}(A)} \right)^{1/\kappa_{I,\mathcal{J}}}, \\ & \left(\frac{t}{m_A} \right)^{1/d} \ell\left(1 \vee \frac{(t/m_A)^{1/d}}{\gamma_d(A)} \right) \Bigg\} \Bigg]. \end{aligned} \tag{11}$$

The comparison with the original bounds follows from $\lambda_A(r) \ge r/\gamma_d(A)$ for $r \ge \gamma_d(A)$, and $\eta_A(t) \ge 1 \vee (t/m_A)^{1/d}/\gamma_d(A)$ for all $t > 0$ (see the proof of Corollary 1.3). Thus the threshold

formulation may give a smaller logarithmic gain, but is easier to use in the explicit quadratic and cubic bounds below.

Extensions. Two extensions are recorded in Appendix A. Standard decoupling [18] transfers the bounded-input estimates, including the logarithmic gain, to diagonal-free non-decoupled chaoses. This includes the quadratic form below. For sparse α-sub-exponential inputs with $1 < \alpha \le \infty$, the same comparison yields a logarithmic power $d\beta$ in the moment denominator and β in the tail rate in the logarithmic regime, where $\beta := 1 - 1/\alpha$. The choice $\alpha = \infty$ recovers the bounded case.

We illustrate these results in the linear, quadratic, and cubic cases. The linear case recovers a familiar Bennett bound; the quadratic and cubic cases give explicit moment bounds where the logarithmic correction is effective and tail bounds valid for every $t > 0$.

The linear case ($d = 1$). For $S := \sum_i a_i X_i$, consider $V^2 := \sum_i a_i^2 p_i > 0$ and $a_{\max} := \max_i |a_i|$. The theorem recovers, up to universal constants, the moment bound implied by Bennett's inequality [7]:

$$\|S\|_{L_r} \lesssim \sqrt{r}\, V + \frac{r a_{\max}}{\ell(1 \vee r a_{\max}^2/V^2)}, \qquad r \ge 2. \tag{12}$$

The quadratic case ($d = 2$). Applying the decoupling extension to the quadratic form $X^{\mathsf{T}} A X$ gives the following sparse analogue of the Hanson–Wright inequality [43] for bounded inputs, with a Bennett-type logarithmic improvement.

Corollary 1.4 (Quadratic Bennett bounds)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix with zero diagonal entries. Let $X_1, \dots, X_n$ be independent and centered random variables with $|X_i| \le 1$ and $\mathbb{E}X_i^2 \le p_i \in (0, 1]$. Set $m_A := \|A\|_\infty$, $D_p := \operatorname{diag}(\sqrt{p_1}, \dots, \sqrt{p_n})$, and*

$$\gamma_2(A) := \max\left\{ \frac{\|AD_p\|_{2,\infty}^2}{m_A^2},\ \left(\frac{\|D_pAD_p\|_{\mathrm{F}}}{m_A}\right)^{2/3},\ \frac{\|D_pAD_p\|_{\mathrm{op}}}{m_A} \right\}.$$

For every $r \ge \max\{2, e\,\gamma_2(A)\}$,

$$\left\| X^{\mathsf{T}} A X \right\|_{L_r} \lesssim \frac{r^2 m_A}{[1 + \log(r/\gamma_2(A))]^2}. \tag{13}$$

Moreover, for every $t > 0$,

$$\begin{aligned} \mathbb{P}\left(\left| X^{\mathsf{T}} A X \right| \ge t \right) \le 2\exp\Bigg[- c \min\Bigg\{ & \frac{t^2}{\|D_pAD_p\|_{\mathrm{F}}^2},\ \frac{t}{\|D_pAD_p\|_{\mathrm{op}}}, \\ & \left(\frac{t}{\|AD_p\|_{2,\infty}}\right)^{2/3},\ \sqrt{\frac{t}{m_A}}\,\ell\left(1 \vee \frac{\sqrt{t/m_A}}{\gamma_2(A)}\right) \Bigg\}\Bigg]. \end{aligned} \tag{14}$$

Here $\gamma_2(A)$ is the threshold in (9) specialized to this symmetric quadratic setting.

Our quadratic bound removes the nonnegativity assumption on A in He, Wang, and Zhu [26, Lemma 6.3]. In their setting of nonnegative A and inputs $X_i = \delta_i \chi_i$, with independent Bernoulli variables δ_i and independent centered amplitudes χ_i bounded by one, our row and operator terms also sharpen their weighted row-sum bounds. Writing $\gamma_{\mathrm{HWZ}} := \max_i \sum_{j\ne i} a_{ij} p_j$, we have

$$\|AD_p\|_{2,\infty}^2 \le \|A\|_\infty\, \gamma_{\mathrm{HWZ}}, \qquad \|D_pAD_p\|_{\mathrm{op}} \le \gamma_{\mathrm{HWZ}}.$$

The first inequality follows from $a_{ij}^2 \le \|A\|_\infty a_{ij}$; the second follows by similarity to AD_p^2 and its maximum row-sum bound. Thus the row and operator terms in $\gamma_2(A)$ are no larger than their HWZ counterparts and can be strictly smaller, while the Frobenius term is unchanged.

The cubic case ($d = 3$). We next specialize Corollary 1.3 to order three, with equal dimensions and a common second-moment bound. For $A \in \mathbb{R}^{n\times n\times n}$, write

$$\|A\|_{\mathrm{inj}} := \sup_{\|x\|_2,\|y\|_2,\|z\|_2\le 1} \Big| \sum_{i,j,k} a_{ijk} x_i y_j z_k \Big|$$

for its injective norm.

The bounds involve six weighted norms of vector slices, matrix slices, and the full tensor. In the notation below, the subscript records the number of free indices. The quantities V use Euclidean or Frobenius norms, W use operator norms, and U_3 uses the injective norm.

Corollary 1.5 (Cubic Bennett bounds)**.** *Suppose that the assumptions of Theorem 1.1 hold with $d = 3$ and $n_1 = n_2 = n_3 = n$, and that $p_i^{(s)} = p \in (0, 1]$ for all $s \in [3]$ and $i \in [n]$. Set $m_A := \|A\|_\infty$ and define*

$$\begin{aligned} V_1 &:= \sqrt{p} \max_{s\in[3]} \max_{\boldsymbol{i}_{\{s\}^c}} \left\| A_{\boldsymbol{i}_{\{s\}}} \right\|_2, \qquad V_2 := p \max_{s<u} \max_{\boldsymbol{i}_{\{s,u\}^c}} \left\| A_{\boldsymbol{i}_{\{s,u\}}} \right\|_{\mathrm{F}}, \\ W_2 &:= p \max_{s<u} \max_{\boldsymbol{i}_{\{s,u\}^c}} \left\| A_{\boldsymbol{i}_{\{s,u\}}} \right\|_{\mathrm{op}}, \quad V_3 := p^{3/2} \|A\|_{\mathrm{F}}, \\ W_3 &:= p^{3/2} \max_{s\in[3]} \|A\|_{\{\{s\},[3]\setminus\{s\}\}}, \quad U_3 := p^{3/2} \|A\|_{\mathrm{inj}}. \end{aligned}$$

Define

$$\gamma_3(A) := \max\left\{ \left(\frac{V_1}{m_A}\right)^2, \left(\frac{V_2}{m_A}\right)^{2/3}, \frac{W_2}{m_A}, \left(\frac{V_3}{m_A}\right)^{2/5}, \left(\frac{W_3}{m_A}\right)^{1/2}, \left(\frac{U_3}{m_A}\right)^{2/3} \right\}. \tag{15}$$

For every $r \ge \max\{2, e\,\gamma_3(A)\}$,

$$\|Z_A(X)\|_{L_r} \lesssim \frac{r^3 m_A}{[1 + \log(r/\gamma_3(A))]^3}. \tag{16}$$

Moreover, for every $t > 0$,

$$\begin{aligned} \mathbb{P}\left(|Z_A(X)| \ge t\right) \le 2\exp\Bigg[- c \min\Bigg\{ &\left(\frac{t}{V_1}\right)^{2/5}, \left(\frac{t}{V_2}\right)^{2/3}, \left(\frac{t}{W_2}\right)^{1/2}, \\ &\frac{t^2}{V_3^2}, \frac{t}{W_3}, \left(\frac{t}{U_3}\right)^{2/3}, \left(\frac{t}{m_A}\right)^{1/3} \ell\left(1 \vee \frac{(t/m_A)^{1/3}}{\gamma_3(A)}\right) \Bigg\}\Bigg]. \end{aligned} \tag{17}$$

In V_1, the vector $A_{\boldsymbol{i}_{\{s\}}}$ is obtained by fixing the other two indices. In V_2 and W_2, the matrix $A_{\boldsymbol{i}_{\{s,u\}}}$ is obtained by fixing the remaining index. For W_3, the full tensor is reshaped into a matrix with index s along the rows and the other two indices along the columns, and we take the largest operator norm over the three choices of s.

Idea of the proof. The proof of Theorem 1.1 combines a moment comparison with an exact calculation of how tensor norms change under replication. After symmetrizing the bounded inputs, we compare each with a suitably scaled sum of independent sparse exponential variables with random signs. Their factorial moment growth allows a logarithmic reduction in the scale of the comparison sum. Expanding the sums repeats the tensor coefficients without changing their

values, a construction we call *replication*. Known moment bounds for the resulting chaos then reduce the proof to choosing how many copies to use.

Example of replication. For a simple example, start with the linear form $aX_1 + bX_2$. Suppose X_1, X_2 are independent and centered, with $|X_j| \le 1$ and $\mathbb{E}X_j^2 \le p_j \in (0, 1]$. For an integer $N \ge 1$, instead of studying X_j's, we define the reference variables by

$$Y_i \coloneqq \delta_i \sigma_i E_i, \qquad i = 1, \dots, 2N,$$

where $\delta_1, \dots, \delta_N \sim \text{Bernoulli}(p_1)$ and $\delta_{N+1}, \dots, \delta_{2N} \sim \text{Bernoulli}(p_2)$, σ_i's are Rademacher random variables, and $E_i \sim \text{Exp}(1)$ have mean one. All these factors are mutually independent. Thus each Y_i is zero or a randomly signed exponential random variable. We compare $aX_1 + bX_2$ with

$$a \sum_{i=1}^{N} Y_i + b \sum_{i=N+1}^{2N} Y_i.$$

The coefficient vector changes from (a, b) to

$$\big(\underbrace{a, \dots, a}_{N \text{ copies}}, \underbrace{b, \dots, b}_{N \text{ copies}}\big),$$

where the largest absolute entry is unchanged but the Euclidean norm grows by a factor of $\sqrt{N}$.

This example illustrates why replication treats the largest coefficient differently from the other norms. To quantify the gain, write $S_1 \coloneqq \sum_{i=1}^{N} Y_i$ and $S_2 \coloneqq \sum_{i=N+1}^{2N} Y_i$. We first compare moments and then estimate the replicated sum. For an even integer $r \ge 2$, the moment comparison in Lemma 3.2 gives

$$\|aX_1 + bX_2\|_{L_r} \le \frac{2\sqrt{e}}{1 + \log N} \, \|aS_1 + bS_2\|_{L_r}.$$

Second, $aS_1 + bS_2$ is a sum of independent random variables with variance $2N(a^2p_1 + b^2p_2)$. The standard Bernstein moment estimate, also the degree-one case of Lemma 3.3, gives

$$\|aS_1 + bS_2\|_{L_r} \le C \left[\sqrt{Nr(a^2p_1 + b^2p_2)} + r \max\{|a|, |b|\} \right].$$

Here the second term, which corresponds to the largest coefficient, is independent of N. Thus we can choose the optimal number of copies N by combining the two bounds:

$$\|aX_1 + bX_2\|_{L_r} \le C \, \frac{\sqrt{Nr(a^2p_1 + b^2p_2)} + r \max\{|a|, |b|\}}{1 + \log N}.$$

For $(a, b) \ne (0, 0)$, the choice $N = \left\lceil \frac{r \max\{a^2, b^2\}}{a^2p_1 + b^2p_2} \right\rceil$ balances the two numerator terms up to a universal constant and gives

$$\|aX_1 + bX_2\|_{L_r} \le \frac{Cr \max\{|a|, |b|\}}{1 + \log\left(\dfrac{r \max\{a^2, b^2\}}{a^2p_1 + b^2p_2} \right)}, \qquad r \ge 2.$$

The argument of the logarithm is at least one because $a^2p_1 + b^2p_2 \le 2 \max\{a^2, b^2\}$ and $r \ge 2$. For an order-d chaos, using N copies in each mode gives the divisor $(1 + \log N)^d$. A weighted slice norm with k free coordinates grows by exactly $N^{k/2}$, while the largest entry is unchanged. When the largest-entry term dominates, we choose the largest replication level for which every enlarged term remains at most $r^d m_A$. This level is $\lambda_A(r)$, and the resulting logarithmic divisor gives the improvement in Theorem 1.1. When $\lambda_A(r) = 1$, we use no replication. The full argument is given in Section 3.

2. Sparse Khatri–Rao embeddings

The squared norm of a sparse Khatri–Rao sketch is an empirical second moment of independent polynomial chaoses. We use the concentration bounds above to control this average. The main result gives an explicit row count under a condition on the support of each tensor slice. We then estimate all normalized inner products in a finite collection, and construct a planted relation for which the row count is sharp up to constants depending only on the tensor order. For this family, the Bennett correction reduces the sufficient projection dimension by a logarithmic factor while keeping the map and sampling probability fixed. A matching lower bound shows that the improved dimension is necessary for this map and its squared-norm estimator; Table 1 summarizes the comparison. General embedding bounds are given in Appendix B.

2.1. **Construction and computational cost.** Throughout this section, take $n_s = n$ in every mode. We identify $\mathbb{R}^{n^d}$ with $\mathbb{R}^{n\times\cdots\times n}$, so vector Euclidean norms are tensor Frobenius norms. Our goal is a random map $\Phi \in \mathbb{R}^{M\times n^d}$ that approximately preserves the norm using $M \ll n^d$ measurements. For a fixed pair of tensors x, y, linearity gives $\Phi x - \Phi y = \Phi(x - y)$. Applying a relative-error guarantee for squared norms to $x - y$ therefore preserves their squared distance, with the same success probability:

$$(1-\varepsilon)\,\|x-y\|_{\mathrm{F}}^2 \le \|\Phi x - \Phi y\|_2^2 \le (1+\varepsilon)\,\|x-y\|_{\mathrm{F}}^2 .$$

A *Khatri–Rao sketch* represents each row as a tensor product of short vectors. This representation avoids forming a length-n^d row and, for a rank-one input, reduces each measurement to a product of d short inner products. The same idea is used in the tensor random projections of Sun, Guo, Tropp, and Udell [46] and the tensor-structured sketches of Chen and Jin [14]. We make the factors sparse to further reduce storage and arithmetic work. Fix a common sampling probability $p \in (0, 1]$ and a number of sketch rows $M \ge 1$. For $s \in [d]$, $m \in [M]$, and $i \in [n]$, let the variables $\xi_{mi}^{(s)}$ be mutually independent, with

$$\mathbb{P}\{\xi_{mi}^{(s)} = 1\} = \mathbb{P}\{\xi_{mi}^{(s)} = -1\} = \frac{p}{2}, \qquad \mathbb{P}\{\xi_{mi}^{(s)} = 0\} = 1 - p.$$

Write $\xi_m^{(s)} \coloneqq (\xi_{mi}^{(s)})_{i=1}^n \in \mathbb{R}^n$. The sketching matrix $\Phi \in \mathbb{R}^{M\times n^d}$ has rows

$$\Phi_{m,:} \coloneqq \frac{1}{\sqrt{Mp^d}}\big(\xi_m^{(1)} \otimes \cdots \otimes \xi_m^{(d)}\big)^{\mathsf{T}}, \qquad m \in [M].$$

On an order-d tensor $x = (x_{\boldsymbol{i}})$, this map acts by

$$(\Phi x)_m \coloneqq \frac{1}{\sqrt{Mp^d}} \sum_{\boldsymbol{i}\in[n]^d} x_{\boldsymbol{i}} \prod_{s=1}^{d} \xi_{mi_s}^{(s)}, \qquad m \in [M]. \tag{18}$$

Independence and centering give

$$\mathbb{E}\langle \Phi x, \Phi y\rangle = \langle x, y\rangle, \qquad \mathbb{E}\,\|\Phi x\|_2^2 = \|x\|_{\mathrm{F}}^2 . \tag{19}$$

We store Φ through its Md sparse factor vectors, recording only nonzero positions and signs. Geometric-gap sampling avoids scanning all coordinates when $p < 1$, so generating and storing the map costs $O(Md(np + 1))$ expected operations and storage words. The cost of applying it depends on how the input is supplied.

For a rank-one tensor supplied through its factors, $x = u^{(1)} \otimes \cdots \otimes u^{(d)}$, each measurement is a product of short inner products,

$$(\Phi x)_m = \frac{1}{\sqrt{Mp^d}} \prod_{s=1}^{d} \left(\sum_{i=1}^{n} \xi_{mi}^{(s)} u_i^{(s)} \right).$$

With coordinate access to the input factors, only sampled coordinates are needed, giving the expected cost

$$O(Md(np+1)) \quad \text{arithmetic operations.} \tag{20}$$

For $np \geq 1$, this is of order p times the cost with dense factors. Reading all input factors adds $O(dn)$ operations.

For a general tensor supplied through entry lookups, each row accesses the entries selected in every mode, averaging $(np)^d$ queries. We assume expected constant-time lookup; a lookup table can be built from a list of nonzero entries in expected time proportional to its length.

For N_{dat} such tensors sharing one map, the total expected cost of generating the map and computing the feature vectors is

$$O_d\Big(M(np+1) + N_{\mathrm{dat}} M[1+(np)^d]\Big). \tag{21}$$

The first term is the shared map-generation cost; the second is the cost of evaluating all tensors. Storing the features takes $O(N_{\mathrm{dat}}M)$ words, and computing their pairwise inner products takes $O(N_{\mathrm{dat}}^2 M)$ arithmetic operations. These counts use unit-cost random draws, entry access, arithmetic, and output writes, excluding bit complexity. At fixed p, reducing M lowers both evaluation costs and feature storage.

2.2. **Tensors with bounded sampled slice counts.** We give a condition for the logarithmic improvement in terms of slice support counts, the numbers of nonzero entries remaining after some indices are fixed. These counts can be computed directly from the list of nonzero entries.

For a real tensor $A = (a_{\boldsymbol{i}})$, write $m_A \coloneqq \|A\|_\infty$. Let $D_I(A)$ denote the maximum number of nonzero entries in any slice obtained by fixing the indices in I,

$$D_I(A) \coloneqq \max_{\boldsymbol{i}_I} \#\{\boldsymbol{i}_{I^c} : a_{\boldsymbol{i}} \neq 0\}, \qquad \varnothing \neq I \subsetneq [d].$$

We require

$$p^{d-|I|} D_I(A) \leq 1 \quad (\varnothing \neq I \subsetneq [d]), \qquad L_A \coloneqq \frac{p^d \|A\|_{\mathrm{F}}^2}{m_A^2} \geq 1. \tag{22}$$

The first condition limits each proper slice to at most one sampled nonzero entry on average, conditional on its fixed coordinates being selected. The quantity L_A is the variance of an unnormalized measurement divided by m_A^2. Both conditions allow entries of either sign. For example, the diagonal tensor $a_{i_1,\dots,i_d} \coloneqq \mathbf{1}\{i_1 = \cdots = i_d\}$ has $D_I(A) = 1$, $m_A = 1$, and $L_A = np^d$, so the conditions hold for $n^{-1/d} \leq p \leq 1$.

The slice counts can be computed by grouping the nonzero entries according to their fixed indices. Together with $\|A\|_{\mathrm{F}}$ and m_A, they give the admissible range of sampling probabilities, provided this interval is nonempty,

$$\left(\frac{m_A^2}{\|A\|_{\mathrm{F}}^2}\right)^{1/d} \leq p \leq \min\left\{1, \min_{\varnothing \neq I \subsetneq [d]} D_I(A)^{-1/(d-|I|)}\right\}. \tag{23}$$

For $L \geq 1$ and $r \geq 2$, define

$$\Gamma_d(L,r) \coloneqq \frac{r^{2d}}{L\,\ell\big(1 \vee r/L^{1/(2d-1)}\big)^{2d}}. \tag{24}$$

Recall that $\ell(u) \coloneqq 1 + \log u$ for $u \geq 1$. The function Γ_d describes the contribution of a rare large row to the squared-norm error.

Theorem 2.1 (Sketching tensors with bounded sampled slice counts)**.** *Fix $d \geq 2$. There is a constant C_d such that the following holds. Let A satisfy* (22) *for the sampling probability p in* (18)*. For every $0 < \varepsilon, \delta < 1$, put $q_\delta \coloneqq \log(e^2/\delta)$. If M is an integer satisfying*

$$M \geq C_d \left[\varepsilon^{-2} q_\delta + \varepsilon^{-1} \Gamma_d(L_A, q_\delta)\right], \tag{25}$$

then

$$\mathbb{P}\left(\left|\frac{\|\Phi A\|_2^2}{\|A\|_{\mathrm{F}}^2} - 1\right| > \varepsilon\right) \leq \delta. \tag{26}$$

There is no additional relation between L_A, δ, and the mode dimension.

The two terms in (25) account for accumulated fluctuations and rare large measurements. Keeping the input, sampling probability, error tolerance, and failure probability fixed, the Bennett correction divides the factor q_δ^{2d}/L_A in the second term of the uncorrected bound by $\ell(1 \vee q_\delta / L_A^{1/(2d-1)})^{2d}$. This gain matters when rare large measurements determine the row count, as in Proposition 2.3. When accumulated fluctuations already dominate, the correction does not change the order of the bound. In particular, if $L_A \geq q_\delta^{2d-1}$, then $C_d \varepsilon^{-2} q_\delta$ rows suffice.

The theorem allows entries of either sign, and its assumptions can be checked from the support and entry magnitudes. For tensors with widely varying magnitudes, using the full weighted norms in Theorem 1.1 can yield finer bounds.

To prove Theorem 2.1, we first bound the weighted tensor norms using the slice counts and the row and column sums of tensor flattenings. The replication estimate then controls the moments of each normalized measurement $Z_m \coloneqq \sqrt{M}(\Phi A)_m / \|A\|_{\mathrm{F}}$. For an even moment order $r \geq 2$ and $M \geq r$, applying an independent-sum moment bound to the squared measurements gives

$$\left\| \frac{1}{M} \sum_{m=1}^{M} (Z_m^2 - 1) \right\|_{L_r} \leq C_d \left[\sqrt{\frac{r}{M}} + \frac{\Gamma_d(L_A, r)}{M} \right].$$

Rounding $\log(e^2/\delta)$ up to an even moment order changes only C_d. Under (25), Markov's inequality then gives failure probability at most δ. Details are in Section 4.

2.3. **Features for aligned relation data.** We extend norm preservation for a single tensor to similarity estimation for a finite collection, using the same random map for all inputs. As in random-feature methods for kernel approximation [41], we approximate normalized inner products by inner products of short feature vectors. For nonzero tensors A, B, define

$$K(A, B) \coloneqq \frac{\langle A, B \rangle}{\|A\|_{\mathrm{F}} \|B\|_{\mathrm{F}}}, \qquad \psi(A) \coloneqq \frac{\Phi A}{\|A\|_{\mathrm{F}}}.$$

For example, tensors may record the presence or strength of multiway interactions among the same indexed entities [32]. For binary tensors, $K(A, B)$ measures their support overlap, normalized by the geometric mean of their support sizes. The shared coordinate labels fix the alignment between tensors. The following result controls all pairwise similarities simultaneously.

Corollary 2.2 (Approximating a relation-overlap Gram matrix)**.** *Let $A_1, \ldots, A_{N_{\mathrm{dat}}}$ be fixed tensors on the same coordinate set $[n]^d$, each satisfying* (22) *for the same p. Fix $0 < \varepsilon, \delta < 1$. If*

$$M \geq C_d \left[\varepsilon^{-2} \log \frac{e^2 N_{\mathrm{dat}}^2}{\delta} + \varepsilon^{-1} \max_{i \in [N_{\mathrm{dat}}]} \Gamma_d\!\left(L_{A_i}, \log \frac{e^2 N_{\mathrm{dat}}^2}{\delta} \right) \right], \tag{27}$$

then, with probability at least $1 - \delta$,

$$|\langle \psi(A_i), \psi(A_j) \rangle - K(A_i, A_j)| \leq \varepsilon \qquad \text{for all } i, j \in [N_{\mathrm{dat}}]. \tag{28}$$

The feature Gram matrix FF^{T}, with $F_{i,:} \coloneqq \psi(A_i)^{\mathsf{T}}$, is positive semidefinite and has rank at most M. It approximates the normalized-input Gram matrix entrywise within ε, and the corresponding squared distances within 4ε. The proof bounds products of measurements directly, so the support condition is needed only for individual tensors. See Section 4 for details.

2.4. **Sharp row counts for a planted relation.** We show that the logarithmic correction is necessary for a family consisting of a block of ones in a sparse random background, as in planted dense-subhypergraph models [20]. The matching bounds below concern our sparse Khatri–Rao map with its squared output norm as the estimator.

Fix $d \geq 2$. Let τ be a sufficiently large positive integer and let $n \geq \tau^2$ be any integer. Set $p \coloneqq \frac{\tau}{n}$ and $b \coloneqq \left\lfloor \frac{n}{8\tau} \right\rfloor$. Define $A \in \{0,1\}^{[n]^d}$ by setting $a_{\boldsymbol{i}} \coloneqq 1$ for $\boldsymbol{i} \in [b]^d$ and taking the remaining entries independently Bernoulli$(1/(8\tau^{d-1}))$, independently of the sketch. The parameter $\tau = np$ is the expected number of nonzero coordinates per projection factor, and b is the side length of the planted block.

Proposition 2.3 (A planted relation with a sharp row count)**.** *For the random tensor A and parameters defined above, there are constants $c > 0$ and $c_d, C_d > 0$, with c absolute, such that, with probability at least $1 - e^{-cn/\tau}$ over A,*

$$L \coloneqq p^d \|A\|_{\mathrm{F}}^2, \qquad \frac{\tau}{16} \leq L \leq \frac{\tau}{4}, \qquad p^{d-|I|} D_I(A) \leq 1 \quad (\forall \varnothing \neq I \subsetneq [d]). \tag{29}$$

Condition on any realization satisfying (29)*. For every $0 < \delta < 1$ with $\tau^{1/d} \leq \log(e^2/\delta) \leq 2\tau^{1/d}$,*

$$M \geq C_d \frac{\tau}{(\log \tau)^{2d}} \quad \Longrightarrow \quad \mathbb{P}_\Phi\left(\left| \frac{\|\Phi A\|_2^2}{\|A\|_{\mathrm{F}}^2} - 1 \right| > \tfrac{1}{2} \right) \leq \delta,$$

whereas

$$1 \leq M \leq c_d \frac{\tau}{(\log \tau)^{2d}} \quad \Longrightarrow \quad \mathbb{P}_\Phi\left(\left| \frac{\|\Phi A\|_2^2}{\|A\|_{\mathrm{F}}^2} - 1 \right| > \tfrac{1}{2} \right) > \delta.$$

Choosing τ of order $(\log n)^d$ gives the following matching row bounds at polynomially small failure probabilities.

Corollary 2.4 (Optimal row count at polynomial confidence)**.** *Fix $d \geq 2$ and $a > 0$. For sufficiently large n, set $\tau \coloneqq \lfloor [\log(e^2 n^a)]^d \rfloor$, and draw A using the construction above. On the event* (29)*, the least row count M for which*

$$\mathbb{P}_\Phi\left(\left| \frac{\|\Phi A\|_2^2}{\|A\|_{\mathrm{F}}^2} - 1 \right| > \tfrac{1}{2} \right) \leq n^{-a}$$

is of order

$$\frac{(\log n)^d}{(\log \log n)^{2d}}, \tag{30}$$

with constants depending only on d, a. Here $p = \tau/n \asymp_{d,a} (\log n)^d/n$, and the event has probability at least $1 - \exp\{-cn/\tau\}$ over A, where $c > 0$ is absolute.

For the planted family, our upper bound gives a sufficient row count, while the matching lower bound shows that fewer rows cannot achieve the stated guarantee, up to constants. We now compare with earlier sufficient bounds, keeping the tensors, sampling probability, error tolerance, and failure probability fixed. Under the same slice-count condition, the bounds of Dai

and Wang [17] or Schudy and Sviridenko [44, Theorem 1.3] give the sufficient row count below, with $L = L_A$,

$$M \geq C_d \left\{ \varepsilon^{-2} \log \frac{e^2}{\delta} + \frac{\varepsilon^{-1} [\log(e^2/\delta)]^{2d}}{L} \right\}. \tag{31}$$

The averaging argument is unchanged, so the gain comes from our sharper bound on each measurement. For the planted family, $L \asymp \tau$ and $\log(e^2/\delta) \asymp \tau^{1/d}$; at relative error $1/2$, Table 1 shows that this gain closes the gap to the lower bound.

TABLE 1. Sufficient and necessary row counts for the same sparse Khatri–Rao map on the planted family.

Bound for the same sparse map	Number of rows
Our sufficient bound	$C_d\tau/(\log\tau)^{2d}$ suffices
Our necessary lower bound	$c_d\tau/(\log\tau)^{2d}$ is necessary
Earlier uncorrected sufficient bound	$C_d\tau$ suffices

With $\tau \asymp (\log n)^d$, the sufficient row count drops from order $(\log n)^d$ to the sharp order in (30). The matching lower bound establishes optimality up to constants for this family, this map, and the squared output-norm estimator at the stated accuracy and confidence. For other tensors, the two upper bounds remain sufficient but need not be sharp.

For context, distribution-only estimates give sufficient row counts for arbitrary fixed tensors, without the slice-count condition. Writing $\theta_p^2 := [p \log(1 + 1/p)]^{-1}$, a standard sub-Gaussian estimate for each measurement followed by Lemma 4.1 gives the sufficient condition

$$M \geq C_d \left\{ \varepsilon^{-2} \theta_p^{4d} \log \frac{e^2}{\delta} + \varepsilon^{-1} \theta_p^{2d} [\log(e^2/\delta)]^d \right\}.$$

For $d = 2$, Chen and Jin [14, Theorem 2.1 and equation (46)] give the sufficient condition

$$M \geq C \left\{ \varepsilon^{-2} p^{-4} \log \frac{e^2}{\delta} + \varepsilon^{-1} p^{-2} [\log(e^2/\delta)]^2 \right\}$$

for arbitrary fixed inputs. These larger sufficient row counts in the planted regime illustrate the benefit of using the tensor's support structure. The comparisons are derived in Section 4.4.

The shared-feature setting follows similarly by averaging products of measurements and taking a union bound over tensor pairs. In the uncorrected comparison, use $L = \min_i L_{A_i}$ and replace δ by $\delta/N_{\mathrm{dat}}^2$. For a polynomial-size collection satisfying (22) with $L_{A_i} \asymp (\log n)^d$, Corollary 2.2 gives the same row-count order at fixed error and polynomially small failure probability. A matching lower bound follows from the diagonal Gram entries if the collection contains a tensor from the planted family.

2.5. **General embeddings.** The method also applies to tensors that do not satisfy the slice-count condition. In Appendix B, we give bounds for preserving a single tensor's norm, pairwise distances in a finite collection, or all norms in a fixed subspace. These bounds use more detailed information about the tensor coefficients. When the projection factors are dense random signs, they recover the standard row-count guarantee.

3. Proofs of the main results

3.1. **The replication estimate.** Recall the weighted norms $\mathcal{S}_{I,\mathcal{J}}(A)$, the moment scales $\Delta_I(A; r)$ from (3), and $\ell(N) := 1 + \log N$.

Proposition 3.1 (Replication bound)**.** *Let A be a real d-tensor and let the mutually independent inputs satisfy* (2)*. For every $r \geq 2$ and $N \geq 1$,*

$$\|Z_A(X)\|_{L_r} \lesssim_d \frac{\max_{I\subseteq[d]} N^{|I^c|/2}\Delta_I(A;r)}{\ell(N)^d}. \tag{32}$$

Consequently, for every $t > 0$,

$$\mathbb{P}(|Z_A(X)| \geq t) \leq 2\exp\left[-c_d \min_{\substack{I\subseteq[d]\\ \mathcal{J}\in\Pi(I^c)}} \left(\frac{t\ell(N)^d}{N^{|I^c|/2}\mathcal{S}_{I,\mathcal{J}}(A)}\right)^{1/\kappa_{I,\mathcal{J}}}\right], \qquad \kappa_{I,\mathcal{J}} = |I| + |\mathcal{J}|/2. \tag{33}$$

The parameter N balances two effects. Replication enlarges a norm with k free coordinates by $N^{k/2}$, but the moment comparison gains the divisor $\ell(N)^d$. In particular, the largest-entry term, for which $k = 0$, is not enlarged. We first prove the estimate for integer N, then extend it by rounding; Section 3.3 chooses N to obtain the main results.

3.2. **Proof of the replication estimate.** The argument combines a coordinatewise moment comparison, a known bound for sparse-exponential chaoses, and an exact calculation of how replication changes the weighted norms.

3.2.1. *A scalar comparison and its multilinear consequence.* Let σ be a Rademacher sign and $E \sim \mathrm{Exp}(1)$, independent of each other. The symmetric exponential variable $\zeta := \sigma E$ satisfies

$$\mathbb{E}\zeta^m = 0 \quad (m \text{ odd}), \qquad \mathbb{E}\zeta^m = m! \quad (m \text{ even}), \qquad \mathbb{E}\exp(|\zeta|/2) = 2.$$

For an integer $N \geq 1$, replace each mode-s coordinate i_s by $(i_s, u_s) \in [n_s] \times [N]$, and define

$$a^{[N]}_{(i_1,u_1),\dots,(i_d,u_d)} := a_{i_1,\dots,i_d}, \qquad Y^{(s)}_{i,u} := \delta^{(s)}_{i,u}\sigma^{(s)}_{i,u}E^{(s)}_{i,u}, \qquad \delta^{(s)}_{i,u} \sim \mathrm{Bernoulli}(p^{(s)}_i).$$

All Bernoulli variables, signs, and exponential variables are mutually independent. The replicated tensor $A^{[N]}$ has dimensions $n_1N \times \cdots \times n_dN$; each replica inherits the original probability $p^{(s)}_i$.

Lemma 3.2 (Replicated moment comparison)**.** *Under the assumptions of Proposition 3.1, for every even integer $r \geq 2$ and integer $N \geq 1$,*

$$\|Z_A(X)\|_{L_r} \leq \frac{(2\sqrt{e})^d}{\ell(N)^d}\,\|Z_{A^{[N]}}(Y)\|_{L_r}.$$

Proof. Step 1: symmetrize the bounded inputs. Let X' be an independent copy of the entire input family and put $V^{(s)}_i = (X^{(s)}_i - X'^{(s)}_i)/2$. These coordinates are independent and symmetric, and

$$|V^{(s)}_i| \leq 1, \qquad \mathbb{E}(V^{(s)}_i)^2 = \tfrac{1}{2}\mathbb{E}(X^{(s)}_i)^2 \leq p^{(s)}_i.$$

For each monomial, independence and centering give

$$\mathbb{E}_{X'}\prod_{s=1}^{d}(X^{(s)}_{i_s} - X'^{(s)}_{i_s}) = \prod_{s=1}^{d} X^{(s)}_{i_s}.$$

Thus $Z_A(X) = \mathbb{E}_{X'}Z_A(X - X')$. Conditional Jensen's inequality, followed by degree-d homogeneity, yields

$$\mathbb{E}|Z_A(X)|^r \leq \mathbb{E}|Z_A(X - X')|^r = 2^{dr}\mathbb{E}|Z_A(V)|^r.$$

Step 2: compare the scalar moments. Put $S_i^{(s)} = \sum_{u=1}^N Y_{i,u}^{(s)}$. For every even integer $m \geq 2$, independence gives the multinomial expansion

$$\mathbb{E}(S_i^{(s)})^m = \sum_{\substack{m_1+\cdots+m_N=m \\ m_u \geq 0}} \frac{m!}{m_1! \cdots m_N!} \prod_{u=1}^N \mathbb{E}(Y_{i,u}^{(s)})^{m_u}.$$

Terms with an odd m_u vanish, and all remaining terms are nonnegative. Keeping the N terms with one exponent equal to m and all others zero gives

$$\mathbb{E}(S_i^{(s)})^m \geq N p_i^{(s)} m! \geq \frac{p_i^{(s)} \ell(N)^m}{e}.$$

The second inequality follows from the exponential series: $\exp(\ell(N)) = eN \geq \ell(N)^m/m!$. On the other hand, boundedness gives $\mathbb{E}|V_i^{(s)}|^m \leq \mathbb{E}(V_i^{(s)})^2 \leq p_i^{(s)}$. Since $e^{m/2-1} \geq 1$, we obtain

$$\mathbb{E}(V_i^{(s)})^m \leq \mathbb{E}\left(\frac{\sqrt{e}\, S_i^{(s)}}{\ell(N)}\right)^m \qquad (m \geq 2 \text{ even}). \tag{34}$$

Both sides have vanishing odd moments.

Step 3: replace the coordinates one at a time. Take the comparison variables independent of V. After conditioning on every coordinate except one, the chaos is $a + bv$, where a, b are fixed real numbers under the conditioning. For symmetric v and even r,

$$\mathbb{E}_v|a + bv|^r = \sum_{\substack{0 \leq m \leq r \\ m \text{ even}}} \binom{r}{m} a^{r-m} b^m \mathbb{E} v^m.$$

Both m and $r-m$ are even, so every coefficient is nonnegative. Consequently, (34) allows $v = V_i^{(s)}$ to be replaced by $\sqrt{e}\, S_i^{(s)}/\ell(N)$ without decreasing the conditional r-th moment. Integrate over the other coordinates and repeat for the entire finite input family. Every monomial contains exactly d coordinates, hence

$$\|Z_A(V)\|_{L_r} \leq \frac{e^{d/2}}{\ell(N)^d} \|Z_A(S)\|_{L_r}.$$

Step 4: identify the replicated chaos. Distributing the sums gives the exact identity

$$Z_A(S) = \sum_{\boldsymbol{i}} a_{\boldsymbol{i}} \prod_{s=1}^d \left(\sum_{u_s=1}^N Y_{i_s,u_s}^{(s)}\right) = \sum_{\boldsymbol{i}} \sum_{\boldsymbol{u} \in [N]^d} a_{\boldsymbol{i}} \prod_{s=1}^d Y_{i_s,u_s}^{(s)} = Z_{A^{[N]}}(Y).$$

Combining this with Steps 1 and 3 proves the lemma. □

3.2.2. *The reference bound and the cost of replication.* The following specialization of Dai and Wang [17, Theorem 1] estimates the comparison chaos.

Lemma 3.3 (Sparse-exponential chaos)**.** *Let $W_i^{(s)} := \delta_i^{(s)} \chi_i^{(s)}$, where all Bernoulli variables and amplitudes are mutually independent, $\delta_i^{(s)} \sim \mathrm{Bernoulli}(p_i^{(s)})$, $\mathbb{E}\chi_i^{(s)} = 0$, and $\mathbb{E}\exp(\left|\chi_i^{(s)}\right|/K_{\exp}) \leq 2$. For every real d-tensor A and every $r \geq 2$,*

$$\|Z_A(W)\|_{L_r} \leq C_d K_{\exp}^d \max_{I \subseteq [d]} \Delta_I(A; r). \tag{35}$$

To obtain this specialization, regard each mode-coordinate pair as a distinct variable. Centering makes all expected derivatives of order below d vanish; the top derivative consists of permuted blocks of A, whose weighted slice norms give $\Delta_I(A;r)$. The number of blocks and partitions depends only on d, so their sum is bounded by a constant times the displayed maximum.

Lemma 3.4 (Exact norm scaling)**.** *For integer $N \geq 1$, evaluate the norms of $A^{[N]}$ using the inherited probabilities $p^{[N],(s)}_{(i,u)} \coloneqq p^{(s)}_i$. Then*

$$\mathcal{S}_{I,\mathcal{J}}(A^{[N]}) = N^{|I^c|/2}\mathcal{S}_{I,\mathcal{J}}(A), \qquad \Delta_I(A^{[N]};r) = N^{|I^c|/2}\Delta_I(A;r).$$

Proof. Fix I, the values of $\boldsymbol{i}_I$, and $\mathcal{J} \in \Pi(I^c)$. Write $U = I^c$ and consider the weighted slice

$$b_{\boldsymbol{i}_U} = a_{\boldsymbol{i}} \prod_{s\in U} \sqrt{p^{(s)}_{i_s}}.$$

The values of the fixed replica indices $\boldsymbol{u}_I$ do not affect this slice. If $U = \emptyset$, its norm is the absolute value of one entry, which is unchanged. Assume $U \neq \emptyset$.

For each block $J \in \mathcal{J}$, take a test array $x^{(J)}_{\boldsymbol{i}_J,\boldsymbol{u}_J}$ with Euclidean norm at most one, and set

$$y^{(J)}_{\boldsymbol{i}_J} = N^{-|J|/2} \sum_{\boldsymbol{u}_J\in[N]^J} x^{(J)}_{\boldsymbol{i}_J,\boldsymbol{u}_J}.$$

Cauchy–Schwarz, applied to the $N^{|J|}$ replica indices, gives

$$\sum_{\boldsymbol{i}_J} |y^{(J)}_{\boldsymbol{i}_J}|^2 \leq \sum_{\boldsymbol{i}_J,\boldsymbol{u}_J} |x^{(J)}_{\boldsymbol{i}_J,\boldsymbol{u}_J}|^2 \leq 1.$$

Because the blocks partition U, direct expansion yields

$$\sum_{\boldsymbol{i}_U,\boldsymbol{u}_U} b_{\boldsymbol{i}_U} \prod_{J\in\mathcal{J}} x^{(J)}_{\boldsymbol{i}_J,\boldsymbol{u}_J} = N^{|U|/2} \sum_{\boldsymbol{i}_U} b_{\boldsymbol{i}_U} \prod_{J\in\mathcal{J}} y^{(J)}_{\boldsymbol{i}_J}.$$

Taking the supremum over admissible $x^{(J)}$ proves that the replicated partition norm is at most $N^{|U|/2}\,\|b\|_{\mathcal{J}}$. Conversely, for any admissible $y^{(J)}$, take

$$x^{(J)}_{\boldsymbol{i}_J,\boldsymbol{u}_J} = N^{-|J|/2} y^{(J)}_{\boldsymbol{i}_J}.$$

These arrays have the same Euclidean norms as the corresponding $y^{(J)}$, and the preceding identity gives the reverse inequality. Hence the norm scales exactly by $N^{|U|/2}$. Maximizing over the fixed original and replica indices proves the assertion for $\mathcal{S}_{I,\mathcal{J}}$; maximizing over partitions after multiplying by $r^{\kappa_{I,\mathcal{J}}}$ proves the assertion for Δ_I. □

3.2.3. *Completion of the replication estimate.*

Proof of Proposition 3.1. For even integers $r \geq 2$ and integers $N \geq 1$, apply the preceding three lemmas with $K_{\exp} = 2$ to obtain

$$\|Z_A(X)\|_{L_r} \lesssim_d \frac{\|Z_{A^{[N]}}(Y)\|_{L_r}}{\ell(N)^d} \lesssim_d \frac{\max_{I\subseteq[d]} \Delta_I(A^{[N]};r)}{\ell(N)^d} = \frac{\max_{I\subseteq[d]} N^{|I^c|/2}\Delta_I(A;r)}{\ell(N)^d}.$$

The first two constants depend only on d.

For real $r \geq 2$ and $N \geq 1$, set $\bar{r} = 2\lceil r/2\rceil$ and $\bar{N} = \lceil N\rceil$. Since $\bar{r} \leq 2r$, $\bar{N} \leq 2N$, and $\kappa_{I,\mathcal{J}} \leq d$, rounding increases the numerator by at most $2^{3d/2}$ and cannot decrease $\ell(N)^d$. Monotonicity of L_r-norms therefore proves (32) for real r, N.

Finally, since $Z_A(X)$ is centered, the standard moment-to-tail implication [23, Proposition 3.3] applied to (32) gives (33). □

3.3. **Choosing the replication level.**

Proof of Theorem 1.1. In Proposition 3.1, group the terms by the number $k = |I^c|$ of free coordinates. The case $k = 0$ is the largest-entry term $\Delta_{[d]}(A; r) = r^d m_A$, and the maximum over the terms with $k \geq 1$ is $B_k(A; r)$. Thus

$$\|Z_A(X)\|_{L_r} \lesssim_d \frac{\max\{r^d m_A, \max_{1\leq k\leq d} N^{k/2} B_k(A; r)\}}{\ell(N)^d}. \tag{36}$$

To keep the numerator equal to $r^d m_A$, it suffices that

$$N^{k/2} B_k(A; r) \leq r^d m_A \quad \text{for every } k, \qquad \text{or equivalently} \qquad N \leq \min_{1\leq k\leq d} \left(\frac{r^d m_A}{B_k(A; r)}\right)^{2/k}.$$

If $\lambda_A(r) > 1$, its definition makes $N = \lambda_A(r)$ the largest admissible common replication level. Substitution into (36) proves (6) and hence also (5). If $\lambda_A(r) = 1$, use $N = 1$ instead. Since $\ell(1) = 1$,

$$\max\{r^d m_A, \max_k B_k(A; r)\} \leq r^d m_A + \max_k B_k(A; r),$$

which is precisely the right-hand side of (5) in this case. □

Proof of Corollary 1.2. Put $q_0 = (t/m_A)^{1/d}$. If $\eta_A(t) = 1$, take $N = 1$ in (33). The entry term gives the rate q_0; every other term gives $(t/\mathcal{S}_{I,\mathcal{J}}(A))^{1/\kappa_{I,\mathcal{J}}}$. Since $\ell(\eta_A(t)) = 1$, this is exactly the claimed minimum in (8).

Now suppose $\eta_A(t) > 1$, and take $N = \eta_A(t)$, $\Lambda = \ell(N)$. For $I \subsetneq [d]$, write $k = |I^c| \geq 1$ and $\kappa = \kappa_{I,\mathcal{J}}$. The definition of $\eta_A(t)$ implies

$$N \leq \left(\frac{m_A q_0^{d-\kappa}}{\mathcal{S}_{I,\mathcal{J}}(A)}\right)^{2/k}, \qquad \text{hence} \qquad N^{k/2} \mathcal{S}_{I,\mathcal{J}}(A) q_0^{\kappa} \leq t.$$

Since $\Lambda \geq 1$ and $\kappa \leq d$,

$$\frac{N^{k/2} \mathcal{S}_{I,\mathcal{J}}(A)(q_0\Lambda)^{\kappa}}{\ell(N)^d} \leq t\Lambda^{\kappa-d} \leq t.$$

The entry term satisfies equality: $m_A(q_0\Lambda)^d/\ell(N)^d = t$. Consequently, every rate in (33) is at least $q_0\Lambda$. This proves the stronger estimate $\mathbb{P}(|Z_A(X)| \geq t) \leq 2e^{-c_d q_0 \Lambda}$, which implies (8) because its displayed minimum is at most $q_0\Lambda$. □

Proof of Corollary 1.3. Write $\gamma = \gamma_d(A)$.

Step 1: the replication level. Fix $I \subsetneq [d]$ and $\mathcal{J} \in \Pi(I^c)$, and put $k = |I^c| \geq 1$ and $\kappa = \kappa_{I,\mathcal{J}}$. By the definition of γ, $\mathcal{S}_{I,\mathcal{J}}(A) \leq m_A \gamma^{d-\kappa}$. For $u = r/\gamma \geq 1$, the candidate of this pair in the minimum defining $\lambda_A(r)$ satisfies

$$\left(\frac{r^d m_A}{r^{\kappa} \mathcal{S}_{I,\mathcal{J}}(A)}\right)^{2/k} \geq u^{2(d-\kappa)/k} \geq u.$$

The last inequality uses $d - \kappa = k - |\mathcal{J}| /2 \geq k/2$, because $|\mathcal{J}| \leq k$, and $u \geq 1$. Taking the minimum over all pairs gives $\lambda_A(r) \geq r/\gamma$ for $r \geq \gamma$. If $r \geq \max\{2, e\gamma\}$, then $\lambda_A(r) \geq e > 1$ and $\ell(\lambda_A(r)) \geq 1 + \log(r/\gamma)$. Substituting this into (6) proves (10).

Step 2: the tail. Put $q_0 = (t/m_A)^{1/d}$. For $I, \mathcal{J}$ as above, the candidate of this pair in the definition of $\eta_A(t)$ equals $(m_A q_0^{d-\kappa}/\mathcal{S}_{I,\mathcal{J}}(A))^{2/k}$. If $q_0 \geq \gamma$, the computation of Step 1, with r replaced by q_0, bounds it below by q_0/γ. Hence, for every $t > 0$,

$$\eta_A(t) \geq 1 \vee \frac{(t/m_A)^{1/d}}{\gamma}.$$

This is an algebraic comparison and does not require $q_0 \geq 2$. Since ℓ is increasing, (8) implies (11). □

For the quadratic corollary we also need forward decoupling for diagonal-free polynomials [18, Theorem 1]. If A is diagonal-free and $Q_A(X) \coloneqq \sum_{\boldsymbol{i}} a_{\boldsymbol{i}} X_{i_1} \cdots X_{i_d}$, then

$$\begin{aligned} \mathbb{P}(|Q_A(X)| \geq t) &\leq C_d \mathbb{P}\left(\left|Z_A(X^{(1)}, \dots, X^{(d)})\right| \geq t/C_d\right), \\ \|Q_A(X)\|_{L_r} &\lesssim_d \left\|Z_A(X^{(1)}, \dots, X^{(d)})\right\|_{L_r}, \qquad r \geq 2, \end{aligned} \tag{37}$$

where the input vectors on the right are independent copies. Only the forward inequality is used; it does not require symmetry of A. The moment comparison follows by integrating the cited tail inequality, with a constant depending only on d.

Proof of Corollaries 1.4 and 1.5. Step 1: reduction to the decoupled case. The cubic chaos is already decoupled. For $d = 2$, the zero diagonal allows us to apply the moment comparison in (37) to $Q_A(X) = X^{\mathsf{T}} A X$. Combining it with (32) gives the same common-N moment estimate for the quadratic form, with a different universal constant. The same moment-to-tail implication and choices of N therefore give the conclusions of Theorem 1.1 and Corollaries 1.2 and 1.3 for Q_A, with $p_i^{(s)} = p_i$. In particular, decoupling is applied before the moment-to-tail conversion, so its constants do not change the arguments of λ_A, η_A, or γ_d.

Step 2: identify all non-entry norms. For $d = 2$, fixing one coordinate leaves a weighted row or column. Symmetry of A makes their maximal Euclidean norms equal to $\|AD_p\|_{2,\infty}$. With neither coordinate fixed, the one-block partition gives the Frobenius norm of $D_p A D_p$, and the two-block partition gives its operator norm. Since $\kappa = 2 - k + |\mathcal{J}|/2$, this yields

$$B_1(A; r) = r^{3/2} \|AD_p\|_{2,\infty}, \qquad B_2(A; r) = \max\{\sqrt{r}\, \|D_p A D_p\|_{\mathrm{F}}, r \|D_p A D_p\|_{\mathrm{op}}\}.$$

For $d = 3$, a slice with k free coordinates has weight $p^{k/2}$, and $\kappa = 3 - k + |\mathcal{J}|/2$. With one free coordinate, the slices are vectors, so the only partition gives V_1 with exponent $\kappa = 5/2$. With two free coordinates, the slices are matrices: the one-block partition gives their weighted Frobenius norm V_2, with $\kappa = 3/2$, while the two-block partition gives their weighted operator norm W_2, with $\kappa = 2$.

With all three coordinates free, one block gives the weighted Frobenius norm V_3, with $\kappa = 1/2$; two blocks give the weighted matrix-flattening norms, whose maximum is W_3, with $\kappa = 1$; and three blocks give the weighted injective norm U_3, with $\kappa = 3/2$. These scales already include the maxima over the applicable fixed coordinates and choices of modes. Therefore

$$\begin{aligned} B_1(A; r) &= r^{5/2} V_1, \\ B_2(A; r) &= \max\{r^{3/2} V_2, r^2 W_2\}, \\ B_3(A; r) &= \max\{\sqrt{r}\, V_3, r W_3, r^{3/2} U_3\}. \end{aligned}$$

Step 3: apply the bounds with a fixed threshold. By (9), a non-entry scale $\mathcal{S}$ with exponent κ contributes $(\mathcal{S}/m_A)^{1/(d-\kappa)}$ to $\gamma_d(A)$. For $d = 2$, the scales $\|AD_p\|_{2,\infty}$, $\|D_p A D_p\|_{\mathrm{F}}$, $\|D_p A D_p\|_{\mathrm{op}}$ have $\kappa = 3/2$, $1/2$, 1, which gives the three terms of $\gamma_2(A)$. For $d = 3$, the scales $V_1, V_2, W_2, V_3, W_3, U_3$ have $\kappa = 5/2$, $3/2$, 2, $1/2$, 1, $3/2$, which gives the six terms in (15). Thus $\gamma_2(A)$ and $\gamma_3(A)$ are the thresholds in (9), and (10) gives (13) and (16). Finally, the non-entry rates in (11) are $(t/\mathcal{S})^{1/\kappa}$. For $d = 2$, their exponents are $2/3$, 2, 1, respectively; for $d = 3$, they are $2/5$, $2/3$, $1/2$, 2, 1, $2/3$. Inserting these rates into (11) proves (14) and (17). □

4. Proofs for sparse Khatri–Rao embeddings

We first derive a moment bound for one measurement from the slice-count assumptions. An elementary independent-sum estimate then converts the moments of the squared measurements into the stated row count. The same argument handles products of measurements and proves the Gram-matrix guarantee. We next prove the planted example and the comparison bounds.

4.1. **Partition norms and the row moment curve.** Fix A as in Theorem 2.1, and abbreviate $L = L_A$. Replacing A by A/m_A, we may assume $\|A\|_\infty = 1$ and $p^d \|A\|_{\mathrm{F}}^2 = L$. For every proper slice with k free coordinates, its weighted Frobenius norm is at most

$$\Big(p^k D_I(A)\Big)^{1/2} \le 1.$$

This also bounds every partition norm of that slice. For a nontrivial full-tensor partition, first split its blocks into two nonempty sets of modes U, V. The flattening of $p^{d/2}|A|$ across U, V has maximum row and column sums bounded, respectively, by

$$p^{d/2} D_U(A) \le p^{d/2-|V|}, \qquad p^{d/2} D_V(A) \le p^{d/2-|U|}.$$

Schur's test therefore bounds the operator norm by one. Taking absolute values and grouping the partition's test arrays into tensor products gives the same bound for every nontrivial partition norm of $p^{d/2}A$. Consequently,

$$\begin{aligned} &\mathcal{S}_{\varnothing,\{[d]\}}(A) = \sqrt{L}, \qquad \mathcal{S}_{\varnothing,\mathcal{J}}(A) \le 1 \quad (|\mathcal{J}| \ge 2), \\ &B_k(A;r) \le r^{d-k/2} \quad (1 \le k < d), \\ &B_d(A;r) \le \max\{\sqrt{rL}, r^{d/2}\}, \qquad B_0(A;r) = r^d. \end{aligned} \tag{38}$$

For one row, write $\xi_i^{(s)} = \xi_{1i}^{(s)}$ and $S_A = \sum_{\boldsymbol{i}} a_{\boldsymbol{i}} \prod_s \xi_{i_s}^{(s)}$. Set $r_0 = L^{1/(2d-1)}$. When $r > r_0$, we increase replication while keeping the Frobenius and slice terms below r^d, thereby gaining the logarithmic divisor. This leads to the following choice in Proposition 3.1, with $N = 1$ when $r \le r_0$,

$$N = \min\left\{r, \max\left\{1, \left(\frac{r^{2d-1}}{L}\right)^{1/d}\right\}\right\}. \tag{39}$$

If $L \ge r^{2d-1}$, then $N = 1$, and every term in the numerator of the replication bound is at most $\sqrt{rL}$. Otherwise, the choice of N ensures $N^{d/2}\sqrt{rL} \le r^d$, while $N \le r$ bounds the remaining terms by

$$N^{k/2} r^{d-k/2} \le r^d, \qquad N^{d/2} r^{d/2} \le r^d.$$

Since $r_0 \ge 1$ and $(2d-1)/d \ge 1$,

$$\log N = \min\left\{\log r, \frac{2d-1}{d}\log(1 \vee r/r_0)\right\} \ge \log(1 \vee r/r_0).$$

Both cases therefore give, for every $r \ge 2$,

$$\|Z\|_{L_r} \le C_d \left[\sqrt{r} + \frac{r^d}{\sqrt{L}\ell(1 \vee r/r_0)^d}\right], \qquad Z = \frac{S_A}{\sqrt{L}}. \tag{40}$$

Here $\mathbb{E}Z^2 = 1$, by independence and centering. For $Y = Z^2 - 1$, applying (40) at $r = 2q$, and using $\ell(1 \vee 2q/r_0) \ge \ell(1 \vee q/r_0)$, yields

$$\|Y\|_{L_q} \le C_d q(1 + \rho_L(q)), \qquad \rho_L(q) = \frac{q^{2d-1}}{L\ell(1 \vee q/r_0)^{2d}}, \qquad q \ge 2. \tag{41}$$

Restoring the original scale of A, the same Z is $\sqrt{M}(\Phi A)_1 / \|A\|_{\mathrm{F}}$.

4.2. **Averaging the measurements.** We use the following even-moment estimate to average the squared rows. See Latała [34] for general moment estimates for independent sums.

Lemma 4.1 (Moments of an independent sum)**.** *Let $Y_1, \dots, Y_M$ be independent identically distributed centered real random variables, and let $r \geq 2$ be even. If $\mathbb{E}|Y_1|^r < \infty$, put*

$$H = \max_{q \in \{2,\dots,r\}} \frac{r}{q} \left(\frac{M}{r}\right)^{1/q} \|Y_1\|_{L_q}.$$

Then $\left\|\sum_{m=1}^M Y_m\right\|_{L_r} \leq 2e^2 H$.

Proof. The case $H = 0$ is immediate. Otherwise, $\mathbb{E}|Y_1|^q \leq H^q (q/r)^q \, r/M$. Since the variables are centered, terms with an index of multiplicity one vanish in the multinomial expansion. Grouping the remaining terms by the number h of distinct indices and taking absolute values gives

$$\mathbb{E}\left|\sum_m Y_m\right|^r \leq H^r \frac{r!}{r^r} \sum_{h=1}^{\min(M, r/2)} \binom{M}{h} \left(\frac{r}{M}\right)^h \sum_{\substack{q_1 + \cdots + q_h = r \\ q_j \geq 2}} \prod_{j=1}^h \frac{q_j^{q_j}}{q_j!}.$$

The expansion is valid because r is even. Using $r! \leq r^r$, $q! \geq (q/e)^q$, and $\binom{M}{h}(r/M)^h \leq r^h/h!$, together with the bound 2^r on the number of compositions for each h, bounds the right-hand side by $H^r e^r 2^r \sum_{h \geq 0} r^h/h! = (2e^2 H)^r$. Taking the r-th root proves the claim. □

We now apply the lemma to the centered squared measurements $Y = Z^2 - 1$. Let $r \geq 2$ be even and $M \geq r$, and set $R = M/r \geq 1$. For independent copies of Y, Lemma 4.1 gives

$$\left\|\sum_{m=1}^M Y_m\right\|_{L_r} \leq C_d r \max_{2 \leq q \leq r} R^{1/q}(1 + \rho_L(q)).$$

The maximum may be taken over real q for this upper bound. For $q \leq r_0$, we have $\rho_L(q) \leq 1$. For $q \geq r_0$, the logarithmic derivative of $\rho_L(q) R^{1/q}$, with respect to q, has the sign of

$$(2d - 1) - \frac{2d}{1 + \log(q/r_0)} - \frac{\log R}{q}.$$

This expression is increasing in q. Thus the maximum on $[\max\{2, r_0\}, r]$, if this interval is nonempty, is attained at an endpoint. Its left endpoint contributes at most $C_d\sqrt{R}$: if $r_0 \geq 2$, then $\rho_L(r_0) = 1$; if $r_0 < 2$, then $\rho_L(2) \leq 2^{2d-1}$, since $L \geq 1$. It follows that

$$\max_{2 \leq q \leq r} R^{1/q}(1 + \rho_L(q)) \leq C_d \left[\sqrt{R} + \rho_L(r) R^{1/r}\right]. \tag{42}$$

For even $r \geq 4$, weighted Young's inequality gives

$$\rho_L(r) R^{1/r} \leq R^{1/2} + \rho_L(r)^{r/(r-2)} \leq R^{1/2} + C_d \rho_L(r).$$

For the last inequality, if $\rho_L(r) \leq 1$, raising it to a power larger than one only decreases it. Otherwise use $\rho_L(r) \leq r^{2d-1}$ and $\sup_{r \geq 4} r^{(4d-2)/(r-2)} < \infty$. The case $r = 2$ follows directly from (41) and independence. Since $r\rho_L(r) = \Gamma_d(L, r)$, division by M proves

$$\left\|\frac{1}{M}\sum_{m=1}^M (Z_m^2 - 1)\right\|_{L_r} \leq C_d \left[\sqrt{\frac{r}{M}} + \frac{\Gamma_d(L, r)}{M}\right], \qquad M \geq r. \tag{43}$$

Proof of Theorem 2.1. We have $\|\Phi A\|_2^2 / \|A\|_{\mathrm{F}}^2 = M^{-1}\sum_m Z_m^2$. Choose $r = 2\lceil \frac{1}{2}\log(e^2/\delta)\rceil$. Then $\log(e^2/\delta) \le r \le 2\log(e^2/\delta)$. Since the denominator in (24) is nondecreasing, $\Gamma_d(L_A, r) \le 2^{2d}\Gamma_d(L_A, \log(e^2/\delta))$. By enlarging C_d in (25), that condition implies $M \ge r$ and makes the right-hand side of (43) at most ε/e. Markov's inequality bounds the failure probability by $e^{-r} \le \delta/e^2 \le \delta$, proving (26). □

Proof of Corollary 2.2. For a row index m, write

$$Z_{m,i} = \frac{\sqrt{M}(\Phi A_i)_m}{\|A_i\|_{\mathrm{F}}}, \qquad Y_{m,ij} = Z_{m,i}Z_{m,j} - K(A_i, A_j).$$

The isotropy identity (19) gives $\mathbb{E}[Z_{m,i}Z_{m,j}] = K(A_i, A_j)$. Although different data points share the same random row, Hölder's inequality and (40) imply

$$\|Y_{m,ij}\|_{L_q} \le \|Z_{m,i}\|_{L_{2q}} \|Z_{m,j}\|_{L_{2q}} + 1 \le C_d q \left[1 + \max_{a\in[N_{\mathrm{dat}}]} \rho_{L_{A_a}}(q)\right].$$

The variables $Y_{m,ij}$ are independent over m. Choose $r = 2\lceil \frac{1}{2}\log(e^2 N_{\mathrm{dat}}^2/\delta)\rceil$. Interchanging the maximum over a with the moment-order maximum, the averaging argument leading to (43) gives, for $M \ge r$,

$$\left\| \frac{1}{M}\sum_{m=1}^{M} Y_{m,ij} \right\|_{L_r} \le C_d \left[\sqrt{\frac{r}{M}} + \frac{\max_a \Gamma_d(L_{A_a}, r)}{M}\right].$$

By the preceding rounding estimate, enlarging C_d in (27) ensures $M \ge r$ and makes the bound at most ε/e. Markov's inequality and a union bound over the ordered pairs give total failure probability at most $N_{\mathrm{dat}}^2 e^{-r} \le \delta$, proving (28). □

4.3. **The planted relation model.**

Proof of Proposition 2.3. We first verify the support and mass bounds, treating the planted box and background separately. For $k = d - |I| \in \{1, \dots, d-1\}$, a background slice has at most n^k possible entries and expected cardinality at most

$$\frac{n^k}{8\tau^{d-1}} = \frac{(n/\tau)^k}{8\tau^{d-1-k}} \le \frac{(n/\tau)^k}{8}.$$

A binomial Chernoff bound puts its cardinality below $(n/\tau)^k/2$, except with probability $e^{-c(n/\tau)^k}$. There are at most $2^d n^{d-1}$ proper slices. Since $\log n/(n/\tau) \le 2\log\tau/\tau$ for $n \ge \tau^2$ and $\tau \ge e$, a union bound shows that all background slices have weighted cardinality at most $1/2$, except with probability $e^{-cn/\tau}$, for sufficiently large τ depending only on d. The planted box adds at most $(pb)^k \le 8^{-k} \le 1/8$. The total background count has mean $\mu = (n^d - b^d)/(8\tau^{d-1})$, with

$$p^d\mu = \frac{\tau}{8} - \frac{(pb)^d}{8\tau^{d-1}}, \qquad \mu \asymp_d \frac{n^d}{\tau^{d-1}}.$$

Thus another Chernoff bound, together with $(pb)^d \le 8^{-d}$, gives $\tau/16 \le L \le \tau/4$ except with probability $2\exp\{-cn^d/\tau^{d-1}\}$. This is absorbed by the stated failure bound, proving (29).

Fix a realization of A satisfying (29). All remaining probabilities are over the sketch.

For the upper bound, apply Theorem 2.1 at failure probability δ, using $L \asymp \tau \asymp_d [\log(e^2/\delta)]^d$. Since $d \ge 2$,

$$\Gamma_d\left(L, \log\frac{e^2}{\delta}\right) \asymp_d \frac{\tau}{(\log\tau)^{2d}}, \qquad \log\frac{e^2}{\delta} = o\left(\frac{\tau}{(\log\tau)^{2d}}\right).$$

At error $1/2$, (25) therefore gives the asserted upper bound, uniformly in $n \ge \tau^2$.

For the lower bound, it suffices that the first squared normalized measurement exceeds $3M/2$. Since the remaining squares are nonnegative, this forces a relative squared-norm error greater than $1/2$. Set $T = [b]$, so the planted box is T^d. Let E_k be the event that exactly k variables in T equal $+1$ in each mode, with all other variables in T zero. For large τ, $1/16 \le bp \le 1/8$ and $p \le 1/\tau$. Hence, whenever $1 \le k \le b/2$,

$$\mathbb{P}_\Phi(E_k) = \left[\binom{b}{k}(p/2)^k(1-p)^{b-k}\right]^d \ge \exp\{-C_d k \log(ek)\}.$$

Let $\mathcal{F}_T$ reveal all factor variables with indices in T. On E_k, the unnormalized first row is $S_A = k^d + U$. Every term of U contains an outside variable, so $\mathbb{E}(U \mid \mathcal{F}_T) = 0$. Decompose U according to the nonempty set J of modes whose coordinates lie outside T. Fix such a set with $j = |J| < d$. Conditional on $\mathcal{F}_T$, on the event E_k its contribution has the form

$$U_J = \sum_{\boldsymbol{i}_J \in ([n]\setminus T)^J} c_{\boldsymbol{i}_J} \prod_{s\in J} \xi^{(s)}_{1 i_s}.$$

The coefficient $c_{\boldsymbol{i}_J}$ counts the selected inside tuples that extend $\boldsymbol{i}_J$ to a nonzero entry of A. Hence

$$0 \le c_{\boldsymbol{i}_J} \le k^{d-j}, \qquad \sum_{\boldsymbol{i}_J} c_{\boldsymbol{i}_J} \le k^{d-j}p^{-j}.$$

The second inequality applies the slice-count condition to each of the k^{d-j} selected inside tuples. Distinct outside monomials are orthogonal, and therefore

$$\mathbb{E}(U_J^2 \mid \mathcal{F}_T) = p^j \sum_{\boldsymbol{i}_J} c_{\boldsymbol{i}_J}^2 \le p^j \max_{\boldsymbol{i}_J} c_{\boldsymbol{i}_J} \sum_{\boldsymbol{i}_J} c_{\boldsymbol{i}_J} \le k^{2(d-j)}.$$

Contributions from distinct sets J are also orthogonal. The part with every mode outside T has conditional variance at most L. Consequently, on E_k,

$$\operatorname{Var}(U \mid \mathcal{F}_T) \le L + \sum_{j=1}^{d-1} \binom{d}{j} k^{2(d-j)} \le L + 2^d k^{2d-2}.$$

Set $k = \lceil (64LM)^{1/(2d)} \rceil$. On E_k, conditional Chebyshev gives

$$\mathbb{P}_\Phi\left(U < -k^d/2 \mid \mathcal{F}_T\right) \le \frac{4L}{k^{2d}} + \frac{2^{d+2}}{k^2} \le \frac{1}{16M} + \frac{2^{d+2}}{k^2}.$$

Since $k^2 \ge c_d \tau^{1/d}$, this is at most $1/2$ for sufficiently large τ. Thus $\mathbb{P}_\Phi\{S_A \ge k^d/2 \mid E_k\} \ge 1/2$. On this event,

$$\frac{\|\Phi A\|_2^2}{\|A\|_{\mathrm{F}}^2} \ge \frac{S_A^2}{ML} \ge \frac{k^{2d}}{4ML} \ge 16.$$

If $M \le c_d \tau/(\log\tau)^{2d}$, choosing c_d small enough gives

$$k \le C_d c_d^{1/(2d)} \tau^{1/d}/\log\tau + 1 \le b/2, \qquad C_d k \log(ek) \le \tau^{1/d}/2 \le \tfrac{1}{2}\log(e^2/\delta).$$

Here $b \ge n/(16\tau) \ge \tau/16$, so the constants are independent of n. In fact, we obtain the stronger estimate

$$\mathbb{P}_\Phi\left(\left|\frac{\|\Phi A\|_2^2}{\|A\|_{\mathrm{F}}^2} - 1\right| > \tfrac{1}{2}\right) \ge \frac{\sqrt{\delta}}{2e} > \delta, \tag{44}$$

where the strict inequality holds for sufficiently large τ, since $\log(e^2/\delta) \ge \tau^{1/d} > 4 + 2\log 2$. This proves the lower bound. □

Proof of Corollary 2.4. For large n, the integer $\tau = \lfloor [\log(e^2 n^a)]^d \rfloor$ satisfies $n \geq \tau^2$ and $\tau^{1/d} \leq \log(e^2 n^a) \leq 2\tau^{1/d}$. Apply Proposition 2.3 with $\delta = n^{-a}$. Its sufficient and necessary row counts have the order in (30), since $\tau \asymp_{d,a} (\log n)^d$ and $\log \tau \asymp_{d,a} \log\log n$. □

4.4. **The comparison bounds.** We give the derivation of (31) so that the averaging step is identical in the two bounds being compared. Rescale the tensor so that $\|A\|_\infty = 1$, and set $L_A = p^d \|A\|_{\mathrm{F}}^2$. Let D_I be the largest number of support entries consistent with fixed coordinates in the nonempty proper mode set I. Assume $p^{d-|I|} D_I \leq C_0$ for all such I, and $L_A \geq 1$.

The variables in S_A are centrally moment bounded with parameter one. The smoothness parameters in Schudy and Sviridenko [44, Theorem 1.3] satisfy

$$\mathrm{Var}(S_A) = L_A, \qquad \mu_j \leq \max_{|I|=j} p^{d-j} D_I \leq C_0 \quad (1 \leq j < d), \qquad \mu_d = 1.$$

A set of polynomial variables with two indices from the same mode belongs to no monomial and contributes zero to the maximum defining μ_j. Integrating that theorem's tail bound yields

$$\|S_A\|_{L_r} \leq C_{d,C_0}\big(\sqrt{rL_A} + r^d\big), \qquad r \geq 2.$$

The same row estimate follows from the uncorrected partition bound, using the codegree estimates in the proof of the new theorem. For $Z = S_A/\sqrt{L_A}$ and $Y = Z^2 - 1$, it follows that

$$\|Y\|_{L_q} \leq C_{d,C_0}\left(q + \frac{q^{2d}}{L_A}\right), \qquad q \geq 2. \tag{45}$$

Let $r \geq 2$ be even and $M \geq r$. Put $R = M/r$. By Lemma 4.1,

$$\left\| \frac{1}{M} \sum_{m=1}^{M} Y_m \right\|_{L_r} \leq \frac{C_{d,C_0}}{R} \max_{2 \leq q \leq r} R^{1/q} \left(1 + \frac{q^{2d-1}}{L_A}\right).$$

The endpoint argument used in (42) applies here with $\rho_L(q)$ replaced by q^{2d-1}/L_A. Using $L_A \geq 1$, the same Young inequality gives, for even $r \geq 4$,

$$\left\| \frac{1}{M} \sum_{m=1}^{M} Y_m \right\|_{L_r} \leq C_{d,C_0} \left(\sqrt{\frac{r}{M}} + \frac{r^{2d}}{ML_A}\right).$$

The case $r = 2$ follows directly from (45). Rounding the confidence order up to the next even integer as in the proof of Theorem 2.1, and then applying Markov's inequality at a constant multiple of this bound, proves (31). For overlap-kernel entries, Hölder's inequality gives the same centered-product moment curve whenever the two weighted masses are at least L. The union bound changes only the chosen confidence order r.

For the distribution-only comparison, let $Z = \sqrt{M}(\Phi A)_1 / \|A\|_{\mathrm{F}}$, so $\mathbb{E}Z^2 = 1$. The normalized factor $\xi/\sqrt{p}$ has sub-Gaussian norm at most $C\theta_p$, where $\theta_p^2 = [p \log(1+1/p)]^{-1}$. Iterating the sub-Gaussian moment bound over the modes gives $\|Z\|_{L_r} \leq C_d \theta_p^d r^{d/2}$. Hence $\|Z^2 - 1\|_{L_q} \leq C_d \theta_p^{2d} q^d$. The same endpoint calculation, now with exponent $d - 1$, gives

$$\left\| \frac{1}{M} \sum_{m=1}^{M} (Z_m^2 - 1) \right\|_{L_r} \leq C_d \theta_p^{2d} \left(\sqrt{\frac{r}{M}} + \frac{r^d}{M}\right).$$

Taking r at the confidence level and applying Markov's inequality gives the stated distribution-only row bound.

Appendix A. Moment bounds beyond the bounded decoupled setting

A.1. **Bounded non-decoupled chaoses.** For a diagonal-free polynomial, decoupling replaces the repeated input vector X by independent copies $X^{(1)}, \dots, X^{(d)}$, at a cost depending only on the degree. Throughout this subsection, evaluate $\mathcal{S}_{I,\mathcal{J}}(A)$, $B_k(A;r)$, and $\lambda_A(r)$ with $p_i^{(s)} = p_i$.

Proposition A.1 (Bounded non-decoupled moment bound)**.** *Let $A = (a_{i_1,\dots,i_d})$ be a real diagonal-free d-tensor, i.e., $a_{i_1,\dots,i_d} = 0$ whenever two indices coincide. Let $X_1, \dots, X_n$ be independent, with*

$$\mathbb{E}X_i = 0, \qquad |X_i| \le 1 \text{ a.s.}, \qquad \mathbb{E}X_i^2 \le p_i, \qquad 0 < p_i \le 1,$$

and define

$$Q_A(X) \coloneqq \sum_{i_1,\dots,i_d=1}^{n} a_{i_1,\dots,i_d} X_{i_1} \cdots X_{i_d}.$$

Writing $m_A \coloneqq \|A\|_\infty$, for every $r \ge 2$ we have

$$\|Q_A(X)\|_{L_r} \lesssim_d \max_{1\le k\le d} B_k(A;r) + \frac{r^d m_A}{\ell(\lambda_A(r))^d}. \tag{46}$$

When $\lambda_A(r) > 1$, the first term is absorbed into the second, as in (6)*.*

Proof. Let $X^{(1)}, \dots, X^{(d)}$ be independent copies of X. By the decoupling inequality (37),

$$\|Q_A(X)\|_{L_r} \lesssim_d \left\| Z_A(X^{(1)}, \dots, X^{(d)}) \right\|_{L_r}.$$

The copies retain the variance bounds p_i, so Theorem 1.1 and (6) give both claims. Forward decoupling requires no symmetry of A. □

For $d = 2$, this is the quadratic form $X^{\mathsf{T}}AX$ with zero diagonal, as used in Corollary 1.4.

A.2. **Sparse ψ_α inputs.** Let $1 < \alpha \le \infty$, with $1/\infty \coloneqq 0$, and put $\beta \coloneqq 1 - 1/\alpha \in (0, 1]$. For finite α, use the standard Orlicz norm

$$\|\chi\|_{\psi_\alpha} \coloneqq \inf\left\{K > 0 : \mathbb{E}\exp\big((|\chi|/K)^\alpha\big) \le 2\right\}; \qquad \|\chi\|_{\psi_\infty} \coloneqq \|\chi\|_{L_\infty}.$$

The cases $\alpha = 2$ and $\alpha = \infty$ correspond to sub-Gaussian and bounded amplitudes.

Proposition A.2 (Bennett-type bound for sparse sub-exponential chaoses)**.** *Let $K > 0$ and $X_i^{(s)} \coloneqq \delta_i^{(s)} \chi_i^{(s)}$, where all Bernoulli random variables and amplitudes are mutually independent and*

$$\delta_i^{(s)} \sim \text{Bernoulli}(p_i^{(s)}), \qquad 0 < p_i^{(s)} \le 1, \qquad \mathbb{E}\chi_i^{(s)} = 0, \qquad \left\|\chi_i^{(s)}\right\|_{\psi_\alpha} \le K.$$

For a real rectangular tensor A, evaluate $\mathcal{S}_{I,\mathcal{J}}(A)$, $B_k(A;r)$, and $\lambda_A(r)$ using the Bernoulli probabilities, and write $m_A \coloneqq \|A\|_\infty$. Then, for every $r \ge 2$,

$$\|Z_A(X)\|_{L_r} \lesssim_d K^d \left[\max_{1\le k\le d} B_k(A;r) + \frac{r^d m_A}{\ell(\lambda_A(r))^{d\beta}}\right]. \tag{47}$$

When $\lambda_A(r) > 1$,

$$\|Z_A(X)\|_{L_r} \lesssim_d \frac{K^d r^d m_A}{\ell(\lambda_A(r))^{d\beta}}. \tag{48}$$

The constants are uniform over $1 < \alpha \le \infty$. The same bounds hold for $Q_A(X)$ with diagonal-free A and $X_i \coloneqq \delta_i \chi_i$, where the selectors and amplitudes are mutually independent, $\delta_i \sim$ Bernoulli(p_i), $\mathbb{E}\chi_i = 0$, and $\|\chi_i\|_{\psi_\alpha} \le K$. Use $p_i^{(s)} = p_i$ in the weighted norms.

Proof. By homogeneity, first take $K = 1$. For finite $\alpha > 1$ and integers $m \geq 2$, the Orlicz assumption gives $\mathbb{E}|\chi_i^{(s)}|^{\alpha m} \leq 2m!$. Monotonicity of moment norms gives

$$\mathbb{E}|X_i^{(s)}|^m = p_i^{(s)}\mathbb{E}|\chi_i^{(s)}|^m \leq p_i^{(s)}(2m!)^{1/\alpha} \leq 2p_i^{(s)}(m!)^{1-\beta}. \tag{49}$$

For $\alpha = \infty$, boundedness gives $\mathbb{E}|X_i^{(s)}|^m \leq p_i^{(s)}$, so the same bound holds.

Use the symmetrized inputs $V = (X - X')/2$ and independent replica sums $S_i^{(s)} = \sum_{u=1}^N Y_{i,u}^{(s)}$ from Lemma 3.2, with integer $N \geq 1$. Convexity gives $\mathbb{E}|V_i^{(s)}|^m \leq \mathbb{E}|X_i^{(s)}|^m$, while $\mathbb{E}(S_i^{(s)})^m \geq Np_i^{(s)}m!$ for even $m \geq 2$. Consequently,

$$\frac{\mathbb{E}|V_i^{(s)}|^m}{\mathbb{E}(S_i^{(s)})^m} \leq \frac{2}{N(m!)^\beta} \leq \frac{2e^\beta}{\ell(N)^{m\beta}} \leq \left(\frac{\sqrt{2e}}{\ell(N)^\beta}\right)^m,$$

where $N(m!)^\beta \geq (Nm!)^\beta \geq e^{-\beta}\ell(N)^{m\beta}$. The symmetrization and coordinate-replacement argument of Lemma 3.2, followed by Lemmas 3.3 and 3.4, therefore gives, after restoring K,

$$\|Z_A(X)\|_{L_r} \lesssim_d K^d \frac{\max_{I\subseteq[d]} N^{|I^c|/2}\Delta_I(A;r)}{\ell(N)^{d\beta}}. \tag{50}$$

Rounding as in Section 3.2.3 extends this estimate to real $r \geq 2$ and $N \geq 1$. Choosing $N = \lambda_A(r)$ as in Theorem 1.1 proves both bounds. The constants are uniform in α, since the scalar comparison uses $\sqrt{2e}$ and $0 < \beta \leq 1$. For $Q_A(X)$, first apply (37); the copies retain p_i and K. □

Remark A.3 (Bounded inputs without a Bernoulli factor). A Bernoulli representation is unnecessary when the same scalar moment bounds hold. Independent centered inputs satisfying

$$|X_i^{(s)}| \leq K \quad \text{a.s.}, \qquad \mathbb{E}(X_i^{(s)})^2 \leq K^2 p_i^{(s)}$$

obey $\mathbb{E}|X_i^{(s)}|^m \leq K^{m-2}\mathbb{E}(X_i^{(s)})^2 \leq K^m p_i^{(s)}$. The proof therefore applies with $\beta = 1$, recovering Theorem 1.1 when $K = 1$.

The moment bound also yields a tail estimate in which the logarithmic gain has power β.

Proposition A.4 (Tail bound in the logarithmic regime)**.** *Let η_A be defined by* (7)*, using the Bernoulli probabilities. Under the assumptions of Proposition A.2, for $t > 0$ with $\eta_A(t/K^d) \geq e$,*

$$\mathbb{P}\big(|Z_A(X)| \geq t\big) \leq 2\exp\left[-c_d\left(\frac{t}{K^d m_A}\right)^{1/d} \ell\big(\eta_A(t/K^d)\big)^\beta\right]. \tag{51}$$

The same bound holds for $Q_A(X)$ under the non-decoupled assumptions. Thus the logarithm enters the tail rate with power $1/2$ for sub-Gaussian amplitudes and power one for bounded amplitudes.

Proof of Proposition A.4. We adapt the proof of Corollary 1.2, putting

$$w = \left(\frac{t}{K^d m_A}\right)^{1/d}, \qquad N = \eta_A(t/K^d), \qquad \Lambda = \ell(N).$$

Since $N > 1$, its definition gives, for every $I, \mathcal{J}$,

$$K^d N^{|I^c|/2}\mathcal{S}_{I,\mathcal{J}}(A)w^{\kappa_{I,\mathcal{J}}} \leq t,$$

Choose $r = cw\Lambda^\beta$, with $0 < c \le 1$ depending only on d. If $r \ge 2$, (50) gives

$$\|Z_A(X)\|_{L_r} \le C_d t \max_{I,\mathcal{J}} c^{\kappa_{I,\mathcal{J}}} \Lambda^{\beta(\kappa_{I,\mathcal{J}}-d)} \le C_d \sqrt{c}\, t,$$

because $1/2 \le \kappa_{I,\mathcal{J}} \le d$. For sufficiently small c, Markov's inequality gives the claimed rate; when $r < 2$, decrease c_d to make the bound trivial. For $Q_A(X)$, apply (37) first and absorb its constant into c, leaving $\eta_A(t/K^d)$ unchanged. □

Appendix B. General embeddings

The support-count condition gives an explicit row bound for a class of tensors. For arbitrary fixed tensors, we use the weighted partition norms in the row moment curve. We give this version below, together with its finite-set and subspace consequences. Throughout this appendix, n and p are arbitrary and the sketch is independent of the input data.

We first define a bound for the moments of one normalized measurement. For $x \ne 0$ and $r \ge 2$, set

$$\mathcal{R}_x(r) \coloneqq 1 + \min\left\{ \left(\frac{r}{p}\right)^{d/2}, \ \frac{1}{p^{d/2}\|x\|_{\mathrm{F}}} \inf_{N\ge 1} \frac{\max_{I\subseteq[d]} N^{|I^c|/2}\Delta_I(x;r)}{\ell(N)^d} \right\}. \tag{52}$$

Here Δ_I is defined in (3) with all second-moment bounds equal to p. The first expression follows from Khintchine's inequality; the second includes the weighted slice norms and the logarithmic correction. Up to a constant depending only on d, each expression bounds $\left\|\sqrt{M}(\Phi x)_1\right\|_{L_r} / \|x\|_{\mathrm{F}}$, as shown below. To control the error after averaging M squared measurements, for an even integer $r \ge 2$, set

$$\mathcal{E}_x(M,r) \coloneqq \frac{1}{M} \max_{q\in\{2,\dots,r\}} \frac{r}{q} \left(\frac{M}{r}\right)^{1/q} \mathcal{R}_x(2q)^2. \tag{53}$$

The expression $\mathcal{E}_x(M,r)$ results from applying Lemma 4.1 to the centered normalized squared measurements $M(\Phi x)_m^2/\|x\|_{\mathrm{F}}^2 - 1$, whose L_q-norms are bounded by $C_d\mathcal{R}_x(2q)^2$. Each term decreases with M, so the sufficient conditions below can be checked by increasing the row count. The replication parameter N in (52) may be chosen separately at each moment order. In particular, no single upper bound on $\mathcal{R}_x(r)/\sqrt{r}$ is needed over the whole range.

Proposition B.1 (Sparse Khatri–Rao embedding)**.** *There is a constant C_d with the following properties. Fix $0 < \varepsilon \le 1$ and $0 < \delta < 1$.*

(1) *For every fixed $x \ne 0$ and every integer $M \ge 1$,*

$$\mathbb{P}\left(\left| \frac{\|\Phi x\|_2^2}{\|x\|_{\mathrm{F}}^2} - 1 \right| > C_d \mathcal{E}_x\left(M, 2\left\lceil \frac{1}{2}\log\frac{e^2}{\delta}\right\rceil\right)\right) \le \delta. \tag{54}$$

Consequently,

$$C_d \mathcal{E}_x\left(M, 2\left\lceil \frac{1}{2}\log\frac{e^2}{\delta}\right\rceil\right) \le \varepsilon \tag{55}$$

is sufficient for relative squared-norm error at most ε, with failure probability at most δ.

(2) *For a fixed set $\mathcal{X} = \{x_1, \dots, x_m\}$ of distinct tensors, with $m \ge 2$, the condition*

$$C_d \max_{i<j} \mathcal{E}_{x_i - x_j}\left(M, 2\left\lceil \frac{1}{2}\log\frac{e^2 m^2}{\delta}\right\rceil\right) \le \varepsilon$$

implies, with probability at least $1-\delta$,

$$(1-\varepsilon)\|x_i - x_j\|_{\mathrm{F}}^2 \le \|\Phi x_i - \Phi x_j\|_2^2 \le (1+\varepsilon)\|x_i - x_j\|_{\mathrm{F}}^2 \quad \text{for all } i < j.$$

(3) *For a fixed k-dimensional subspace W, with $k \geq 1$, the condition*

$$C_d \sup_{x \in W \setminus \{0\}} \mathcal{E}_x\left(M, 2\left\lceil \frac{1}{2} \log \frac{e^2 9^k}{\delta} \right\rceil\right) \leq \varepsilon$$

implies, with probability at least $1-\delta$,

$$(1-\varepsilon)\|x\|_{\mathrm{F}}^2 \leq \|\Phi x\|_2^2 \leq (1+\varepsilon)\|x\|_{\mathrm{F}}^2 \quad \textit{for all } x \in W.$$

The finite-set condition involves pairwise differences. The subspace condition is uniform over W, and therefore also covers vectors chosen after the sketch has been observed.

Proof. Put

$$Z_m(x) = \sqrt{M}(\Phi x)_m = p^{-d/2} \sum_{\boldsymbol{i}} x_{\boldsymbol{i}} \prod_{s=1}^{d} \xi_{mi_s}^{(s)}.$$

Independence and centering give $\mathbb{E}[Z_m(x)Z_m(y)] = \langle x, y\rangle$. Applying Proposition 3.1 for every $N \geq 1$ gives the second bound in (52). For the first, write $\xi_{mi}^{(s)} = b_{mi}^{(s)}\sigma_{mi}^{(s)}$, where the $b_{mi}^{(s)}$ are independent Bernoulli(p) variables and the $\sigma_{mi}^{(s)}$ are independent Rademacher variables, independent of the selectors. Conditional on the selectors, repeated Khintchine inequalities and Minkowski's inequality give

$$\|Z_m(x)\|_{L_r(\text{signs})} \leq p^{-d/2} r^{d/2} \left(\sum_{\boldsymbol{i}} x_{\boldsymbol{i}}^2 \prod_{s=1}^{d} b_{mi_s}^{(s)}\right)^{1/2} \leq p^{-d/2} r^{d/2} \|x\|_{\mathrm{F}}.$$

This conditional bound is uniform in the Bernoulli variables. Combining the two estimates yields

$$\|Z_m(x)\|_{L_r} \leq C_d \mathcal{R}_x(r) \|x\|_{\mathrm{F}}. \tag{56}$$

Consequently, since $\mathcal{R}_x(r) \geq 1$,

$$\left\| Z_m(x)^2 - \|x\|_{\mathrm{F}}^2 \right\|_{L_q} \leq C_d \mathcal{R}_x(2q)^2 \|x\|_{\mathrm{F}}^2. \tag{57}$$

For every even integer $r \geq 2$, Lemma 4.1 and Markov's inequality bound the probability that the relative squared-norm error exceeds $C_d \mathcal{E}_x(M, r)$ by e^{-r}, after absorbing the Markov factor e into C_d. Taking $r = 2\lceil \frac{1}{2}\log(e^2/\delta)\rceil$ proves (54), since $e^{-r} \leq \delta/e^2 \leq \delta$. For the finite-set claim, take $r = 2\lceil \frac{1}{2}\log(e^2 m^2/\delta)\rceil$ and use a union bound over fewer than m^2 differences.

For the subspace claim, let $\mathcal{N}$ be a $1/4$-net of the unit sphere of W, with $|\mathcal{N}| \leq 9^k$. Apply the preceding even-moment bound on this net with $r = 2\lceil \frac{1}{2}\log(e^2 9^k/\delta)\rceil$. For the self-adjoint operator $\Theta = (\Phi|_W)^*(\Phi|_W) - I_W$, the usual net argument gives $\|\Theta\|_{\mathrm{op}} \leq 2\max_{x \in \mathcal{N}} |\langle \Theta x, x\rangle|$. Indeed, approximate a unit vector attaining the operator norm by a net point; the two resulting error terms sum to at most $\frac{1}{2}\|\Theta\|_{\mathrm{op}}$. A union bound and an enlargement of C_d complete the proof. □

We next check that the general bound recovers the standard row count for dense Rademacher tensor-product maps, where $p = 1$. In this case, (52) gives $\mathcal{R}_x(2q)^2 \leq C_d q^d$ for every tensor x. For an even integer $r \geq 2$, if $M \geq r$, then

$$\mathcal{E}_x(M, r) \leq C_d \left(\sqrt{\frac{r}{M}} + \frac{r^d}{M}\right). \tag{58}$$

This follows from the endpoint argument in (42) and the subsequent Young inequality, now with the power q^{d-1}; the case $r = 2$ is immediate. Taking $r = 2\lceil \frac{1}{2}\log(e^2/\delta)\rceil$ and enlarging C_d to absorb the rounding shows that

$$M \geq C_d \left[\varepsilon^{-2} \log \frac{e^2}{\delta} + \varepsilon^{-1} \left(\log \frac{e^2}{\delta} \right)^d \right]$$

is sufficient for relative squared-norm error at most ε for a fixed tensor, with failure probability at most δ, matching the dependence in [5, Theorem 2]. The first term ensures $M \geq r$.

Acknowledgements

GPT-5.6 assisted in developing the proof. The authors independently verified the argument and take full responsibility for the final proof. K.W. is supported by Hong Kong RGC grant GRF 16305526. Y.Z. was partially supported by the Simons Grant MPS-TSM-00013944 and NSF DMS-2606337. This work was carried out while Y.Z. was visiting the Simons Institute for the Theory of Computing during the Spectral Theory Beyond Graphs program in Fall 2026.

References

[1] R. Adamczak. A note on the Hanson-Wright inequality for random vectors with dependencies. *Electronic Communications in Probability*, 72:1–13, 2015.

[2] R. Adamczak and R. Latała. Tail and moment estimates for chaoses generated by symmetric random variables with logarithmically concave tails. *Annales de l'IHP Probabilités et statistiques*, 48(4):1103–1136, 2012.

[3] R. Adamczak, R. Latała, and R. Meller. Moments of Gaussian chaoses in Banach spaces. *Electronic Journal of Probability*, 26(11):1–36, 2021.

[4] R. Adamczak and P. Wolff. Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probability Theory and Related Fields*, 162:531–586, 2015.

[5] T. D. Ahle and J. B. T. Knudsen. Almost optimal tensor sketch. *arXiv preprint arXiv:1909.01821*, 2019.

[6] S. Bamberger, F. Krahmer, and R. Ward. The Hanson–Wright inequality for random tensors. *Sampling Theory, Signal Processing, and Data Analysis*, 20(2):14, 2022.

[7] G. Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.

[8] L. Beretta and C. Musco. A tight analysis of Khatri–Rao oblivious subspace embeddings. *arXiv preprint arXiv:2608.28094v2*, 2026.

[9] S. Boucheron, O. Bousquet, G. Lugosi, and P. Massart. Moment inequalities for functions of independent random variables. *The Annals of Probability*, 33(2):514–560, 2005.

[10] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford, 2013.

[11] Z. Bujanović, L. Grubišić, D. Kressner, and H. Y. Lam. Subspace embedding with random Khatri–Rao products and its application to eigensolvers. *IMA Journal of Numerical Analysis*, 46(3):1328–1355, 2026.

[12] S. Chatterjee. The missing log in large deviations for triangle counts. *Random Structures & Algorithms*, 40(4):437–451, 2012.

[13] S. Chatterjee. An introduction to large deviations for random graphs. *Bulletin of the American Mathematical Society*, 53(4):617–642, 2016.

[14] K. Chen and R. Jin. Tensor-structured sketching for constrained least squares. *SIAM J. Matrix Anal. Appl.*, 42(4):1703–1731, 2021.

[15] X. Chen and Y. Yang. Hanson–Wright inequality in Hilbert spaces with application to K-means clustering for non-Euclidean data. *Bernoulli*, 27(1):586–614, 2021.

[16] G. Dai, Y. He, K. Wang, and Y. Zhu. A note on the improved sparse Hanson–Wright inequalities. arXiv preprint arXiv:2505.20799, 2025.

[17] G. Dai and K. Wang. Sparsity-adaptive concentration inequalities for random polynomials. *arXiv preprint arXiv:2606.24090v1*, 2026.

[18] V. H. de la Peña and S. J. Montgomery-Smith. Decoupling inequalities for the tail probabilities of multivariate U-statistics. *The Annals of Probability*, 23(2):806–816, 1995.

[19] R. DeMarco and J. Kahn. Tight upper tail bounds for cliques. *Random Structures & Algorithms*, 41(4):469–487, 2012.
[20] A. Dhawan, C. Mao, and A. S. Wein. Detection of dense subhypergraphs by low-degree polynomials. *Random Structures & Algorithms*, 66(1):e21279, 2025.
[21] E. Giné, R. Latała, and J. Zinn. Exponential and moment inequalities for U-statistics. In *High Dimensional Probability II*, pages 13–38. Springer, 2000.
[22] E. D. Gluskin and S. Kwapień. Tail and moment estimates for sums of independent random variables with logarithmically concave tails. *Studia Mathematica*, 114(3):303–309, 1995.
[23] F. Götze, H. Sambale, and A. Sinulis. Concentration inequalities for polynomials in α-sub-exponential random variables. *Electronic Journal of Probability*, 26:1–22, 2021.
[24] D. L. Hanson and F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3):1079–1083, 1971.
[25] M. Harel, F. Mousset, and W. Samotij. Upper tails via high moments and entropic stability. *Duke Mathematical Journal*, 171(10):2089–2192, 2022.
[26] Y. He, K. Wang, and Y. Zhu. Sparse Hanson–Wright inequalities with applications. *Electron. J. Probab.*, 31(39):1–49, 2026.
[27] P. Hitczenko, S. J. Montgomery-Smith, and K. Oleszkiewicz. Moment inequalities for sums of certain independent symmetric random variables. *Studia Math*, 123(1):15–42, 1997.
[28] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58(301):13–30, 1963.
[29] S. Janson and A. Ruciński. The infamous upper tail. *Random Structures & Algorithms*, 20(3):317–342, 2002.
[30] J. H. Kim and V. H. Vu. Concentration of multivariate polynomials and its applications. *Combinatorica*, 20(3):417–434, 2000.
[31] Y. Klochkov and N. Zhivotovskiy. Uniform Hanson-Wright type concentration inequalities for unbounded entries via the entropy method. *Electronic Journal of Probability*, 25(22):1–30, 2020.
[32] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.
[33] K. Kolesko and R. Latała. Moment estimates for chaoses generated by symmetric random variables with logarithmically convex tails. *Statistics & Probability Letters*, 107:210–214, 2015.
[34] R. Latała. Estimation of moments of sums of independent real random variables. *The Annals of Probability*, 25(3):1502–1513, 1997.
[35] R. Latała. Tail and moment estimates for some types of chaos. *Studia mathematica*, 135:39–53, 1999.
[36] R. Latała. Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.*, 34(6):2315–2331, 2006.
[37] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
[38] E. Lubetzky and Y. Zhao. On the variational problem for upper tails in sparse random graphs. *Random Structures & Algorithms*, 50(3):420–436, 2017.
[39] S. J. Montgomery-Smith. The distribution of Rademacher sums. *Proceedings of the American Mathematical Society*, 109(2):517–522, 1990.
[40] S. Park, X. Wang, and J. Lim. Sparse Hanson–Wright inequality for a bilinear form of sub-Gaussian variables. *Stat*, 12(1):e539, 2023.
[41] A. Rahimi and B. Recht. Random features for large-scale kernel machines. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007.
[42] B. Rakhshan and G. Rabusseau. Tensorized random projections. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 3306–3316. PMLR, 2020.
[43] M. Rudelson and R. Vershynin. Hanson–Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18(82):1–9, 2013.
[44] W. Schudy and M. Sviridenko. Bernstein-like concentration and moment inequalities for polynomials of independent random variables: multilinear case. *arXiv preprint arXiv:1109.5193*, 2011. Version 2, revised June 8, 2012.
[45] W. Schudy and M. Sviridenko. Concentration and moment inequalities for polynomials of independent random variables. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 437–446. SIAM, 2012.
[46] Y. Sun, Y. Guo, J. A. Tropp, and M. Udell. Tensor random projection for low memory dimension reduction. *arXiv preprint arXiv:2105.00105*, 2021.

[47] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.
[48] R. Vershynin. Concentration inequalities for random tensors. *Bernoulli*, 26(4):3139–3162, 2020.
[49] V. Vu and K. Wang. Random weighted projections, random quadratic forms and random eigenvectors. *Random Structures & Algorithms*, 47(4):792–821, 2015.
[50] S. Zhou. Sparse Hanson–Wright inequalities for subgaussian quadratic forms. *Bernoulli*, 25(3):1603–1639, 2019.
[51] S. Zhou. Concentration of measure bounds for matrix-variate data with missing values. *Bernoulli*, 30(1):198–226, 2024.

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
Email address: guozhengdai@ust.hk

Department of Mathematics, University of California, San Diego, La Jolla, CA 92093 USA
Email address: yih130@ucsd.edu

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
Email address: kewang@ust.hk

Department of Mathematics, University of Southern California, 3620 Vermont Avenue, Los Angeles, CA 90089, USA
Email address: yizhezhu@usc.edu