Sparsity-adaptive concentration inequalities for random polynomials

Guozheng Dai* and Ke Wang†

June 24, 2026

Abstract

We prove concentration inequalities for polynomials of independent, sparse α-sub-exponential random variables. Specifically, we consider $X_i = \delta_i \xi_i$, where the Bernoulli selectors δ_i are independent with parameters p_i, and the variables ξ_i are independent α-sub-exponential random variables (not necessarily centered). For any polynomial $f : \mathbb{R}^n \to \mathbb{R}$ of degree at most D and any $0 < \alpha \leq 1$, we establish an L_r-moment bound for $f(X) - \mathbb{E}f(X)$ in terms of partition norms of sparsity-weighted expected derivative tensors. The weights count distinct coordinates rather than multiplicities and therefore distinguish diagonal, partially diagonal, and off-diagonal contributions. This captures the sparse scaling in both collective fluctuation regimes and extreme-coordinate regimes.

When all sparsity parameters are equal to one, our result recovers the polynomial concentration inequality of Götze, Sambale, and Sinulis. In degree two, it recovers sparse Hanson-Wright bounds. As applications, we derive deviation inequalities for the distance between a sparse simple random tensor and a fixed subspace, and obtain lower bounds for the smallest singular value of matrices whose columns are independent sparse simple random tensors.

1 Introduction

Concentration inequalities are a fundamental tool in probability theory and its applications. A classical example is the Gaussian concentration inequality [6]: if G is a standard Gaussian vector in $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ is 1-Lipschitz, then for every $t > 0$,

$$\mathbb{P}\big\{|f(G) - \mathbb{E}f(G)| \geq t\big\} \leq 2e^{-t^2/2}.$$

Such inequalities play a central role in the study of Gaussian processes, high-dimensional geometry, random matrices, and related areas. However, many natural functions arising in probability, combinatorics, and theoretical computer science are not Lipschitz. A basic class of such examples is given by polynomials of independent random variables. For instance, subgraph counts in random graphs can be written as polynomial functions of independent edge indicators;

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

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

their concentration and large-deviation behavior have been extensively studied (see, for instance, [22, 9, 10, 31]).

This paper studies concentration inequalities for random polynomials with sparse independent inputs. Since tail bounds can often be obtained from L_r-moment estimates through Markov's inequality, much of the literature on polynomial concentration is formulated in terms of sharp bounds for L_r-norms; see, for example, Proposition 3.3 in [16]. We first recall several lines of work most relevant to the present paper.

The most classical case is the linear form $\sum_{i=1}^n a_i\xi_i$, where $a = (a_i)_{i\le n}$ is a fixed vector and $\{\xi_i\}_{i=1}^n$ is a sequence of independent random variables. Montgomery-Smith [32] gave two-sided concentration bounds for $\sum_{i=1}^n a_i\xi_i$ when ξ_i are Rademacher variables (i.e. $\xi_i = \pm 1$ with equal probability), Gluskin and Kwapień [15] obtained the optimal L_r-norm estimate under log-concave tails, and Hitczenko, Montgomery-Smith and Oleszkiewicz [21] derived the optimal bound for the log-convex-tailed case (i.e. $\log \mathbb{P}\{|\xi_i| \ge t\}$ is convex or concave, respectively). A general optimal estimate for $\|\sum_i a_i\xi_i\|_{L_r}$, under the assumption that the variables are symmetric or non-negative, was obtained by Latała [26]. These results provide a mature understanding of the linear regime and serve as a benchmark for higher-order inequalities.

The next fundamental case is the quadratic form $\xi^\top A\xi - \mathbb{E}\xi^\top A\xi$, which is the subject of Hanson-Wright type inequalities. The classical Hanson-Wright inequality [18, 34] gives a two-level tail bound governed by the Hilbert–Schmidt and operator norms of A. It has become a standard tool in high-dimensional probability, with applications to random matrices, covariance estimation, compressed sensing, randomized linear algebra, graph theory, and statistical learning. Considerable work has been devoted to extensions beyond the classical Rademacher or sub-Gaussian settings, including quadratic forms of dependent variables, non-isotropic random vectors, unbounded entries, and tensor-valued analogues; see, for example, [27, 38, 1, 23, 30, 11, 22, 16, 8] and references therein.

Sparse quadratic forms with entries $X_i = \delta_i\xi_i$ where $\delta_i \sim \mathrm{Bernoulli}(p_i)$ have attracted particular attention recently [14, 39, 40, 19, 36, 35, 20, 12] (we refer to [20] for a detailed discussion), as dense Hanson-Wright inequalities lose the correct dependence on sparsity parameters p_i. The authors of [20] developed Hanson-Wright inequalities for sparse independent α-sub-exponential variables with applications to random matrices and norm concentration. Subsequently, [12] obtained a refined sparse Hanson-Wright inequality for $0 < \alpha \le 1$. The present paper extends this improved sparse quadratic theory to arbitrary fixed degree.

Beyond quadratic forms, a natural object is the order-d tetrahedral chaos

$$\sum_{i_1,\ldots,i_d} a_{i_1\ldots i_d}\xi_{i_1}\ldots\xi_{i_d},$$

where the d-tensor $A = (a_{i_1\ldots i_d})_{1\le i_1,\ldots,i_d\le n}$ is diagonal free (i.e., $a_{i_1\ldots i_d} = 0$ whenever two indices coincide) and $\xi = \{\xi_i, i \le n\}$ is a sequence of independent mean-zero random variables. Optimal L_r-norms for this chaos have been established by Latała [28] in the Gaussian case, by Adamczak and Latała [2] under log-concave tails (sharp for $d \le 3$), and by Kolesko and Latała [24] under log-convex tails.

Several related developments address tensor-valued or structured higher-order models. Vershynin [37] proved concentration inequalities for simple random tensors with independent sub-Gaussian entries, with explicit dependence on both dimension and tensor order. Bamberger, Krahmer and Ward [5] established a Hanson-Wright inequality for random tensors. Adamczak, Latała and Meller [3] obtained moment and tail estimates for Gaussian chaoses with values in Banach spaces. The general moment inequalities of Boucheron, Bousquet, Lugosi and Massart [7] apply to broad classes of functions of independent random variables and include moment inequalities for Rademacher chaoses.

These results primarily address either centered chaoses, dense tensor models, or general functions through abstract difference operators. While centering by subtracting the mean is conceptually straightforward, converting the resulting bounds into neat formulas is surprisingly complex, with computational difficulty growing steeply in the degree d. This challenge is already apparent for $d = 3$ and becomes prohibitive for higher orders.

A broader framework for polynomial concentration was developed by Adamczak-Wolff [4] and by Götze-Sambale-Sinulis [16]. These works study general polynomials $f(\xi_1, \dots, \xi_n)$ of independent, not necessarily centered, random variables, with bounds expressed in terms of partition norms of the expected derivative tensors $\mathbb{E}f^{(d)}(\xi)$ for $1 \le d \le D$. Adamczak and Wolff [4] treated the sub-Gaussian setting, while Götze, Sambale and Sinulis [16] obtained corresponding inequalities for α-sub-exponential variables with $0 < \alpha \le 1$. The appearance of all derivative orders is essential: lower-order derivatives may dominate fluctuations even for high-degree polynomials.

The present paper develops a sparse analogue of this polynomial concentration theory. We consider independent random variables $X_i = \delta_i \xi_i$ for $i = 1, \dots, n$, where the Bernoulli selectors δ_i have parameters p_i and the ξ_i are independent α-sub-exponential with $0 < \alpha \le 1$. For a polynomial $f : \mathbb{R}^n \to \mathbb{R}$ of degree at most D, we prove moment and tail bounds for $f(X) - \mathbb{E}f(X)$ in terms of sparsity-weighted versions of the expected derivative tensors. The case $\alpha \ge 1$ requires different analytical tools and will be addressed in future work.

We also apply our results to sparse simple random tensors, obtaining deviation inequalities for the distance between a sparse simple random tensor and a fixed subspace. Using leave-one-out arguments and the negative second moment identity, this yields lower bounds for the smallest singular value of matrices with independent sparse simple random tensor columns.

The rest of the paper is organized as follows. In Section 2, we state the main sparse polynomial concentration theorem and its tail consequence. Section 3 presents applications to sparse simple random tensors, including distance estimates to fixed subspaces and lower bounds for the smallest singular value of random tensor matrices. Section 4 is devoted to the proofs. The appendix contains auxiliary tensor-norm estimates needed for the application section.

2 Main results

We study concentration inequalities for polynomials of independent sparse α-sub-exponential random variables. Let $X_i = \delta_i \xi_i$ for $1 \le i \le n$, where the variables $\{\delta_i, \xi_i : i \in [n]\}$ are mutually independent, $\delta_i \sim \mathrm{Bernoulli}(p_i)$ with $0 < p_i \le 1$, and ξ_i is α-sub-exponential. Our goal is to bound

$$f(X) - \mathbb{E}f(X), \qquad X = (X_1, \dots, X_n),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a polynomial of degree D.

Recall that a random variable ξ is called *α-sub-exponential* if, for all $t \ge 0$,

$$\mathbb{P}\{|\xi| \ge t\} \le Ce^{-ct^\alpha},$$

where $C, c > 0$. We use the standard Orlicz functional

$$\|\xi\|_{\Psi_\alpha} := \inf\left\{t > 0 : \mathbb{E}\exp\left(\frac{|\xi|^\alpha}{t^\alpha}\right) \le 2\right\}.$$

For $\alpha \ge 1$, this is a norm; for $0 < \alpha < 1$, it is no longer a norm, but we keep the same notation.

Since $|X_i| \le |\xi_i|$, one has

$$\|X_i\|_{\Psi_\alpha} \le \|\xi_i\|_{\Psi_\alpha}.$$

Thus the dense polynomial concentration inequalities of Adamczak-Wolff [4] and Götze-Sambale-Sinulis [16] may be applied directly to $f(X) - \mathbb{E}f(X)$. However, this direct approach ignores the fact that X_i vanishes with probability $1-p_i$, and therefore generally gives suboptimal dependence on the sparsity parameters p_i (see Remark 2.4 in [12] for an example). The purpose of this paper is to derive concentration bounds that preserve this sparsity information.

We now introduce the notation for the main theorem, adapted primarily from [16]. For a vector $x = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$, we write $\|x\|_r = (\sum_i |x_i|^r)^{1/r}$ for the ℓ_r norm. For a random variable ξ, we write $\|\xi\|_{L_r} = (\mathbb{E}|\xi|^r)^{1/r}$ for its L_r norm.

Unless otherwise stated, we denote by $C, C_1, c, c_1, \ldots$ universal constants independent of any parameters and the dimension n. Similarly, $C(\delta), c(\delta), \ldots$ denote constants depending only on the parameter δ. Their values can change from line to line. We write $f \lesssim g$ if $f \le Cg$ for some universal constant C, and $f \lesssim_\delta g$ if $f \le C(\delta)g$ for some constant $C(\delta)$ depending only on δ. We use $f \asymp g$ if $f \lesssim g$ and $g \lesssim f$, and similarly $f \asymp_\delta g$.

Let $[n] := \{1, \ldots, n\}$. For $d \ge 1$, a multi-index $\mathbf{i} = (i_1, \ldots, i_d) \in [n]^d$ selects one coordinate in each of the d directions. A d-tensor is an array $A = (a_\mathbf{i})_{\mathbf{i}\in[n]^d}$. For two d-tensors $A = (a_\mathbf{i})$ and $B = (b_\mathbf{i})$, their Hadamard product is the d-tensor $A \circ B$ with entries

$$(A \circ B)_\mathbf{i} = a_\mathbf{i} b_\mathbf{i}, \qquad \mathbf{i} \in [n]^d.$$

Given an index set $I \subseteq [d]$, we write $i_I = (i_j)_{j\in I}$. For a d-tensor A, the notation A_{i_I} denotes the tensor obtained by fixing the coordinates in I^c and allowing the coordinates in I to vary. Equivalently, i_I records the free coordinates of the slice. For example, if $d = 4$, $I = \{2, 4\}$, and $i_1 = 1, i_3 = 2$, then $A_{i_I} = (a_{1\,i_2\,2\,i_4})_{i_2,i_4}$. In particular, in the main theorem below, the expression $(\cdot)_{\mathbf{i}_{I^c}}$ denotes the slice obtained after fixing the coordinates indexed by I.

Let Π_d be the set of all partitions of $[d]$. More generally, for $I \subseteq [d]$, let $\Pi(I)$ be the set of all partitions of I. If $\mathcal{J} = \{J_1, \ldots, J_k\} \in \Pi(I)$, then $|\mathcal{J}| = k$ denotes the number of blocks. We use the convention that $\Pi(\emptyset) = \{\emptyset\}$, and the corresponding tensor norm of a scalar is its absolute value.

Given a partition $\mathcal{J} = \{J_1, \ldots, J_k\} \in \Pi_d$, we define

$$\|A\|_\mathcal{J} := \sup\left\{\sum_{\mathbf{i}\in[n]^d} a_\mathbf{i} \prod_{\ell=1}^k x^{(\ell)}_{\mathbf{i}_{J_\ell}} : \sum_{\mathbf{i}_{J_\ell}} \left(x^{(\ell)}_{\mathbf{i}_{J_\ell}}\right)^2 \le 1,\ \ell = 1, \ldots, k\right\}.$$

The same definition applies to sliced tensors A_{i_I} and partitions $\mathcal{J} \in \Pi(I)$.

These norms increase as partitions become coarser. A partition $\mathcal{I} = \{I_1, \ldots, I_\mu\}$ is finer than a partition $\mathcal{J} = \{J_1, \ldots, J_\nu\}$ if each block of $\mathcal{J}$ is a union of blocks from $\mathcal{I}$. For example,

$\{\{1,2\},\{3\},\{4\}\}$ is finer than $\{\{1,2,4\},\{3\}\}$. This gives the monotonicity

$$\|A\|_{\mathrm{op}} = \|A\|_{\{\{1\},\ldots,\{d\}\}} \le \|A\|_{\mathcal{J}} \le \|A\|_{\{[d]\}} = \|A\|_{\mathrm{HS}} = \Big(\sum_{\mathbf{i}} a_{\mathbf{i}}^2\Big)^{1/2}.$$

For a multi-index $\mathbf{i} = (i_1, \ldots, i_d) \in [n]^d$, define

$$\mathfrak{p}_{\mathbf{i}}^{(d)} := \sqrt{p_{j_1} \cdots p_{j_\nu}},$$

where $j_1, \ldots, j_\nu$ are the distinct coordinate values appearing in $\mathbf{i}$. Thus repeated coordinates are counted only once. We write

$$\mathfrak{p}^{(d)} = \big(\mathfrak{p}_{\mathbf{i}}^{(d)}\big)_{\mathbf{i}\in[n]^d}$$

for the corresponding d-tensor. For $I \subseteq [d]$, define the partial sparsity weight

$$\mathfrak{p}_I^{(d)}(i_I) := \sqrt{p_{k_1} \cdots p_{k_\mu}},$$

where $k_1, \ldots, k_\mu$ are the distinct coordinate values appearing in $i_I = (i_j)_{j\in I}$. We use the convention $\mathfrak{p}_{\emptyset}^{(d)} = 1$. The role of the quotient $\mathfrak{p}^{(d)}/\mathfrak{p}_I^{(d)}$ will be illustrated after the theorem in the cubic case.

We first state the general result. For $1 \le d \le D$, $I \subseteq [d]$, and $\mathcal{J} \in \Pi([d] \setminus I)$, define

$$M_{d,I,\mathcal{J}} := \max_{i_I} \left(\left\| \Big(\mathbb{E} f^{(d)}(X) \circ \mathfrak{p}^{(d)}\Big)_{\mathbf{i}_{I^c}} \right\|_{\mathcal{J}} \frac{1}{\mathfrak{p}_I^{(d)}(i_I)} \right).$$

With the slicing convention above, the subscript $\mathbf{i}_{I^c}$ indicates that the coordinates in I^c remain free, while the coordinates in I are fixed and maximized over.

THEOREM 1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a polynomial of degree D. Let $X_1, \cdots, X_n$ be a sequence of independent random variables such that $X_i = \delta_i \cdot \xi_i$, where the variables $\{\delta_i, \xi_i : i \in [n]\}$ are mutually independent, $\delta_i \sim \mathrm{Bernoulli}(p_i)$, and ξ_i is an α-sub-exponential random variable satisfying $\|\xi_i\|_{\Psi_\alpha} \le 1$. Assume $0 < \alpha \le 1$. Then, for any $r \ge 1$,*

$$\|f(X) - \mathbb{E}f(X)\|_{L_r} \lesssim_{D,\alpha} \sum_{d=1}^{D} \sum_{I\subseteq[d]} \sum_{\mathcal{J}\in\Pi([d]\setminus I)} r^{|\mathcal{J}|/2+|I|/\alpha} M_{d,I,\mathcal{J}}.$$

The cubic case. We now illustrate Theorem 1 in the first genuinely higher-order case. Since $D = 2$ recovers the sparse Hanson-Wright bounds from [12], the cubic case $D = 3$ is the simplest setting where the new higher-order structure appears.

Assume, for this example, that $p_1 = \cdots = p_n = p$, that the variables X_i are centered, and that

$$f(x) = \frac{1}{3!} \sum_{i,j,k=1}^{n} B_{ijk} x_i x_j x_k,$$

where B is symmetric and diagonal-free, i.e., $B_{ijk} = 0$ whenever $|\{i,j,k\}| < 3$. Then $f^{(3)} \equiv B$, while $\mathbb{E}f^{(1)}(X) = \mathbb{E}f^{(2)}(X) = 0$. For any index vector $\mathbf{i}$ corresponding to a nonzero entry of B, we have $\frac{\mathfrak{p}_{\mathbf{i}}^{(3)}}{\mathfrak{p}_I^{(3)}(i_I)} = p^{(3-|I|)/2}$.

Theorem 1 yields that, for every $r \geq 1$,

$$\|f(X) - \mathbb{E}f(X)\|_{L_r} \lesssim_\alpha p^{3/2} \left[r^{1/2}\|B\|_{\{[3]\}} + r \sum_{\substack{\mathcal{J}\in\Pi([3]) \\ |\mathcal{J}|=2}} \|B\|_{\mathcal{J}} + r^{3/2}\|B\|_{\{\{1\},\{2\},\{3\}\}} \right]$$

$$+ p \sum_{\substack{I\subseteq[3] \\ |I|=1}} \left[r^{1/2+1/\alpha} \max_{i_I} \|B_{\mathbf{i}_{I^c}}\|_{\{I^c\}} + r^{1+1/\alpha} \max_{i_I} \|B_{\mathbf{i}_{I^c}}\|_{\{\{j\}:j\in I^c\}} \right]$$

$$+ \sqrt{p}\, r^{1/2+2/\alpha} \sum_{\substack{I\subseteq[3] \\ |I|=2}} \max_{i_I} \|B_{\mathbf{i}_{I^c}}\|_{\{I^c\}} + r^{3/\alpha} \max_{\mathbf{i}_{[3]}} |B_{i_1 i_2 i_3}|.$$

To interpret these partition norms, note that

$$\|B\|_{\{[3]\}} = \|B\|_{\mathrm{HS}}, \qquad \|B\|_{\{\{1\},\{2\},\{3\}\}} = \|B\|_{\mathrm{op}},$$

and the middle terms $\|B\|_{\mathcal{J}}$ for $|\mathcal{J}| = 2$ interpolate between them. When coordinates are fixed to form slices $\|B_{\mathbf{i}_{I^c}}\|$, different structures emerge. If $|I| = 1$, fixing one coordinate yields matrix slices with norms $\|B_{\mathbf{i}_{I^c}}\|_{\{I^c\}}$ (Hilbert-Schmidt norm) and $\|B_{\mathbf{i}_{I^c}}\|_{\{\{j\}:j\in I^c\}}$ (operator norm). If $|I| = 2$, fixing two coordinates yields vector fibers with ℓ_2 norm $\|B_{\mathbf{i}_{I^c}}\|_{\{I^c\}}$. If $|I| = 3$, fixing all coordinates yields the maximum entry $\|B\|_{\max} = \max_{i_1,i_2,i_3} |B_{i_1 i_2 i_3}|$.

This display illustrates the structure of the general theorem. If no coordinate is fixed, all three Bernoulli selectors contribute to the averaging scale, yielding the factor $p^{3/2}$. If one coordinate is fixed, the remaining matrix slice carries the factor p. If two coordinates are fixed, the remaining vector fiber carries the factor $\sqrt{p}$. If all three coordinates are fixed, there is no uniform sparsity gain in the high-moment regime, and the corresponding term is controlled by $\|B\|_{\max}$. Each fixed coordinate costs a heavy-tail factor $r^{1/\alpha}$, while the remaining free coordinates are measured by partition norms.

Theorem 1 and the standard moment-tail relationship (see, e.g., Proposition 3.3 in [16]) immediately yield the following exponential tail bound for $f(X)$. We omit the proof via Markov's inequality as it is standard.

COROLLARY 1. *Under the assumptions of Theorem 1, there exist constants $C_{D,\alpha}, c_{D,\alpha} > 0$ such that, for any $t > 0$,*

$$\mathbb{P}\{|f(X) - \mathbb{E}f(X)| \geq t\} \leq C_{D,\alpha} \exp\left[-c_{D,\alpha} \min_{\substack{1\leq d\leq D,\ I\subseteq[d], \\ \mathcal{J}\in\Pi([d]\setminus I),\ M_{d,I,\mathcal{J}}>0}} \left(\frac{t}{M_{d,I,\mathcal{J}}} \right)^{\frac{1}{|\mathcal{J}|/2+|I|/\alpha}} \right].$$

If all $M_{d,I,\mathcal{J}}$ vanish, then $f(X) - \mathbb{E}f(X) = 0$ almost surely and the bound is trivial.

2.1 Discussion of Theorem 1

When $p_1 = \cdots = p_n = 1$, all sparsity weights are equal to one. Hence Theorem 1 reduces to the polynomial concentration inequality of Götze, Sambale, and Sinulis [16] for α-sub-exponential inputs. In this sense, Theorem 1 is a sparse analogue of their result.

The case $D = 2$ recovers the sparse Hanson-Wright inequality proved in [12]. More precisely, for $\mathbb{E}\xi_i = 0$ and symmetric coefficients $a_{ij} = a_{ji}$, Theorem 1 yields a four-term estimate for $f(X) = \sum_{i,j} a_{ij} X_i X_j$:

$$\|f(X) - \mathbb{E}f(X)\|_{L_r} \lesssim_\alpha r^{2/\alpha} \max_{i,j} |a_{ij}| + r^{1/2+1/\alpha} \max_i \Big(\sum_{j=1}^n a_{ij}^2 p_j \Big)^{1/2}$$
$$+ r \left\| (a_{ij} \sqrt{p_i p_j})_{i,j} \right\|_{\mathrm{op}} + \sqrt{r} \Big(\sum_{i=1}^n a_{ii}^2 p_i + \sum_{i \neq j} a_{ij}^2 p_i p_j \Big)^{1/2}.$$

This illustrates the sparse Hanson-Wright phenomenon: averaged coordinates contribute powers of $p_i^{1/2}$, while coordinates controlled through high moments do not carry a uniform sparsity gain (see [20, 12] for detailed discussions).

For general D, the same mechanism is encoded by the quotient

$$\mathfrak{p}^{(d)} / \mathfrak{p}_I^{(d)}.$$

The numerator $\mathfrak{p}^{(d)}$ records the Bernoulli weights for all distinct coordinates in a derivative tensor. The denominator $\mathfrak{p}_I^{(d)}$ removes the sparsity gain from the coordinates fixed in the high-moment part. This removal is necessary: for instance, a single monomial like $X_1^D = \delta_1 \xi_1^D$ has L_r-norm of order $r^{D/\alpha}$ in the regime $r \gtrsim D \log(1/p_1)$, so no positive power of p_1 can multiply the corresponding extreme-coordinate term uniformly in r.

The appearance of all derivative orders $1 \le d \le D$ is also unavoidable for non-centered polynomials. For example, if $X_i = \delta_i \sim \text{Bernoulli}(p)$, $S = \sum_{i=1}^n X_i$, and $f(x) = (\sum_i x_i)^D$, then the leading fluctuation of S^D is of order $(np)^{D-1}\sqrt{np}$. This scale is captured by the first derivative $\mathbb{E}f^{(1)}(X)$, not by the top derivative alone. Thus the lower-order derivative tensors in Theorem 1 are a structural feature of polynomial concentration for non-centered inputs, already present in the dense results of Adamczak-Wolff [4] and Götze-Sambale-Sinulis [16].

Finally, we comment on the proof strategy. Our proof follows the polynomial concentration framework from the dense setting (see [1, 16]): one expands the polynomial into centered monomial chaoses, applies decoupling and symmetrization, reduces the variables to Weibull-type models through a contraction principle, and then invokes moment estimates for chaoses with log-convex tails. The new difficulty in the sparse setting is that the Bernoulli selectors must be kept throughout this reduction. Their contribution depends on how many distinct coordinates remain averaged and how many coordinates are fixed in the high-moment regime. This is exactly what produces the distinct-coordinate weight $\mathfrak{p}^{(d)}$ and its quotient by $\mathfrak{p}_I^{(d)}$. The key technical contribution is to propagate these weights through the partition-norm estimates and then convert the resulting coefficient tensors back into the expected derivative tensors $\mathbb{E}f^{(d)}(X)$.

3 Applications: Subspace Distance and Smallest Singular Values

In this section we illustrate Theorem 1 through a geometric concentration result for sparse simple random tensors. Let

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

where the vectors $X^{(1)}, \dots, X^{(d)} \in \mathbb{R}^n$ have independent sparse α-sub-exponential entries.

A natural problem is to understand the distance from X to a fixed subspace $H \subseteq \mathbb{R}^{n^d}$. Writing $P_{H^\perp}$ for the orthogonal projection onto $H^\perp$, we have

$$\operatorname{dist}(X, H) = \|P_{H^\perp} X\|_2.$$

For $d = 1$, this is a Lipschitz function of the underlying independent entries. For $d \geq 2$, however, the map

$$(X^{(1)}, \dots, X^{(d)}) \mapsto \|P_{H^\perp}(X^{(1)} \otimes \cdots \otimes X^{(d)})\|_2$$

is no longer globally Lipschitz in the scalar inputs. Its square is a polynomial of degree $2d$. More generally, for a deterministic matrix $B \in \mathbb{R}^{n^d \times n^d}$,

$$\|BX\|_2^2 = X^\top B^\top B X.$$

Thus Theorem 1 applies naturally to this problem and yields a deviation inequality which keeps track of the sparsity parameter.

The following theorem gives the resulting concentration estimate. Throughout this section, the tensor order d is fixed, and constants may depend on d and α.

THEOREM 2 (Distance theorem for sparse simple random tensors)**.** *Let $0 < \alpha \leq 1$, $p \in (0, 1]$, and let*

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

where $X_i^{(k)} = \delta_i^{(k)} \xi_i^{(k)}$, the variables $\delta_i^{(k)}$ are i.i.d. Bernoulli(p), and the variables $\xi_i^{(k)}$ are independent, mean zero, variance one, and satisfy $\|\xi_i^{(k)}\|_{\Psi_\alpha} \leq 1$. All variables are assumed to be mutually independent. Set

$$\Lambda_{n,p,d} := \max_{0 \leq s \leq d-1} (np)^{s/2}.$$

Then for every deterministic nonzero matrix $B \in \mathbb{R}^{n^d \times n^d}$ and every $t \geq 0$,

$$\begin{aligned}
&\mathbb{P}\left\{ \left| \|BX\|_2 - p^{d/2}\|B\|_{\mathrm{HS}} \right| \geq t \right\} \\
&\qquad \leq 2 \exp\left[-c_{\alpha,d} \min\left\{ \frac{t^2}{\Lambda_{n,p,d}^2 \|B\|_{\mathrm{op}}^2}, \left(\frac{t}{n^{(d-1)/2}\|B\|_{\mathrm{op}}} \right)^\alpha, \left(\frac{t}{\|B\|_{\mathrm{op}}} \right)^{\alpha/d} \right\} \right].
\end{aligned}$$

In particular, if $np \geq 1$, then

$$\Lambda_{n,p,d} = (np)^{(d-1)/2}.$$

Taking $B = P_{H^\perp}$ in Theorem 2 gives the following distance estimate. If $H \subseteq \mathbb{R}^{n^d}$ is a fixed subspace and $k := \dim(H^\perp)$, then $\|P_{H^\perp}\|_{\mathrm{op}} = 1$ and $\|P_{H^\perp}\|_{\mathrm{HS}} = \sqrt{k}$. Consequently, for every $t \geq 0$,

$$\begin{aligned}
&\mathbb{P}\left\{ \left| \operatorname{dist}(X, H) - p^{d/2}\sqrt{k} \right| \geq t \right\} \\
&\qquad \leq 2 \exp\left[-c_{\alpha,d} \min\left\{ \frac{t^2}{\Lambda_{n,p,d}^2}, \left(\frac{t}{n^{(d-1)/2}} \right)^\alpha, t^{\alpha/d} \right\} \right].
\end{aligned}$$

The distance estimate of the preceding form serves as a core tool for studying random matrices with tensor structures. Indeed, by the standard leave-one-out argument and the negative second moment identity, lower bounds on

$$\mathrm{dist}(\mathbf{X}_j, H_j), \qquad H_j := \mathrm{span}\{\mathbf{X}_i : i \neq j\},$$

imply lower bounds for the smallest singular value of the column matrix.

For $d = 1$, this problem reduces to the smallest singular value of an ordinary rectangular random matrix, which has been extensively investigated. For independent sub-Gaussian entries, Rudelson and Vershynin [33] proved lower bounds of the optimal order. In the sparse regime, Götze and Tikhomirov [17] demonstrated $\sigma_{\min}(X) \geq \sqrt{mp}$ under certain moment conditions. For higher-order tensors ($d \geq 2$), Vershynin [37] established that $\sigma_{\min}(X) \geq \sqrt{\epsilon}$ for dense sub-Gaussian tensor matrices when $m = (1-\epsilon)n^d$.

As a direct consequence of Theorem 2, we extend these frameworks simultaneously to the sparse and heavy-tailed setting, providing a non-trivial lower bound for sparse simple random tensor matrices.

COROLLARY 2 (Smallest singular value of sparse simple tensor matrices)**.** *Let $\mathbf{X}_1, \ldots, \mathbf{X}_m$ be independent copies of*

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

as in Theorem 2, and let

$$\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_m) \in \mathbb{R}^{n^d \times m}.$$

Suppose that

$$\frac{c_1 (d \log n)^{2/\alpha}}{np^d} \leq \varepsilon < 1$$

and

$$m \leq (1-\varepsilon) n^d.$$

Then

$$\mathbb{P}\left\{ \sigma_{\min}(\mathbf{X}) > \frac{1}{2} p^{d/2} \sqrt{\frac{\varepsilon}{1-\varepsilon}} \right\} \geq 1 - 2\exp\left[-c_2 (p^d \varepsilon n)^{\alpha/2} \right],$$

where $c_1, c_2 > 0$ depend only on α and d.

Remarks on the applications. Theorem 2 applies sparse polynomial concentration to quadratic forms in simple random tensors. If $X = X^{(1)} \otimes \cdots \otimes X^{(d)} \in \mathbb{R}^{n^d}$, then $\|BX\|_2^2 = X^\top B^\top BX$ is a degree-$2d$ polynomial in the underlying independent entries. The proof in Section 4.2.1 establishes a more general moment estimate (see Theorem 3) for such tensor quadratic forms, but we defer that technical statement to the proof section since it requires additional partial-trace notation. For geometric applications, Theorem 2 provides the main tool.

Theorem 2 gives a sparse, heavy-tailed analogue of tensor Hanson–Wright concentration. A closely dense result is the tensor Hanson-Wright inequality of Bamberger, Krahmer and Ward [5] for quadratic forms in simple random tensors with independent centered sub-Gaussian entries. Our setting differs: the entries are sparse and only α-sub-exponential with $0 < \alpha \leq 1$. Accordingly, the typical distance scale becomes $p^{d/2}\|B\|_{\mathrm{HS}}$, and the leading fluctuation contains the sparse factor $\Lambda_{n,p,d} = \max_{0\le s\le d-1}(np)^{s/2}$. Thus our result complements rather than improves the dense sub-Gaussian theory.

The tail bound in Theorem 2 has three regimes corresponding to

$$t^2/(\Lambda_{n,p,d}^2\|B\|_{\mathrm{op}}^2), \quad (t/(n^{(d-1)/2}\|B\|_{\mathrm{op}}))^{\alpha}, \quad (t/\|B\|_{\mathrm{op}})^{\alpha/d}.$$

The first captures Gaussian-type fluctuations, while the second and third reflect heavy-tailed behavior from products of α-sub-exponential variables. These additional regimes are natural for degree-d tensor products but absent in purely sub-Gaussian linear models.

The constants in our results depend on the tensor order d, so they are most useful when d is fixed. This differs from dense sub-Gaussian tensor results like Vershynin [37], where the dependence on d is tracked explicitly. The advantage of our approach is its applicability to sparse inputs and the heavy-tailed regime $0 < \alpha \le 1$.

Corollary 2 should be viewed as a consequence of general polynomial concentration rather than an optimal smallest-singular-value theorem. For ordinary matrices ($d = 1$), sharper bounds exist through specialized methods. The value of Corollary 2 is that the same distance argument works for tensor-product columns with sparse α-sub-exponential entries. The probability estimate becomes meaningful once $p^d\varepsilon n$ is at least logarithmic in n, which motivates the assumption $\varepsilon \ge c_1(d\log n)^{2/\alpha}/(np^d)$.

4 Proofs of our main results

4.1 Proof of Theorem 1

To begin, we introduce the notation and auxiliary lemmas that will be used throughout this section. For any multiindex $\mathbf{i}$, denote $|\mathbf{i}| = \sum_j i_j$. Write

$$[n]^{\underline{d}} := \left\{\mathbf{i} \in [n]^d : i_1, \dots, i_d \text{ are pairwise distinct}\right\}$$

and

$$I_{m,d} := \{(i_1, \cdots, i_m) \in \mathbb{N}_+^m : |\mathbf{i}| = d\}.$$

For two d-tensors $A = (a_{\mathbf{i}})$ and $B = (b_{\mathbf{i}})$, their (Frobenius) inner product is

$$\langle A, B\rangle := \sum_{\mathbf{i}\in[n]^d} a_{\mathbf{i}}\, b_{\mathbf{i}}.$$

The first property we shall introduce is the following contraction principle.

Lemma 4.1 (Lemma 3.1 in [16]). *For every $k \in \mathbb{N}$, $\alpha > 0$ and $r \ge 1$ the following holds. Let $Y_1, \dots, Y_n$ be independent, symmetric random variables satisfying $\|Y_i\|_{\Psi_{\alpha/k}} \le M$. Then*

$$\Big\|\sum_{i=1}^n a_i Y_i\Big\|_{L_r} \lesssim_{\alpha,k,M} \Big\|\sum_{i=1}^n a_i\, w_{i,1}\cdots w_{i,k}\Big\|_{L_r},$$

where the $w_{i,j}$ are i.i.d. symmetric Weibull random variables with shape parameter α.

Recall that a random variable ξ is said to have a log-convex tail if the function

$$t \mapsto \log \mathbb{P}\{|\xi| \geq t\}$$

is convex for $t \geq 0$. We will also need the following properties which bounds the L_r-norms of multilinear forms in random variables with log-convex tails.

LEMMA 4.2 (Theorem 1 in [24]). *Let $(X_i^{(j)})_{i\leq n, j\leq d}$ be independent symmetric r.v.'s with log-convex tails such that $\mathbb{E}|X_i^{(j)}|^2 = 1$ for all i, j. For any multiindexed matrix $(a_{\mathbf{i}})_{\mathbf{i}\in[n]^d}$ and any $r \geq 2$ we have*

$$\Big\| \sum_{\mathbf{i}} a_{\mathbf{i}} X_{i_1}^{(1)} \cdots X_{i_d}^{(d)} \Big\|_{L_r} \asymp_d \sum_{I\subseteq[d]} \sum_{\mathcal{J}\in\Pi(I^c)} r^{|\mathcal{J}|/2} \big(\sum_{\mathbf{i}_I} \|(a_{\mathbf{i}})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}^r \prod_{j\in I} \|X_{i_j}^{(j)}\|_{L_r}^r \big)^{1/r}.$$

To state the next property we fix some notation. For any subset $S \subseteq [n]^d$ let $\mathbf{1}_S$ be the indicator tensor whose entry at position $\mathbf{i} \in [n]^d$ is 1 if $\mathbf{i} \in S$ and 0 otherwise. Given a partition $\mathcal{K} = \{K_1, \ldots, K_\nu\}$ of $[d]$, we set

$$L(\mathcal{K}) = \big\{ \mathbf{i} \in [n]^d : i_k = i_l \iff \exists\, j \text{ such that } k, l \in K_j \big\}.$$

In words, $L(\mathcal{K})$ collects all multi-indices whose level-set partition is exactly $\mathcal{K}$.

LEMMA 4.3 (Lemma 3.4 in [16]). *Let $A = (a_{\mathbf{i}})_{\mathbf{i}\in[n]^d}$ be a d-tensor, $I \subseteq [d]$ and $\mathbf{i}_I \in [n]^{|I|}$ fixed. Then for any $\mathcal{K} \in \Pi_d$ and $\mathcal{J} \in \Pi(I^c)$,*

$$\|(A \circ \mathbf{1}_{L(\mathcal{K})})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \leq 2^{|\mathcal{K}|(|\mathcal{K}|-1)/2} \|A_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}.$$

Proof of Theorem 1. Without loss of generality, we can represent any order D polynomial f as

$$f(x) = \sum_{d=1}^{D} \sum_{\nu=1}^{d} \sum_{\mathbf{k}\in I_{\nu,d}} \sum_{\mathbf{i}\in[n]^{\underline{\nu}}} c_{(i_1,k_1),\cdots,(i_\nu,k_\nu)}^{(d)} x_{i_1}^{k_1} x_{i_2}^{k_2} \cdots x_{i_\nu}^{k_\nu} + c_0,$$

where the constant satisfies that $c_{(i_1,k_1),\cdots,(i_\nu,k_\nu)}^{(d)} = c_{(i_{\pi_1},k_{\pi_1}),\cdots,(i_{\pi_\nu},k_{\pi_\nu})}^{(d)}$ for any permutation $(\pi_1, \cdots, \pi_\nu)$ of $[\nu]$. Note that

$$\begin{aligned}
&f(X) - \mathbb{E}f(X) \\
=&\sum_{d=1}^{D} \sum_{\nu=1}^{d} \sum_{\mathbf{k}\in I_{\nu,d}} \sum_{\mathbf{i}\in[n]^{\underline{\nu}}} c_{(i_1,k_1),\cdots,(i_\nu,k_\nu)}^{(d)} \prod_{j\in[\nu]} (X_{i_j}^{k_j} - \mathbb{E}X_{i_j}^{k_j} + \mathbb{E}X_{i_j}^{k_j}) \\
=&\sum_{d=1}^{D} \sum_{\nu=1}^{d} \sum_{\mathbf{k}\in I_{\nu,d}} \sum_{\mathbf{i}\in[n]^{\underline{\nu}}} c_{(i_1,k_1),\cdots,(i_\nu,k_\nu)}^{(d)} \left(\sum_{\emptyset\neq J\subseteq[\nu]} \prod_{j\in J} (X_{i_j}^{k_j} - \mathbb{E}X_{i_j}^{k_j}) \prod_{j\notin J} \mathbb{E}X_{i_j}^{k_j} \right).
\end{aligned}$$

Hence, by rearranging the terms and using the triangle inequality, we have

$$\big|f(X) - \mathbb{E}f(X)\big| \leq \sum_{d=1}^{D} \sum_{\nu=1}^{d} \sum_{\mathbf{k}\in I_{\nu,d}} \Big| \sum_{\mathbf{i}\in[n]^{\underline{\nu}}} a_{\mathbf{i}}^{\mathbf{k}} (X_{i_1}^{k_1} - \mathbb{E}X_{i_1}^{k_1}) \cdots (X_{i_\nu}^{k_\nu} - \mathbb{E}X_{i_\nu}^{k_\nu}) \Big|,$$

where

$$a^{\mathbf{k}}_{\mathbf{i}} = a^{(k_1,\cdots,k_\nu)}_{i_1,\cdots,i_\nu} := \sum_{m=\nu}^{D} \sum_{\substack{k_{\nu+1},\cdots,k_m>0\\ k_1+\cdots+k_m\le D}} \sum_{\substack{i_{\nu+1},\cdots,i_m\\ (i_1,\cdots,i_m)\in[n]^{\underline{m}}}} \binom{m}{\nu} c^{(k_1+\cdots+k_m)}_{(i_1,k_1),\cdots,(i_m,k_m)} \prod_{\beta=\nu+1}^{m} \mathbb{E}X^{k_\beta}_{i_\beta}.$$

Indeed, in the rearrangement of the sum, for a given degree m with ν many terms of $(X_i^k - \mathbb{E}X_i^k)$, we will just consider the pattern $\prod_{j=1}^{\nu}(X^{k_j}_{i_j} - \mathbb{E}X^{k_j}_{i_j})\prod_{j=\nu+1}^{m}\mathbb{E}X^{k_j}_{i_j}$. Then for given m,ν, there will be $\binom{m}{\nu}$ many products corresponds to this pattern in the sum due to symmetry.

Let $X^{(1)},\cdots,X^{(D)}$ be independent copies of the random vector X. Let $(\varepsilon_i^{(j)}), 1\le i\le n, 1\le j\le D$ be a sequence of i.i.d. Rademacher variables independent of $X^{(1)},\cdots X^{(D)}$. By the standard decoupling inequalities (see Theorem 3.1.1 in [25]) and symmetrization technique (see Section 6.1 in [29]), we have for $r\ge 1$

$$\begin{aligned}
&\big\|f(X)-\mathbb{E}f(X)\big\|_{L_r}\\
\le&\sum_{d=1}^{D}\sum_{\nu=1}^{d}\sum_{\mathbf{k}\in I_{\nu,d}}\Big\|\sum_{\mathbf{i}\in[n]^{\underline{\nu}}} a^{\mathbf{k}}_{\mathbf{i}}((X^{(1)}_{i_1})^{k_1}-\mathbb{E}(X^{(1)}_{i_1})^{k_1})\cdots((X^{(\nu)}_{i_\nu})^{k_\nu}-\mathbb{E}(X^{(\nu)}_{i_\nu})^{k_\nu})\Big\|_{L_r}\\
\lesssim&\sum_{d=1}^{D}\sum_{\nu=1}^{d}\sum_{\mathbf{k}\in I_{\nu,d}}\Big\|\sum_{\mathbf{i}\in[n]^{\underline{\nu}}} a^{\mathbf{k}}_{\mathbf{i}}\varepsilon^{(1)}_{i_1}(X^{(1)}_{i_1})^{k_1}\cdots\varepsilon^{(\nu)}_{i_\nu}(X^{(\nu)}_{i_\nu})^{k_\nu}\Big\|_{L_r}\\
\lesssim&\sum_{d=1}^{D}\sum_{\nu=1}^{d}\sum_{\mathbf{k}\in I_{\nu,d}}\Big\|\sum_{\mathbf{i}\in[n]^{\underline{\nu}}} a^{\mathbf{k}}_{\mathbf{i}}\delta^{(1)}_{i_1}\varepsilon^{(1)}_{i_1}(\xi^{(1)}_{i_1})^{k_1}\cdots\delta^{(\nu)}_{i_\nu}\varepsilon^{(\nu)}_{i_\nu}(\xi^{(\nu)}_{i_\nu})^{k_\nu}\Big\|_{L_r}.
\end{aligned}$$

Due to that $\|(\xi_i^{(j)})^k\|_{\Psi_{\alpha/k}} = \|\xi_i^{(j)}\|^k_{\Psi_\alpha}\le 1, j=1,\cdots,D$, we have by an iteration of Lemma 4.1

$$\begin{aligned}
&\big\|f(X)-\mathbb{E}f(X)\big\|_{L_r}\\
\lesssim&\sum_{d=1}^{D}\sum_{\nu=1}^{d}\sum_{\mathbf{k}\in I_{\nu,d}}\Big\|\sum_{\mathbf{i}\in[n]^{\underline{\nu}}} a^{\mathbf{k}}_{\mathbf{i}}(\delta^{(1)}_{i_1}w^{(1)}_{i_1,1}\cdots w^{(1)}_{i_1,k_1})\cdots(\delta^{(\nu)}_{i_\nu}w^{(\nu)}_{i_\nu,1},\cdots w^{(\nu)}_{i_\nu,k_\nu})\Big\|_{L_r},
\end{aligned}\tag{4.1}$$

where $\{w^{(j)}_{i,k}, 1\le i,j,k\le D\}$ is a sequence of i.i.d. symmetric Weibull variables with shape parameter α, and the extra factor of ε will not change its distribution.

We next define a d-tensor $A_d = (a^{(d)}_{\mathbf{i}})_{\mathbf{i}\in[n]^d}$ as follows. Given any $\mathbf{i}=(i_1,\cdots,i_d)$, there are distinct elements $j_1,\cdots,j_\nu$ with each j_l appearing exactly k_l times in $\mathbf{i}$. Then, we set $a^{(d)}_{i_1,\cdots,i_d} := a^{(k_1,\cdots,k_\nu)}_{j_1,\cdots,j_\nu}$ (as $a^{\mathbf{k}}_{\mathbf{j}}$ is invariant under permutation on $[\nu]$). Obviously, we have $k_1+\cdots+k_\nu = d$. For any $\mathbf{k}=(k_1,\cdots,k_\nu)\in I_{\nu,d}$, denote by $\mathcal{K}(\mathbf{k}) = \{K_1,\cdots,K_\nu\}\in\Pi_d$ such that $K_l = \{\sum_{i=1}^{l-1}k_i+1, \sum_{i=1}^{l-1}k_i+2,\cdots,\sum_{i=1}^{l}k_i\}, l=1,\cdots,\nu$. For convenience, given a $\mathbf{k}\in I_{\nu,d}$, let

$$\begin{aligned}
&Y^{(1)} = \big(w^{(1)}_{i,1}\cdot\delta^{(1)}_i\big)_{i\le n}, Y^{(2)} = \big(w^{(1)}_{i,2}\big)_{i\le n},\cdots, Y^{(k_1)} = \big(w^{(1)}_{i,k_1}\big)_{i\le n},\\
&Y^{(k_1+1)} = \big(w^{(2)}_{i,1}\cdot\delta^{(2)}_i\big)_{i\le n}, Y^{(k_1+2)} = \big(w^{(2)}_{i,2}\big)_{i\le n},\cdots, Y^{(k_1+k_2)} = \big(w^{(2)}_{i,k_2}\big)_{i\le n},\\
&\cdots\\
&Y^{(\sum_{j=1}^{\nu-1}k_j+1)} = \big(w^{(\nu)}_{i,1}\cdot\delta^{(\nu)}_i\big)_{i\le n}, Y^{(\sum_{j=1}^{\nu-1}k_j+2)} = \big(w^{(\nu)}_{i,2}\big)_{i\le n}\cdots, Y^{(d)} = \big(w^{(\nu)}_{i,k_\nu}\big)_{i\le n}.
\end{aligned}$$

Note that

$$-\log \mathbb{P}\{|\delta_{i_1}^{(1)} w_{i_1,1}^{(1)}| \geq t\} = \begin{cases} 0, & t = 0 \\ t^\alpha - \log p_i, & t > 0 \end{cases}$$

is a concave function for $t \geq 0$. Since Lemma 4.2 is stated under the normalization $\mathbb{E}|X_i^{(j)}|^2 = 1$, we apply it below to the normalized variables $Y_i^{(j)}/\|Y_i^{(j)}\|_2$; the deterministic factors $\|Y_i^{(j)}\|_2$ are then absorbed into the coefficient tensor. Hence, we can further bound (4.1) by Lemma 4.2

$$\begin{aligned}
&\sum_{d=1}^{D}\sum_{\nu=1}^{d}\sum_{\mathbf{k}\in I_{\nu,d}} \Big\|\Big\langle A_d \circ \mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}, \otimes_{j=1}^d Y^{(j)}\Big\rangle\Big\|_{L_r} \\
\lesssim_d &\sum_{d=1}^{D}\sum_{\nu=1}^{d}\sum_{\mathbf{k}\in I_{\nu,d}} \sum_{I\subseteq[d]} \sum_{\mathcal{J}\in\Pi([d]\setminus I)} r^{|\mathcal{J}|/2}\Big(\sum_{\mathbf{i}_I}\Big(\big\|\big(A_d\circ \mathfrak{p}^{(d)}\circ \mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}}^r \\
&\times \prod_{j\in I}\|Y_{i_j}^{(j)}\|_{L_r}^r \prod_{j\in I\cap S(\mathbf{k})} \frac{1}{\sqrt{p_{i_j}^r}}\Big)\Big)^{1/r},
\end{aligned}$$

where $\mathfrak{p}^{(d)}$ is defined as in Theorem 1, and $S(\mathbf{k})$ is defined as follows: for a vector $\mathbf{k} = (k_1, \ldots, k_\nu)$ we set $S(\mathbf{k}) = \{1, k_1+1, k_1+k_2+1, \ldots, k_1+\cdots+k_{\nu-1}+1\}$.

Let $\tilde{Y}^{(1)}, \cdots, \tilde{Y}^{(d)} \overset{\text{i.i.d.}}{\sim} Y^{(2)}$. Then, by Young's inequality, we have for $r \geq 3$

$$\begin{aligned}
&\Big(\sum_{\mathbf{i}_I}\big\|\big(A_d\circ \mathfrak{p}^{(d)}\circ \mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}}^r \prod_{j\in I}\|Y_{i_j}^{(j)}\|_{L_r}^r \prod_{j\in I\cap S(\mathbf{k})} \frac{1}{\sqrt{p_{i_j}^r}}\Big)^{1/r} \\
=&\Big(\sum_{\mathbf{i}_I}\big\|\big(A_d\circ \mathfrak{p}^{(d)}\circ \mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}}^r \prod_{j\in I}\|\tilde{Y}_{i_j}^{(j)}\|_{L_r}^r \prod_{j\in I\cap S(\mathbf{k})} \frac{p_{i_j}}{\sqrt{p_{i_j}^r}}\Big)^{1/r} \\
\lesssim_d&\Big(e^{2r/(r-2)}\max_{\mathbf{i}_I}\big\|\big(A_d\circ \mathfrak{p}^{(d)}\circ \mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}} \prod_{j\in I}\|\tilde{Y}_{i_j}^{(j)}\|_{L_r} \prod_{j\in I\cap S(\mathbf{k})} \frac{1}{\sqrt{p_{i_j}}}\Big)^{(r-2)/r} \\
&\times\Big(e^{-2r}\sum_{\mathbf{i}_I}\big\|\big(A_d\circ \mathfrak{p}^{(d)}\circ \mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}}^2 \prod_{j\in I}\|\tilde{Y}_{i_j}^{(j)}\|_{L_r}^2\Big)^{1/r} \\
\leq&\frac{r-2}{r}\Big(e^{2r/(r-2)}\max_{\mathbf{i}_I}\big\|\big(A_d\circ \mathfrak{p}^{(d)}\circ \mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}} \prod_{j\in I}\|\tilde{Y}_{i_j}^{(j)}\|_{L_r} \prod_{j\in I\cap S(\mathbf{k})} \frac{1}{\sqrt{p_{i_j}}}\Big) \\
&+\frac{2}{r}\Big(e^{-2r}\sum_{\mathbf{i}_I}\big\|\big(A_d\circ \mathfrak{p}^{(d)}\circ \mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}}^2 \prod_{j\in I}\|\tilde{Y}_{i_j}^{(j)}\|_{L_r}^2\Big)^{1/2} \\
\lesssim_{\alpha,d}&\max_{\mathbf{i}_I}\Big(\big\|\big(A_d\circ \mathfrak{p}^{(d)}\circ \mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}} \prod_{j\in I}\|\tilde{Y}_{i_j}^{(j)}\|_{L_r} \prod_{j\in I\cap S(\mathbf{k})} \frac{1}{\sqrt{p_{i_j}}}\Big) \\
&+\|A_d\circ \mathfrak{p}^{(d)}\circ \mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\|_{\mathrm{HS}},
\end{aligned}$$

where the last inequality we use the fact that

$$\prod_{j\in I}\|\tilde{Y}_{i_j}^{(j)}\|_{L_r}^2 \lesssim_{\alpha,d} r^{2d/\alpha}, \quad \sum_{\mathbf{i}_I}\big\|\big(A_d\circ \mathfrak{p}^{(d)}\circ \mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}}^2 \leq \|A_d\circ \mathfrak{p}^{(d)}\circ \mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\|_{\mathrm{HS}}^2.$$

Note that, when $I=\emptyset$ and $\mathcal{J}=\{[d]\}$, we have

$$\big\|\big(A_d\circ\mathfrak{p}^{(d)}\circ\mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}}\prod_{j\in I}\|\tilde{Y}_{i_j}^{(j)}\|_{L_r}\prod_{j\in I\cap S(\mathbf{k})}\frac{1}{\sqrt{p_{i_j}}}=\|A_d\circ\mathfrak{p}^{(d)}\circ\mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\|_{\mathrm{HS}}.$$

Hence, (4.1) can be further bounded by

$$\begin{aligned}
&\sum_{d=1}^{D}\sum_{I\subseteq[d]}\sum_{\mathcal{J}\in\Pi([d]\setminus I)}r^{|\mathcal{J}|/2+|I|/\alpha}\sum_{\nu=1}^{d}\sum_{\mathbf{k}\in I_{\nu,d}}\max_{\mathbf{i}_I}\Big(\big\|\big(A_d\circ\mathfrak{p}^{(d)}\circ\mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}}\prod_{j\in I\cap S(\mathbf{k})}\frac{1}{\sqrt{p_{i_j}}}\Big)\\
\leq&\sum_{d=1}^{D}\sum_{I\subseteq[d]}\sum_{\mathcal{J}\in\Pi([d]\setminus I)}r^{|\mathcal{J}|/2+|I|/\alpha}\sum_{\nu=1}^{d}\sum_{\mathbf{k}\in I_{\nu,d}}\max_{\mathbf{i}_I}\Big(\big\|\big(A_d\circ\mathfrak{p}^{(d)}\circ\mathbf{1}_{L(\mathcal{K}(\mathbf{k}))}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}}\frac{1}{\mathfrak{p}_I^{(d)}(\mathbf{i}_I)}\Big)\\
\lesssim_D&\sum_{d=1}^{D}\sum_{I\subseteq[d]}\sum_{\mathcal{J}\in\Pi([d]\setminus I)}r^{|\mathcal{J}|/2+|I|/\alpha}\max_{\mathbf{i}_I}\Big(\big\|\big(A_d\circ\mathfrak{p}^{(d)}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}}\frac{1}{\mathfrak{p}_I^{(d)}(\mathbf{i}_I)}\Big),
\end{aligned}$$

where $\mathfrak{p}_I^{(d)}$ was defined in Theorem 1 and the second inequality is due to Lemma 4.3.

To finish the proof, it remains to replace $\big\|\big(A_d\circ\mathfrak{p}^{(d)}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}}$ by $\big\|\big(\mathbb{E}f^{(d)}(X)\circ\mathfrak{p}^{(d)}\big)_{\mathbf{i}_{I^c}}\big\|_{\mathcal{J}}$. To this end, first fix any multi-index $\mathbf{i}=(i_1,\cdots,i_d)\in[n]^d$ whose distinct values are $j_1,\ldots,j_\nu$ occurring $l_1,\ldots,l_\nu$ times, respectively. Then, we have

$$\begin{aligned}
\mathbb{E}\frac{\partial^d f}{\partial x_{i_1}\cdots\partial x_{i_d}}(X)=&\sum_{k_1\geq l_1,\cdots,k_\nu\geq l_\nu}\sum_{m=\nu}^{D}\sum_{\substack{k_{\nu+1},\cdots,k_m>0\\ k_1+\cdots+k_m\leq D}}\sum_{\substack{j_{\nu+1},\cdots,j_m\\ (j_1,\cdots,j_m)\in[n]^{\underline{m}}}}\\
&\Bigg(\binom{m}{\nu}\nu!c^{(k_1+\cdots+k_m)}_{(j_1,k_1),\cdots,(j_m,k_m)}\Big(\prod_{\beta=1}^{\nu}\mathbb{E}X_{j_\beta}^{k_\beta-l_\beta}\frac{k_\beta!}{(k_\beta-l_\beta)!}\Big)\Big(\prod_{\beta=\nu+1}^{m}\mathbb{E}X_{j_\beta}^{k_\beta}\Big)\Bigg)\\
=&\nu!l_1!\cdots l_\nu!a_{\mathbf{i}}^{(d)}+R_{\mathbf{i}}^{(d)},
\end{aligned}\tag{4.2}$$

where $a_{\mathbf{i}}^{(d)}$ is the $\mathbf{i}$-th entry of A_d defined above and $R_{\mathbf{i}}^{(d)}$ corresponds to the set of indices $\mathbf{k}$ such that, $\exists i\in[\nu],k_i>l_i$.

For the case $d=D$, we have $l_1+\cdots+l_\nu=D$. Hence, we have $k_1=l_1,\cdots,k_\nu=l_\nu$ which yields that

$$\mathbb{E}\frac{\partial^D f}{\partial x_{i_1}\cdots\partial x_{i_d}}(X)=\nu!l_1!\cdots l_\nu!a_{\mathbf{i}}^{(D)}.\tag{4.3}$$

Note that $A_D=\sum_{\mathcal{K}\in\Pi_D}A_D\circ\mathbf{1}_{L(\mathcal{K})}$. Hence, for any $I\subseteq[d]$ and $\mathcal{J}\in\Pi(I^c)$, we have by the triangle inequality

$$\begin{aligned}
&\|(A_D\circ\mathfrak{p}^{(D)})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}\\
\leq&\sum_{\mathcal{K}\in\Pi_D}\|(A_D\circ\mathfrak{p}^{(D)}\circ\mathbf{1}_{L(\mathcal{K})})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}\leq\sum_{\mathcal{K}\in\Pi_D}\|(\mathbb{E}f^{(D)}(X)\circ\mathfrak{p}^{(D)}\circ\mathbf{1}_{L(\mathcal{K})})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}\\
\leq&\sqrt{2}^{D(D-1)}\cdot|\Pi_D|\cdot\|(\mathbb{E}f^{(D)}(X)\circ\mathfrak{p}^{(D)})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}\lesssim_D\|(\mathbb{E}f^{(D)}(X)\circ\mathfrak{p}^{(D)})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}},
\end{aligned}\tag{4.4}$$

where the third inequality is due to Lemma 4.3 and the second inequality is due to that, for any fixed $\mathcal{K} = \{K_1, \cdots, K_\nu\} \in \Pi_D$, by virtue of (4.3), we have

$$\mathbb{E} f^{(D)}(X) \circ \mathbf{1}_{L(\mathcal{K})} = \nu! |K_1|! \cdots |K_\nu|! \cdot A_D \circ \mathbf{1}_{L(\mathcal{K})}.$$

To ensure a streamlined presentation, we temporarily assume the validity of the following key estimate, deferring its detailed derivation to Appendix A. For every $d = 1, \ldots, D-1$, every subset $I \subseteq [d]$, and every partition $\mathcal{J} \in \Pi([d] \setminus I)$, we have

$$\|(R^{(d)} \circ \mathfrak{p}^{(d)})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \lesssim_D \sum_{k=d+1}^{D} \sum_{\substack{\mathcal{K} \in \Pi([k] \setminus I) \\ |\mathcal{K}| \geq |\mathcal{J}|}} \|(A_k \circ \mathfrak{p}^{(k)})_{\mathbf{i}_{I^c}}\|_{\mathcal{K}}. \tag{4.5}$$

In the specific case where $d = D-1$, invoking (4.2), (4.4), and (4.5) yields

$$\begin{aligned}
&\|(A_{D-1} \circ \mathfrak{p}^{(D-1)})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \\
\lesssim_D &\|(\mathbb{E} f^{(D-1)}(X) \circ \mathfrak{p}^{(D-1)})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} + \|(R^{(D-1)} \circ \mathfrak{p}^{(D-1)})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \\
\lesssim_D &\|(\mathbb{E} f^{(D-1)}(X) \circ \mathfrak{p}^{(D-1)})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} + \sum_{\substack{\mathcal{K} \in \Pi([D] \setminus I) \\ |\mathcal{K}| \geq |\mathcal{J}|}} \|(\mathbb{E} f^{(D)}(X) \circ \mathfrak{p}^{(D)})_{\mathbf{i}_{I^c}}\|_{\mathcal{K}}.
\end{aligned}$$

Consequently, a backward induction argument leads to

$$\begin{aligned}
&\sum_{d=1}^{D} \sum_{I \subseteq [d]} \sum_{\mathcal{J} \in \Pi([d] \setminus I)} r^{|\mathcal{J}|/2 + |I|/\alpha} \max_{\mathbf{i}_I} \left(\left\| \left(A_d \circ \mathfrak{p}^{(d)}\right)_{\mathbf{i}_{I^c}} \right\|_{\mathcal{J}} \frac{1}{\mathfrak{p}_I^{(d)}(\mathbf{i}_I)} \right) \\
\lesssim_D &\sum_{d=1}^{D} \sum_{I \subseteq [d]} \sum_{\mathcal{J} \in \Pi([d] \setminus I)} r^{|\mathcal{J}|/2 + |I|/\alpha} \max_{\mathbf{i}_I} \left(\left\| \left(\mathbb{E} f^{(d)}(X) \circ \mathfrak{p}^{(d)}\right)_{\mathbf{i}_{I^c}} \right\|_{\mathcal{J}} \frac{1}{\mathfrak{p}_I^{(d)}(\mathbf{i}_I)} \right),
\end{aligned}$$

which completes the proof for $r \geq 3$.

For $1 \leq r < 3$, by monotonicity of L_r-norms,

$$\|f(X) - \mathbb{E} f(X)\|_{L_r} \leq \|f(X) - \mathbb{E} f(X)\|_{L_3}.$$

Since all powers $r^{|\mathcal{J}|/2 + |I|/\alpha}$ are bounded below by constants on $1 \leq r < 3$ and above by constants depending only on D, α, the $r = 3$ bound implies the desired bound after changing the constants.

□

4.2 Proofs of Theorem 2 and Corollary 2

4.2.1 Concentration inequalities for random tensors

Our proof of the distance theorem (Theorem 2) uses moment estimates for a particular high-order quadratic-type chaos. We first state these estimates in a form that is convenient for later use.

Let $X^{(1)}, \ldots, X^{(d)}$ be independent random vectors in $\mathbb{R}^n$. For each $k \in [d]$ and $i \in [n]$, assume that

$$X_i^{(k)} = \delta_i^{(k)} \xi_i^{(k)},$$

where $\delta_i^{(k)} \sim \mathrm{Bernoulli}(p_i)$, $0 < p_i \leq 1$, and $\xi_i^{(k)}$ is mean zero, has variance one, and is α-sub-exponential with

$$\|\xi_i^{(k)}\|_{\Psi_\alpha} \leq 1$$

for some fixed $0 < \alpha \leq 1$. All random variables appearing above are assumed to be independent. We write

$$X = (X^{(1)}, \ldots, X^{(d)}) \in \mathbb{R}^{nd}$$

for the concatenated random vector.

Let $A = (A_{i_1,\ldots,i_{2d}})_{(i_1,\ldots,i_{2d})\in[n]^{2d}}$ be a deterministic $2d$-tensor, and consider the polynomial

$$f(X) = \sum_{i_1,\ldots,i_{2d}=1}^{n} A_{i_1,\ldots,i_d,i_{d+1},\ldots,i_{2d}} \prod_{k=1}^{d} X_{i_k}^{(k)} X_{i_{d+k}}^{(k)}. \tag{4.6}$$

This polynomial is homogeneous of degree two in each block $X^{(k)}$, and has total degree $2d$. In the proof of Theorem 2, A will be chosen so that $f(X)$ represents the squared norm of a projected random tensor. Thus the moment estimates for (4.6) provide the input needed for the distance theorem.

To apply the general polynomial concentration result, Theorem 1, the main technical point is to estimate the expected derivative tensors

$$\mathbb{E}f^{(m)}(X), \qquad 1 \leq m \leq 2d,$$

in the relevant tensor partition norms. The block-homogeneous structure of (4.6) allows these norms to be written in more concrete terms, using matrix operator norms and certain partial traces of the tensor A. We therefore introduce the notation and weights that will be used in the statement and proof of the main result of this subsection.

We identify the coordinate $X_i^{(k)}$ with the block-coordinate pair (k, i). Equivalently, one may identify (k, i) with $(k-1)n + i \in [nd]$. For a multi-index

$$\gamma = ((k_1, i_1), \ldots, (k_m, i_m)) \in ([d] \times [n])^m,$$

define the sparse weight

$$\mathfrak{P}_\gamma^{(m)} = \left(\prod_{(k,i)\in \mathrm{supp}(\gamma)} p_i \right)^{1/2},$$

where $\mathrm{supp}(\gamma)$ denotes the set of distinct block-coordinate pairs appearing in γ. For $S \subseteq [m]$, the marginal weight is

$$\mathfrak{P}_S^{(m)}(\gamma_S) = \left(\prod_{(k,i)\in \mathrm{supp}(\gamma_S)} p_i \right)^{1/2},$$

where $\gamma_S = ((k_j, i_j), j \in S)$. This is the same weight as in Theorem 1 after the identification $(k, i) \leftrightarrow (k-1)n + i$.

Reduced weights for the free variables. For $l \in [d]$ and $j = (j_1, \dots, j_{2l}) \in [n]^{2l}$, define the reduced block-coordinate map

$$\vartheta_l(s, j_s) := \begin{cases} (s, j_s), & 1 \le s \le l, \\ (s-l, j_s), & l+1 \le s \le 2l. \end{cases}$$

For $I_2 \subseteq [2l]$, define

$$\widetilde{\mathfrak{P}}^{(2l)}_{I_2}(j_{I_2}) := \left(\prod_{(m,a) \in \{\vartheta_l(s,j_s) \colon s \in I_2\}} p_a \right)^{1/2}.$$

The product is taken over distinct pairs (m, a). In particular, if two equal coordinates occur in the same reduced block, they contribute only one factor $\sqrt{p}$; if the same coordinate value occurs in two different reduced blocks, it contributes twice, since the underlying random variables belong to different independent copies.

We write $\widetilde{\mathfrak{P}}^{(2l)}(j) := \widetilde{\mathfrak{P}}^{(2l)}_{[2l]}(j)$. Equivalently, for $j = (a, b) = (a_1, \dots, a_l, b_1, \dots, b_l)$,

$$\widetilde{\mathfrak{P}}^{(2l)}(a, b) = \prod_{m=1}^{l} q(a_m, b_m), \qquad q(a_m, b_m) := \begin{cases} \sqrt{p_{a_m} p_{b_m}}, & a_m \neq b_m, \\ \sqrt{p_{a_m}}, & a_m = b_m. \end{cases}$$

Weighted partial traces. Fix $I = \{k_1, \dots, k_l\} \subseteq [d]$, where $k_1 < \cdots < k_l$. For $a = (a_1, \dots, a_l) \in [n]^l$, $b = (b_1, \dots, b_l) \in [n]^l$, $u = (u_t)_{t \in [d] \setminus I} \in [n]^{d-l}$, and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_l) \in \{0,1\}^l$, define

$$\Theta_{I,\varepsilon}(a, b, u) = (i_1, \dots, i_{2d}) \in [n]^{2d}$$

as follows. For each unselected block $t \in [d] \setminus I$, set

$$i_t = i_{d+t} = u_t.$$

For each selected block $t = k_m \in I$, set

$$(i_{k_m}, i_{d+k_m}) = \begin{cases} (a_m, b_m), & \varepsilon_m = 0, \\ (b_m, a_m), & \varepsilon_m = 1. \end{cases}$$

Define the p-weighted partial trace

$$T_I^p A(a, b) := \sum_{u \in [n]^{d-l}} \left(\prod_{t \in [d] \setminus I} p_{u_t} \right) \sum_{\varepsilon \in \{0,1\}^l} A_{\Theta_{I,\varepsilon}(a,b,u)}.$$

Finally, define the reduced tensor

$$\mathcal{A}_I^{(2l,p)}(a, b) := \widetilde{\mathfrak{P}}^{(2l)}(a, b)\, T_I^p A(a, b).$$

The factor $\prod_{t \notin I} p_{u_t}$ comes from the contraction

$$\mathbb{E} X_{i_t}^{(t)} X_{i_{d+t}}^{(t)} = p_{i_t} \mathbf{1}_{\{i_t = i_{d+t}\}},$$

whereas $\widetilde{\mathfrak{P}}^{(2l)}(a, b)$ is the derivative weight from Theorem 1 for the free variables.

THEOREM 3. *Let f be the polynomial defined in* (4.6)*. Then, for every $r \geq 2$,*

$$\|f(X) - \mathbb{E}f(X)\|_{L_r} \lesssim_{\alpha,d} \sum_{l=1}^{d} \sum_{I_2 \subseteq [2l]} \sum_{\mathcal{J} \in \Pi([2l] \setminus I_2)} r^{|\mathcal{J}|/2 + |I_2|/\alpha} \max_{j_{I_2}} \sum_{\substack{I_1 \subseteq [d] \\ |I_1| = l}} \frac{\left\| \left(\mathcal{A}_{I_1}^{(2l,p)} \right)_{j_{I_2^c}} \right\|_{\mathcal{J}}}{\widetilde{\mathfrak{P}}_{I_2}^{(2l)}(j_{I_2})}. \tag{4.7}$$

We use the conventions $\widetilde{\mathfrak{P}}_{\varnothing}^{(2l)} = 1$ and $\Pi(\varnothing) = \{\varnothing\}$.

Proof. Applying Theorem 1 to this polynomial, whose degree is at most $2d$, gives

$$\|f(X) - \mathbb{E}f(X)\|_{L_r} \lesssim_{\alpha,d} \sum_{m=1}^{2d} \sum_{S \subseteq [m]} \sum_{\mathcal{J} \in \Pi([m] \setminus S)} r^{|\mathcal{J}|/2 + |S|/\alpha} \max_{\gamma_S} \frac{\left\| \left(\mathbb{E}f^{(m)}(X) \circ \mathfrak{P}^{(m)} \right)_{\gamma_{S^c}} \right\|_{\mathcal{J}}}{\mathfrak{P}_S^{(m)}(\gamma_S)}. \tag{4.8}$$

Here $\gamma = ((k_1, i_1), \ldots, (k_m, i_m))$ is a derivative multi-index in $([d] \times [n])^m$.

We next identify the nonzero entries of $\mathbb{E}f^{(m)}(X)$. For a derivative multi-index γ, let

$$N_t(\gamma) := \#\{s \in [m] : k_s = t\}, \qquad t \in [d].$$

Since the polynomial (4.6) has degree at most two in each block $X^{(t)}$, we have

$$\mathbb{E}\partial_\gamma f(X) = 0$$

unless

$$N_t(\gamma) \in \{0, 2\} \qquad \text{for every } t \in [d].$$

Indeed, if $N_t(\gamma) \geq 3$, the derivative with respect to the t-th block vanishes. If $N_t(\gamma) = 1$, then after differentiation one centered factor from the t-th block remains, and its expectation is zero.

Consequently only even orders $m = 2l$ contribute. Let

$$I = \{k_1, \ldots, k_l\} \subseteq [d], \qquad k_1 < \cdots < k_l,$$

be the set of active blocks, namely the blocks differentiated exactly twice. For $a = (a_1, \ldots, a_l)$ and $b = (b_1, \ldots, b_l)$, consider the canonical derivative

$$\partial_{I,a,b} := \partial_{z_{k_1,a_1}} \cdots \partial_{z_{k_l,a_l}} \partial_{z_{k_1,b_1}} \cdots \partial_{z_{k_l,b_l}}.$$

For each selected block k_m,

$$\begin{aligned} &\partial_{z_{k_m,a_m}} \partial_{z_{k_m,b_m}} \left(z_{k_m,i_{k_m}} z_{k_m,i_{d+k_m}} \right) \\ &\qquad = \mathbf{1}_{\{(i_{k_m}, i_{d+k_m}) = (a_m, b_m)\}} + \mathbf{1}_{\{(i_{k_m}, i_{d+k_m}) = (b_m, a_m)\}}. \end{aligned}$$

This identity also gives the correct factor 2 when $a_m = b_m$. For each unselected block $t \notin I$,

$$\mathbb{E}X_{i_t}^{(t)} X_{i_{d+t}}^{(t)} = p_{i_t} \mathbf{1}_{\{i_t = i_{d+t}\}}.$$

Therefore

$$
\begin{aligned}
\mathbb{E}\,\partial_{I,a,b} f(X) &= \sum_{u\in[n]^{d-l}} \left(\prod_{t\in[d]\setminus I} p_{u_t}\right) \sum_{\varepsilon\in\{0,1\}^l} A_{\Theta_{I,\varepsilon}(a,b,u)} \\
&= T_I^p A(a,b).
\end{aligned}
\tag{4.9}
$$

Multiplying (4.9) by the derivative weight from Theorem 1 yields

$$
\left(\mathbb{E} f^{(2l)}(X) \circ \mathfrak{P}^{(2l)}\right)_{I,a,b} = \widetilde{\mathfrak{P}}^{(2l)}(a,b)\, T_I^p A(a,b) = \mathcal{A}_I^{(2l,p)}(a,b),
$$

for the canonical ordering of the derivative coordinates.

It remains only to account for the fact that the tensor $\mathbb{E} f^{(2l)}(X) \circ \mathfrak{P}^{(2l)}$ is indexed by all possible orderings of the $2l$ derivative coordinates. Every nonzero entry is obtained from one of the canonical tensors $\mathcal{A}_I^{(2l,p)}$, $|I| = l$, by a coordinate permutation and by the corresponding embedding into the appropriate block coordinates. The number of such permutations and embeddings depends only on d.

Partition norms are invariant under coordinate relabeling, up to relabeling the underlying partition. Moreover, under the same relabeling the denominator $\mathfrak{P}_S^{(2l)}(\gamma_S)$ becomes precisely the reduced marginal weight $\widetilde{\mathfrak{P}}_{\pi(S)}^{(2l)}(j_{\pi(S)})$ for the corresponding subset of coordinates. Since the right-hand side of (4.7) sums over all subsets $I_2 \subseteq [2l]$ and all partitions $\mathcal{J} \in \Pi([2l] \setminus I_2)$, these finitely many relabelings are absorbed into the constant depending only on d. Hence, for each $l = 1, \ldots, d$, the total contribution of the $m = 2l$ term in (4.8) is bounded by

$$
C_d \sum_{I_2\subseteq[2l]} \sum_{\mathcal{J}\in\Pi([2l]\setminus I_2)} r^{|\mathcal{J}|/2+|I_2|/\alpha} \max_{j_{I_2}} \sum_{\substack{I_1\subseteq[d]\\ |I_1|=l}} \frac{\left\| \left(\mathcal{A}_{I_1}^{(2l,p)}\right)_{j_{I_2^c}} \right\|_{\mathcal{J}}}{\widetilde{\mathfrak{P}}_{I_2}^{(2l)}(j_{I_2})}.
$$

Summing over $l = 1, \ldots, d$ proves (4.7). $\square$

While Theorem 3 gives a precise moment bound, the sum over all partitions is cumbersome in applications. We therefore specialize to the homogeneous setting, where the selection probabilities are uniform. Bounding the partition norms by tensor operator norms then yields the following corollary. This bound separates the sparse Gaussian-type fluctuation from the heavy-tailed fractional-exponential decay, and it is the main tool in the proof of our subspace distance theorem.

COROLLARY 3. *Assume the setting of Theorem 3. Suppose in addition that*

$$
p_1 = \cdots = p_n = p \in (0,1].
$$

Define

$$
\Gamma_{n,p,d} := p^{d/2} \max_{0\le s\le d-1} (np)^{s/2}.
$$

Then, for every $r \geq 2$,

$$
\begin{aligned}
\|f(X) - \mathbb{E} f(X)\|_{L_r} \lesssim_{\alpha,d} r^{1/2}\Gamma_{n,p,d}\, \|A\|_{[2d]} \\
+ \sum_{l=1}^{d} p^{d-l} \sum_{k=2}^{2l} r^{k/\alpha} n^{d-k/2} \|A\|_{[d],[2d]\setminus[d]}.
\end{aligned}
\tag{4.10}
$$

Proof. We apply Theorem 3. Fix $l \in [d]$, $I_1 \subseteq [d]$ with $|I_1| = l$, a subset $I_2 \subseteq [2l]$, and a partition $\mathcal{J} \in \Pi([2l] \setminus I_2)$. Put

$$q := |I_2|, \qquad \kappa := |\mathcal{J}|.$$

In the homogeneous case, the reduced tensor admits the decomposition

$$\mathcal{A}_{I_1}^{(2l,p)} = p^{d-l}\, \widetilde{\mathfrak{P}}^{(2l)} \circ \sum_{\varepsilon \in \{0,1\}^l} T_{I_1,\varepsilon} A, \tag{4.11}$$

where $T_{I_1,\varepsilon} A$ is the unweighted partial trace defined in Appendix B. We split the contribution in Theorem 3 according to whether $\kappa + q = 1$ or $\kappa + q \geq 2$.

First consider the terms with $\kappa + q = 1$. Since $\mathcal{J}$ is a partition of $[2l] \setminus I_2$, this can occur only when $I_2 = \varnothing$ and $\mathcal{J} = \{[2l]\}$. Hence the norm is the Hilbert–Schmidt norm. By (4.11), the pointwise bound $\widetilde{\mathfrak{P}}^{(2l)}(j) \leq p^{l/2}$, and (B.10),

$$\begin{aligned}
\left\| \mathcal{A}_{I_1}^{(2l,p)} \right\|_{[2l]} &\leq p^{d-l} \sum_{\varepsilon \in \{0,1\}^l} \left\| \widetilde{\mathfrak{P}}^{(2l)} \circ T_{I_1,\varepsilon} A \right\|_{[2l]} \\
&\leq p^{d-l/2} \sum_{\varepsilon \in \{0,1\}^l} \| T_{I_1,\varepsilon} A \|_{[2l]} \\
&\lesssim_d p^{d-l/2} n^{(d-l)/2} \|A\|_{[2d]}.
\end{aligned}$$

Therefore, after summing over the finitely many choices of l and I_1, the total contribution of the terms with $\kappa + q = 1$ is bounded by

$$\begin{aligned}
C_{\alpha,d} r^{1/2} \max_{1 \leq l \leq d} p^{d-l/2} n^{(d-l)/2} \|A\|_{[2d]} &= C_{\alpha,d} r^{1/2} p^{d/2} \max_{0 \leq s \leq d-1} (np)^{s/2} \|A\|_{[2d]} \\
&= C_{\alpha,d} r^{1/2} \Gamma_{n,p,d} \|A\|_{[2d]}.
\end{aligned} \tag{4.12}$$

It remains to handle the terms with $\kappa + q \geq 2$. We first remove the free weight $\widetilde{\mathfrak{P}}^{(2l)}$. If $I_2 = \varnothing$, then (B.3) and $p \leq 1$ give

$$\begin{aligned}
\left\| \mathcal{A}_{I_1}^{(2l,p)} \right\|_{\mathcal{J}} &\leq p^{d-l} \sum_{\varepsilon \in \{0,1\}^l} \left\| \widetilde{\mathfrak{P}}^{(2l)} \circ T_{I_1,\varepsilon} A \right\|_{\mathcal{J}} \\
&\lesssim_d p^{d-l/2} \sum_{\varepsilon \in \{0,1\}^l} \| T_{I_1,\varepsilon} A \|_{\mathcal{J}}.
\end{aligned} \tag{4.13}$$

If $I_2 \neq \varnothing$, then (B.4) gives, uniformly in j_{I_2},

$$\begin{aligned}
\frac{\left\| \left(\mathcal{A}_{I_1}^{(2l,p)} \right)_{j_{I_2^c}} \right\|_{\mathcal{J}}}{\widetilde{\mathfrak{P}}_{I_2}^{(2l)}(j_{I_2})} &\leq p^{d-l} \sum_{\varepsilon \in \{0,1\}^l} \frac{\left\| \left(\widetilde{\mathfrak{P}}^{(2l)} \circ T_{I_1,\varepsilon} A \right)_{j_{I_2^c}} \right\|_{\mathcal{J}}}{\widetilde{\mathfrak{P}}_{I_2}^{(2l)}(j_{I_2})} \\
&\lesssim_d p^{d-l} \sum_{\varepsilon \in \{0,1\}^l} \left\| (T_{I_1,\varepsilon} A)_{j_{I_2^c}} \right\|_{\mathcal{J}}.
\end{aligned} \tag{4.14}$$

Combining (4.13) and (4.14), and then applying (B.9), we obtain, for all terms with $\kappa + q \geq 2$,

$$\frac{\left\| \left(\mathcal{A}_{I_1}^{(2l,p)} \right)_{j_{I_2^c}} \right\|_{\mathcal{J}}}{\widetilde{\mathfrak{P}}_{I_2}^{(2l)}(j_{I_2})} \lesssim_d p^{d-l} n^{d-(\kappa+q)/2} \|A\|_{[d],[2d]\setminus[d]}.$$

Here, when $I_2 = \varnothing$, the denominator is interpreted as one.

The corresponding factor in Theorem 3 is $r^{\kappa/2+q/\alpha}$. Since $0 < \alpha \leq 1$ and $r \geq 2$,

$$r^{\kappa/2+q/\alpha} \leq r^{(\kappa+q)/\alpha}.$$

Set $k := \kappa + q$. Since we are in the case $k \geq 2$ and always have $k \leq 2d$, summing over all choices of l, I_1, I_2, and $\mathcal{J}$ only changes the implicit constant by a factor depending on d. Hence the total contribution of the terms with $\kappa + q \geq 2$ is bounded by

$$C_{\alpha,d} \sum_{l=1}^{d} p^{d-l} \sum_{k=2}^{2l} r^{k/\alpha} n^{d-k/2} \|A\|_{[d],[2d]\setminus[d]}. \tag{4.15}$$

Combining (4.12) and (4.15) proves (4.10). □

4.2.2 Distance to subspaces for random tensors

Proof of Theorem 2. Assume $B \neq 0$ and set

$$M := B^\top B.$$

We view M as a $2d$-tensor by writing

$$A_{i_1,\ldots,i_d,i_{d+1},\ldots,i_{2d}} := M_{(i_1,\ldots,i_d),(i_{d+1},\ldots,i_{2d})}.$$

Then

$$X^\top M X = \sum_{i_1,\ldots,i_{2d}=1}^{n} A_{i_1,\ldots,i_{2d}} \prod_{k=1}^{d} X_{i_k}^{(k)} X_{i_{d+k}}^{(k)}.$$

We first record the quadratic-form consequence of Corollary 3. For every $r \geq 2$,

$$\left\| X^\top M X - \mathbb{E} X^\top M X \right\|_{L_r} \lesssim_{\alpha,d} r^{1/2} \Gamma_{n,p,d} \|M\|_{\mathrm{HS}} + \sum_{k=2}^{2d} r^{k/\alpha} n^{d-k/2} \|M\|_{\mathrm{op}}. \tag{4.16}$$

Since the coordinates of the simple tensor are isotropic up to the factor p^d, we have

$$\mathbb{E} X^\top M X = p^d \operatorname{tr}(M) = p^d \|B\|_{\mathrm{HS}}^2. \tag{4.17}$$

Put

$$Y := X^\top M X - \mathbb{E} X^\top M X, \qquad b := p^{d/2} \|B\|_{\mathrm{HS}}.$$

For non-negative real numbers a, b with $b > 0$,

$$|a - b| \le \min\left\{ \frac{|a^2 - b^2|}{b}, \sqrt{|a^2 - b^2|} \right\}.$$

Applying this with $a = \|BX\|_2$ and using (4.17), we obtain

$$\Big\| \|BX\|_2 - p^{d/2}\|B\|_{\mathrm{HS}} \Big\|_{L_r} \le \min\left\{ \frac{\|Y\|_{L_r}}{b}, \left\| \sqrt{|Y|} \right\|_{L_r} \right\}.$$

Since $r \ge 2$,

$$\left\| \sqrt{|Y|} \right\|_{L_r} = \|Y\|_{L_{r/2}}^{1/2} \le \|Y\|_{L_r}^{1/2}.$$

Let

$$g(x) := \min\{x/b, \sqrt{x}\}, \qquad x \ge 0.$$

The function g is non-negative, increasing, concave, and satisfies $g(0) = 0$, hence it is subadditive. Applying g to the right-hand side of (4.16), using $g(x) \le x/b$ for the Hilbert–Schmidt term and $g(x) \le \sqrt{x}$ for the remaining terms, we get

$$\begin{aligned} \Big\| \|BX\|_2 - p^{d/2}\|B\|_{\mathrm{HS}} \Big\|_{L_r} \lesssim_{\alpha,d} r^{1/2} \frac{\Gamma_{n,p,d}}{p^{d/2}} \frac{\|B^\top B\|_{\mathrm{HS}}}{\|B\|_{\mathrm{HS}}} \\ + \sum_{k=2}^{2d} r^{k/(2\alpha)} n^{d/2-k/4} \sqrt{\|B^\top B\|_{\mathrm{op}}}. \end{aligned}$$

Using

$$\|B^\top B\|_{\mathrm{HS}} \le \|B\|_{\mathrm{op}}\|B\|_{\mathrm{HS}}, \qquad \|B^\top B\|_{\mathrm{op}} = \|B\|_{\mathrm{op}}^2,$$

this becomes

$$\Big\| \|BX\|_2 - p^{d/2}\|B\|_{\mathrm{HS}} \Big\|_{L_r} \lesssim_{\alpha,d} \|B\|_{\mathrm{op}} \left(r^{1/2}\Lambda_{n,p,d} + \sum_{k=2}^{2d} r^{k/(2\alpha)} n^{d/2-k/4} \right). \tag{4.18}$$

By the standard moment-to-tail conversion lemma, see, e.g., [13, Lemma 2.1], (4.18) implies that, for every $t > 0$,

$$\begin{aligned} \mathbb{P}\left\{ \Big| \|BX\|_2 - p^{d/2}\|B\|_{\mathrm{HS}} \Big| > t \right\} \le 2\exp\Bigg(- c(\alpha, d) \min\Bigg\{ \frac{t^2}{\Lambda_{n,p,d}^2 \|B\|_{\mathrm{op}}^2}, \\ \min_{2 \le k \le 2d} \left(\frac{t}{n^{d/2-k/4}\|B\|_{\mathrm{op}}} \right)^{2\alpha/k} \Bigg\} \Bigg). \end{aligned} \tag{4.19}$$

It remains to simplify the second minimum. Let

$$x := \frac{t}{n^{d/2}\|B\|_{\mathrm{op}}}.$$

For $2 \le k \le 2d$,

$$\left(\frac{t}{n^{d/2-k/4}\|B\|_{\text{op}}}\right)^{2\alpha/k} = n^{\alpha/2}x^{2\alpha/k}.$$

If $x \le 1$, the minimum over $2 \le k \le 2d$ is attained at $k = 2$, giving

$$n^{\alpha/2}x^{\alpha} = \left(\frac{t}{n^{(d-1)/2}\|B\|_{\text{op}}}\right)^{\alpha}.$$

If $x > 1$, the minimum is attained at $k = 2d$, giving

$$n^{\alpha/2}x^{\alpha/d} = \left(\frac{t}{\|B\|_{\text{op}}}\right)^{\alpha/d}.$$

Thus the second minimum in (4.19) is bounded below by

$$\min\left\{\left(\frac{t}{n^{(d-1)/2}\|B\|_{\text{op}}}\right)^{\alpha}, \left(\frac{t}{\|B\|_{\text{op}}}\right)^{\alpha/d}\right\}.$$

□

4.2.3 The Singularity of simple tensor matrices

Proof of Corollary 2. Let $N = n^d$. For each $j \in [m]$, set

$$H_j := \text{span}\{\mathbf{X}_i : i \ne j\}, \qquad Q_j := P_{H_j^{\perp}}, \qquad k_j := \text{rank}(Q_j).$$

Since $m \le (1-\varepsilon)N$, we have

$$k_j = \text{codim}(H_j) \ge N - m + 1 \ge \varepsilon N.$$

Put

$$\eta := p^d \varepsilon n.$$

By the assumption on ε, after choosing $c_1 = c_1(\alpha, d)$ sufficiently large, we may assume that $\eta \ge 1$. Moreover, since $\varepsilon < 1$, the range of parameters is non-empty only if $np^d \ge c_1(d\log n)^{2/\alpha}$. Hence, increasing c_1 if necessary, $np \ge np^d \ge 1$, and therefore

$$\Lambda_{n,p,d} = \max_{0 \le s \le d-1}(np)^{s/2} = (np)^{(d-1)/2}.$$

Conditioning on all columns except $\mathbf{X}_j$, the projection Q_j is fixed and $\mathbf{X}_j$ is independent of Q_j. Applying Theorem 2 with $B = Q_j$ and using

$$\|Q_j\|_{\text{op}} = 1, \qquad \|Q_j\|_{\text{HS}} = \sqrt{k_j},$$

we get

$$\begin{aligned}
&\mathbb{P}\left\{\text{dist}(\mathbf{X}_j, H_j) \le \frac{1}{2}p^{d/2}\sqrt{\varepsilon N} \,\middle|\, H_j\right\} \\
&\quad \le \mathbb{P}\left\{\|Q_j\mathbf{X}_j\|_2 \le \frac{1}{2}p^{d/2}\sqrt{k_j} \,\middle|\, H_j\right\} \\
&\quad \le 2\exp\left[-c(\alpha, d)\min\left\{\frac{p^d k_j}{(np)^{d-1}}, \left(\frac{p^{d/2}\sqrt{k_j}}{n^{(d-1)/2}}\right)^{\alpha}, \left(p^{d/2}\sqrt{k_j}\right)^{\alpha/d}\right\}\right].
\end{aligned}$$

Since $k_j \geq \varepsilon n^d$, $0 < p \leq 1$, $0 < \varepsilon < 1$, and $\eta \geq 1$, the three terms in the minimum are bounded from below, up to constants depending only on α, d, by

$$p\varepsilon n, \qquad (p^d \varepsilon n)^{\alpha/2}, \qquad p^{\alpha/2}\varepsilon^{\alpha/(2d)} n^{\alpha/2}.$$

Each of these is at least $c_{\alpha,d}\eta^{\alpha/2}$. Indeed,

$$p\varepsilon n \geq p^d \varepsilon n = \eta \geq \eta^{\alpha/2},$$

and

$$p^{\alpha/2}\varepsilon^{\alpha/(2d)} n^{\alpha/2} \geq p^{d\alpha/2}\varepsilon^{\alpha/2} n^{\alpha/2} = \eta^{\alpha/2}.$$

Therefore,

$$\mathbb{P}\left\{\operatorname{dist}(\mathbf{X}_j, H_j) \leq \frac{1}{2} p^{d/2}\sqrt{\varepsilon N}\right\} \leq 2\exp\left(-c_{\alpha,d}\eta^{\alpha/2}\right).$$

Taking a union bound over $j = 1, \ldots, m$, and using $m \leq N = n^d$, gives

$$\begin{aligned}\mathbb{P}\left\{\exists j \in [m] : \operatorname{dist}(\mathbf{X}_j, H_j) \leq \frac{1}{2} p^{d/2}\sqrt{\varepsilon N}\right\} &\leq 2n^d \exp\left(-c_{\alpha,d}\eta^{\alpha/2}\right) \\ &\leq 2\exp\left(-c_2(\alpha, d)\eta^{\alpha/2}\right),\end{aligned}$$

where the last inequality follows from

$$\eta^{\alpha/2} = (p^d \varepsilon n)^{\alpha/2} \geq c_1^{\alpha/2} d \log n$$

and from choosing c_1 sufficiently large.

On the complementary event,

$$\operatorname{dist}(\mathbf{X}_j, H_j) > \frac{1}{2} p^{d/2}\sqrt{\varepsilon N}, \qquad j = 1, \ldots, m.$$

Let $s_\ell(\mathbf{X})$ denote the singular values of $\mathbf{X}$. By the negative second moment identity,

$$\sum_{\ell=1}^{m} s_\ell(\mathbf{X})^{-2} = \sum_{j=1}^{m} \operatorname{dist}(X_j, H_j)^{-2}.$$

Thus

$$\sigma_{\min}(\mathbf{X})^{-2} \leq \sum_{\ell=1}^{m} s_\ell(\mathbf{X})^{-2} = \sum_{j=1}^{m} \operatorname{dist}(\mathbf{X}_j, H_j)^{-2} < \frac{4m}{p^d \varepsilon N} \leq \frac{4(1-\varepsilon)}{p^d \varepsilon}.$$

Equivalently,

$$\sigma_{\min}(\mathbf{X}) > \frac{1}{2} p^{d/2}\sqrt{\frac{\varepsilon}{1-\varepsilon}}.$$

Since $\eta = p^d \varepsilon n$, the desired estimate follows:

$$\mathbb{P}\left\{\sigma_{\min}(\mathbf{X}) > \frac{1}{2} p^{d/2}\sqrt{\frac{\varepsilon}{1-\varepsilon}}\right\} \geq 1 - 2\exp\left(-c_2(\alpha, d)(p^d \varepsilon n)^{\alpha/2}\right).$$

$\square$

A Supplementary Proof of Theorem 1

In this appendix we prove the estimate used in the proof of Theorem 1: for every $d = 1, \dots, D-1$, every $I \subseteq [d]$, every fixed $\mathbf{i}_I$, and every $\mathcal{J} \in \Pi([d] \setminus I)$,

$$\left\| (R^{(d)} \circ \mathfrak{p}^{(d)})_{\mathbf{i}_{I^c}} \right\|_{\mathcal{J}} \lesssim_{\alpha,D} \sum_{q=d+1}^{D} \sum_{\substack{\mathcal{K} \in \Pi([q] \setminus I) \\ |\mathcal{K}| \geq |\mathcal{J}|}} \left\| (A_q \circ \mathfrak{p}^{(q)})_{\mathbf{i}_{I^c}} \right\|_{\mathcal{K}}. \tag{A.1}$$

Here, when $q > d$, the set $I \subseteq [d]$ is identified with the same subset of $[q]$.

It is enough to prove the localized estimate

$$\left\| (R^{(d)} \circ \mathfrak{p}^{(d)} \circ \mathbf{1}_{L(\mathcal{I})})_{\mathbf{i}_{I^c}} \right\|_{\mathcal{J}} \lesssim_{\alpha,D} \sum_{q=d+1}^{D} \sum_{\substack{\mathcal{K} \in \Pi([q] \setminus I) \\ |\mathcal{K}| \geq |\mathcal{J}|}} \left\| (A_q \circ \mathfrak{p}^{(q)})_{\mathbf{i}_{I^c}} \right\|_{\mathcal{K}}$$

for every $\mathcal{I} \in \Pi_D$, since

$$R^{(d)} = \sum_{\mathcal{I} \in \Pi_D} R^{(d)} \circ \mathbf{1}_{L(\mathcal{I})}$$

and the number of partitions of $[d]$ depends only on d.

Fix

$$\mathcal{I} = \{I_1, \dots, I_\nu\} \in \Pi_d, \qquad l_\beta := |I_\beta|, \quad \beta = 1, \dots, \nu.$$

For $\mathbf{i} \in L(\mathcal{I})$, denote by $j_\beta = j_\beta(\mathbf{i})$ the common value of the coordinates $\{i_r, r \in I_\beta\}$.

Let

$$\mathbf{k} = (k_1, \dots, k_\nu) \in \mathbb{N}_+^\nu$$

satisfy

$$k_\beta \geq l_\beta \quad \text{for all } \beta \in [\nu], \qquad \mathbf{k} \neq (l_1, \dots, l_\nu), \qquad |\mathbf{k}| := k_1 + \cdots + k_\nu \leq D.$$

In particular, $q = |\mathbf{k}| > d$.

Define the d-tensor

$$S_{\mathcal{I}}^{(d,\mathbf{k})} = \left(s_{\mathbf{i}}^{(d,\mathbf{k})} \right)_{\mathbf{i} \in [n]^d}$$

by

$$s_{\mathbf{i}}^{(d,\mathbf{k})} = \mathbf{1}_{\{\mathbf{i} \in L(\mathcal{I})\}}\, a_{j_1,\dots,j_\nu}^{\mathbf{k}} \prod_{\beta=1}^{\nu} \mathbb{E} X_{j_\beta}^{k_\beta - l_\beta},$$

where j_β is the value of $\mathbf{i}$ on the level set I_β, and $a_{j_1,\dots,j_\nu}^{\mathbf{k}}$ denotes the coefficient tensor used in the definition of A_q on level sets of sizes $k_1, \dots, k_\nu$.

By the definition of $R^{(d)}$, we have

$$R^{(d)} \circ \mathbf{1}_{L(\mathcal{I})} = \sum_{\substack{\mathbf{k} \in \mathbb{N}_+^\nu \\ k_\beta \geq l_\beta,\ \mathbf{k} \neq (l_1,\dots,l_\nu),\ |\mathbf{k}| \leq D}} \nu! \prod_{\beta=1}^{\nu} \frac{k_\beta!}{(k_\beta - l_\beta)!} S_{\mathcal{I}}^{(d,\mathbf{k})}. \tag{A.2}$$

The sum is finite and all combinatorial coefficients are bounded by constants depending only on D. Hence it is enough to prove, for every such $\mathbf{k}$,

$$\left\|(S_{\mathcal{I}}^{(d,\mathbf{k})} \circ \mathfrak{p}^{(d)})_{\mathbf{i}_{I^c}}\right\|_{\mathcal{J}} \lesssim_{\alpha,D} \sum_{\substack{\mathcal{K}\in\Pi([q]\setminus I)\\ |\mathcal{K}|\geq|\mathcal{J}|}} \left\|(A_q \circ \mathfrak{p}^{(q)})_{\mathbf{i}_{I^c}}\right\|_{\mathcal{K}}, \qquad q = |\mathbf{k}|. \tag{A.3}$$

Fix such a vector $\mathbf{k}$. Choose a partition

$$\widetilde{\mathcal{I}} = \{\widetilde{I}_1, \ldots, \widetilde{I}_\nu\} \in \Pi_q$$

such that

$$I_\beta \subseteq \widetilde{I}_\beta, \qquad |\widetilde{I}_\beta| = k_\beta, \qquad \beta = 1, \ldots, \nu.$$

For each β, let

$$\rho_\beta := \min I_\beta.$$

Define a map $\pi : [q] \to [d]$ by

$$\pi(r) = \begin{cases} r, & r \leq d, \\ \rho_\beta, & r \in \widetilde{I}_\beta \setminus I_\beta. \end{cases}$$

Thus, if $u \in L(\widetilde{\mathcal{I}})$, then

$$u_r = u_{\pi(r)}, \qquad r = d+1, \ldots, q.$$

Define the q-tensor

$$\widetilde{S}^{(q,\mathbf{k})} = \left(\widetilde{s}_u^{(q,\mathbf{k})}\right)_{u\in[n]^q}$$

by

$$\widetilde{s}_u^{(q,\mathbf{k})} = s_{u_{[d]}}^{(d,\mathbf{k})} \mathbf{1}_{\{u\in L(\widetilde{\mathcal{I}})\}}. \tag{A.4}$$

By construction of A_q, for $u \in L(\widetilde{\mathcal{I}})$,

$$\widetilde{s}_u^{(q,\mathbf{k})} = (A_q)_u \prod_{\beta=1}^{\nu} \mathbb{E}X_{u_{\rho_\beta}}^{k_\beta - l_\beta}. \tag{A.5}$$

Moreover, on $L(\widetilde{\mathcal{I}})$, the coordinates $d+1, \ldots, q$ only repeat old coordinates. Therefore the set of distinct values of u is the same as the set of distinct values of $u_{[d]}$, and hence

$$\mathfrak{p}_u^{(q)} = \mathfrak{p}_{u_{[d]}}^{(d)}, \qquad u \in L(\widetilde{\mathcal{I}}). \tag{A.6}$$

We now construct a partition of $[q] \setminus I$ from $\mathcal{J}$. Write

$$\mathcal{J} = \{J_1, \ldots, J_\mu\} \in \Pi([d] \setminus I).$$

Initialize

$$K_a := J_a, \qquad a = 1, \ldots, \mu, \qquad K_{\mu+1} := \varnothing.$$

For every $r = d+1, \dots, q$, let $t = \pi(r) \in [d]$. If $t \in [d] \setminus I$, let $a \in \{1, \dots, \mu\}$ be the unique index such that $t \in J_a$, and put

$$K_a := K_a \cup \{r\}.$$

If $t \in I$, put

$$K_{\mu+1} := K_{\mu+1} \cup \{r\}.$$

If $K_{\mu+1} = \varnothing$, we discard it. The resulting family, still denoted by $\mathcal{K}$, is a partition of $[q] \setminus I$, and

$$|\mathcal{K}| \in \{|\mathcal{J}|, |\mathcal{J}|+1\}. \tag{A.7}$$

We claim that

$$\left\| (S_{\mathcal{I}}^{(d,\mathbf{k})} \circ \mathfrak{p}^{(d)})_{\mathbf{i}_{I^c}} \right\|_{\mathcal{J}} \le \left\| (\widetilde{S}^{(q,\mathbf{k})} \circ \mathfrak{p}^{(q)})_{\mathbf{i}_{I^c}} \right\|_{\mathcal{K}}. \tag{A.8}$$

Let

$$x^{(a)} = \left(x^{(a)}_{i_{J_a}} \right)_{i_{J_a}}, \qquad a = 1, \dots, \mu,$$

be test tensors with

$$\|x^{(a)}\|_{\mathrm{HS}} \le 1.$$

We define test tensors for the partition $\mathcal{K}$. For $a = 1, \dots, \mu$, set

$$y^{(a)}_{u_{K_a}} = x^{(a)}_{u_{J_a}} \prod_{r \in K_a \setminus [d]} \mathbf{1}_{\{u_r = u_{\pi(r)}\}}.$$

This is well-defined because if $r \in K_a \setminus [d]$, then $\pi(r) \in J_a \subseteq K_a$. Also,

$$\|y^{(a)}\|_{\mathrm{HS}} \le \|x^{(a)}\|_{\mathrm{HS}} \le 1,$$

since the indicator factors determine the extra coordinates $K_a \setminus [d]$ once u_{J_a} is fixed.

If $K_{\mu+1} \neq \varnothing$, define

$$y^{(\mu+1)}_{u_{K_{\mu+1}}} = \prod_{r \in K_{\mu+1}} \mathbf{1}_{\{u_r = i_{\pi(r)}\}}.$$

Then

$$\|y^{(\mu+1)}\|_{\mathrm{HS}} \le 1,$$

because this tensor has at most one nonzero entry.

Using (A.4) and (A.6), we obtain

$$\begin{aligned}
&\sum_{i_{[d]\setminus I}} (S_{\mathcal{I}}^{(d,\mathbf{k})} \circ \mathfrak{p}^{(d)})_i \prod_{a=1}^{\mu} x^{(a)}_{i_{J_a}} \\
&\quad = \sum_{u_{[q]\setminus I}} (\widetilde{S}^{(q,\mathbf{k})} \circ \mathfrak{p}^{(q)})_u \prod_{a=1}^{\mu} y^{(a)}_{u_{K_a}} \cdot \begin{cases} 1, & K_{\mu+1} = \varnothing, \\ y^{(\mu+1)}_{u_{K_{\mu+1}}}, & K_{\mu+1} \neq \varnothing. \end{cases}
\end{aligned}$$

Taking the supremum over all admissible $x^{(a)}$'s proves (A.8).

It remains to compare $\widetilde{S}^{(q,\mathbf{k})} \circ \mathfrak{p}^{(q)}$ with $A_q \circ \mathfrak{p}^{(q)}$. For $r = 1, \dots, q$, define vectors $v_r \in \mathbb{R}^n$ by

$$v_{\rho_\beta}(a) = \mathbb{E} X_a^{k_\beta - l_\beta}, \qquad a \in [n], \quad \beta = 1, \dots, \nu,$$

and

$$v_r(a) = 1$$

for all remaining r. Since $\|X_a\|_{\Psi_\alpha} \le 1$ and $0 \le k_\beta - l_\beta \le D$, we have

$$\|v_r\|_\infty \lesssim_{\alpha,D} 1, \qquad r = 1, \dots, q.$$

By (A.5),

$$\widetilde{S}^{(q,\mathbf{k})} \circ \mathfrak{p}^{(q)} = \Big(A_q \circ \mathfrak{p}^{(q)} \circ \mathbf{1}_{L(\widetilde{\mathcal{I}})}\Big) \circ \bigotimes_{r=1}^{q} v_r.$$

Therefore, applying Lemma 3.4 (5) in [16] to the tensor $A_q \circ \mathfrak{p}^{(q)} \circ \mathbf{1}_{L(\widetilde{\mathcal{I}})}$, we get

$$\left\|(\widetilde{S}^{(q,\mathbf{k})} \circ \mathfrak{p}^{(q)})_{\mathbf{i}_{I^c}}\right\|_{\mathcal{K}} \lesssim_{\alpha,D} \left\|(A_q \circ \mathfrak{p}^{(q)} \circ \mathbf{1}_{L(\widetilde{\mathcal{I}})})_{\mathbf{i}_{I^c}}\right\|_{\mathcal{K}}.$$

Then Lemma 4.3 gives

$$\left\|(A_q \circ \mathfrak{p}^{(q)} \circ \mathbf{1}_{L(\widetilde{\mathcal{I}})})_{\mathbf{i}_{I^c}}\right\|_{\mathcal{K}} \lesssim_{D} \left\|(A_q \circ \mathfrak{p}^{(q)})_{\mathbf{i}_{I^c}}\right\|_{\mathcal{K}}.$$

Combining the last two displays with (A.8), we obtain

$$\left\|(S_{\mathcal{I}}^{(d,\mathbf{k})} \circ \mathfrak{p}^{(d)})_{\mathbf{i}_{I^c}}\right\|_{\mathcal{J}} \lesssim_{\alpha,D} \left\|(A_q \circ \mathfrak{p}^{(q)})_{\mathbf{i}_{I^c}}\right\|_{\mathcal{K}}.$$

By (A.7), this implies

$$\left\|(S_{\mathcal{I}}^{(d,\mathbf{k})} \circ \mathfrak{p}^{(d)})_{\mathbf{i}_{I^c}}\right\|_{\mathcal{J}} \lesssim_{\alpha,D} \sum_{\substack{\mathcal{K}' \in \Pi([q]\setminus I) \\ |\mathcal{K}'| \ge |\mathcal{J}|}} \left\|(A_q \circ \mathfrak{p}^{(q)})_{\mathbf{i}_{I^c}}\right\|_{\mathcal{K}'}.$$

This proves (A.3).

Finally, summing over the finitely many admissible $\mathbf{k}$'s in (A.2), and then over the finitely many level-set partitions $\mathcal{I} \in \Pi_D$, proves (A.1). Hence (4.5) follows.

B Supplementary estimates for Corollary 3

In this appendix we collect the deterministic tensor-norm estimates used in the proof of Corollary 3. Throughout this appendix, set

$$\mathsf{L} := [d], \qquad \mathsf{R} := d + [d] = \{d+1, \dots, 2d\}.$$

Thus

$$\|A\|_{\mathsf{L},\mathsf{R}} = \|A\|_{[d],[2d]\setminus[d]}.$$

The unweighted trace tensors

Let

$$I = \{k_1, \dots, k_l\} \subseteq [d], \qquad k_1 < \cdots < k_l,$$

and let $a = (a_1, \dots, a_l), b = (b_1, \dots, b_l) \in [n]^l$. For $u = (u_t)_{t\in[d]\setminus I} \in [n]^{d-l}$ and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_l) \in \{0,1\}^l$, recall the filling map

$$\Theta_{I,\varepsilon}(a, b, u) = (i_1, \dots, i_{2d}) \in [n]^{2d},$$

defined by

$$i_t = i_{d+t} = u_t, \qquad t \in [d] \setminus I,$$

and, for $t = k_m \in I$,

$$(i_{k_m}, i_{d+k_m}) = \begin{cases} (a_m, b_m), & \varepsilon_m = 0, \\ (b_m, a_m), & \varepsilon_m = 1. \end{cases}$$

We define the unweighted partial trace associated with one selected-pair swap pattern by

$$T_{I,\varepsilon}A(a, b) := \sum_{u\in[n]^{d-l}} A_{\Theta_{I,\varepsilon}(a,b,u)}, \qquad a, b \in [n]^l. \tag{B.1}$$

In the homogeneous case $p_1 = \cdots = p_n = p$, the reduced tensor from Theorem 3 can be written as

$$\mathcal{A}_I^{(2l,p)} = p^{d-l}\,\widetilde{\mathfrak{P}}^{(2l)} \circ \sum_{\varepsilon\in\{0,1\}^l} T_{I,\varepsilon}A.$$

For later use, define the bijection

$$\rho_{I,\varepsilon} : [2l] \to I \cup (I + d)$$

by

$$\rho_{I,\varepsilon}(m) = \begin{cases} k_m, & \varepsilon_m = 0, \\ d + k_m, & \varepsilon_m = 1, \end{cases} \qquad \rho_{I,\varepsilon}(l + m) = \begin{cases} d + k_m, & \varepsilon_m = 0, \\ k_m, & \varepsilon_m = 1, \end{cases} \qquad m \in [l]. \tag{B.2}$$

If $\mathcal{J}$ is a partition of a subset of $[2l]$, then $\rho_{I,\varepsilon}(\mathcal{J})$ denotes the partition obtained by applying $\rho_{I,\varepsilon}$ to every block of $\mathcal{J}$.

LEMMA B.1 (Lemma 23 in [5]). *Let $B = (B_i)_{i\in[n]^m}$ be an m-tensor and let $\mathcal{K} = \{K_1, \dots, K_s\}$ be a partition of $[m]$. Suppose that $K_s = K'_s \cup K'_{s+1}$ is a disjoint union. Then*

$$\|B\|_{K_1,\dots,K_{s-1},K'_s,K'_{s+1}} \le \|B\|_{K_1,\dots,K_s} \le n^{\min\{|K'_s|,|K'_{s+1}|\}/2}\|B\|_{K_1,\dots,K_{s-1},K'_s,K'_{s+1}}.$$

LEMMA B.2. *Assume $p_1 = \cdots = p_n = p \in (0, 1]$. Let $l \in [d]$ and let $B = (B_j)_{j\in[n]^{2l}}$ be a $2l$-tensor.*

1. *For every partition $\mathcal{J} \in \Pi([2l])$,*

$$\left\|\widetilde{\mathfrak{P}}^{(2l)} \circ B\right\|_{\mathcal{J}} \lesssim_d p^{l/2}\|B\|_{\mathcal{J}}. \tag{B.3}$$

2. *Let* $I_2 \subseteq [2l]$*, fix* $j_{I_2} \in [n]^{I_2}$*, and let* $\mathcal{J} \in \Pi([2l] \setminus I_2)$*. Then*

$$\frac{\|(\widetilde{\mathfrak{P}}^{(2l)} \circ B)_{j_{I_2^c}}\|_{\mathcal{J}}}{\widetilde{\mathfrak{P}}^{(2l)}_{I_2}(j_{I_2})} \lesssim_d \|B_{j_{I_2^c}}\|_{\mathcal{J}}. \tag{B.4}$$

Proof. Write the multi-index as $j = (a, b) = (a_1, \dots, a_l, b_1, \dots, b_l)$. In the homogeneous case, the global weight tensor factors completely over the block structure:

$$\widetilde{\mathfrak{P}}^{(2l)}(a, b) = \prod_{m=1}^{l} q(a_m, b_m),$$

where the pairwise weight function is given by

$$q(a_m, b_m) = p\,\mathbf{1}_{\{a_m \neq b_m\}} + \sqrt{p}\,\mathbf{1}_{\{a_m = b_m\}} = p + (\sqrt{p} - p)\mathbf{1}_{\{a_m = b_m\}}.$$

Expanding the product over $m = 1, \dots, l$, the weight tensor can be expressed as a finite linear combination of indicator functions associated with specific equality patterns:

$$\widetilde{\mathfrak{P}}^{(2l)}(a, b) = \sum_{S \subseteq [l]} p^{l-|S|}(\sqrt{p} - p)^{|S|}\mathbf{1}_{E_S}(a, b),$$

where $E_S = \{(a, b) : a_m = b_m \text{ for all } m \in S\}$. Geometrically, each set E_S is an intersection of generalized diagonals. According to Lemma 3.4 in [16], the operation of taking the Hadamard product with the indicator function of a generalized diagonal (or intersections thereof) constitutes a bounded multiplier on any partition norm $\|\cdot\|_{\mathcal{J}}$. Therefore, there exists a constant $C_d > 0$ depending only on the tensor dimension d such that $\|\mathbf{1}_{E_S} \circ B\|_{\mathcal{J}} \le C_d\|B\|_{\mathcal{J}}$ for all $S \subseteq [l]$.

Since $p \in (0, 1]$, we have $\sqrt{p} - p \ge 0$. The total variation (the sum of the absolute values of the coefficients) converges exactly to $p^{l/2}$ by the binomial theorem:

$$\sum_{S \subseteq [l]} \left|p^{l-|S|}(\sqrt{p} - p)^{|S|}\right| = \sum_{S \subseteq [l]} p^{l-|S|}(\sqrt{p} - p)^{|S|} = \big(p + (\sqrt{p} - p)\big)^{l} = p^{l/2}.$$

Applying the triangle inequality for the partition norm gives:

$$\big\|\widetilde{\mathfrak{P}}^{(2l)} \circ B\big\|_{\mathcal{J}} \le \sum_{S \subseteq [l]} p^{l-|S|}(\sqrt{p} - p)^{|S|}\big\|\mathbf{1}_{E_S} \circ B\big\|_{\mathcal{J}} \lesssim_d p^{l/2}\|B\|_{\mathcal{J}}.$$

This establishes (B.3).

For the sliced estimate (B.4), we fix a subset of coordinates $I_2 \subseteq [2l]$ with values j_{I_2}. The condition weight multiplier applied to the remaining free coordinates $j_{I_2^c}$ is the quotient

$$W(j_{I_2^c}) = \frac{\widetilde{\mathfrak{P}}^{(2l)}(j)}{\widetilde{\mathfrak{P}}^{(2l)}_{I_2}(j_{I_2})}.$$

Due to the independence of the blocks, this quotient preserves the product structure, yielding $W(j_{I_2^c}) = \prod_{m=1}^{l} w_m$. For any given block m, the factor w_m falls into one of three mutually exclusive cases depending on the fixed coordinates:

1. **Neither coordinate is fixed** ($a_m, b_m \notin I_2$): $w_m = q(a_m, b_m) = p + (\sqrt{p} - p)\mathbf{1}_{\{a_m=b_m\}}$. The sum of absolute coefficients is $p + (\sqrt{p} - p) = \sqrt{p} \le 1$.

2. **Exactly one coordinate is fixed** (e.g., $b_m = x_0 \in j_{I_2}$ is fixed and $a_m = x$ is free): Dividing the joint weight by the fixed marginal weight yields $w_m = \sqrt{p} + (1 - \sqrt{p})\mathbf{1}_{\{x=x_0\}}$. The sum of absolute coefficients is $\sqrt{p} + (1 - \sqrt{p}) = 1$.

3. **Both coordinates are fixed** ($a_m, b_m \in I_2$): The factor identically cancels out, yielding $w_m = 1$. The sum of absolute coefficients is 1.

Expanding the product $W(j_{I_2^c}) = \prod_{m=1}^{l} w_m$ generates a new linear combination $W(j_{I_2^c}) = \sum_K \tilde{c}_K \mathbf{1}_{E'_K}$, where each E'_K represents either an equality pattern strictly among the free coordinates, or an equality forcing a free coordinate to match a fixed value (which corresponds to a coordinate projection after slicing). The total variation of this expansion is simply the product of the absolute coefficient sums of the individual blocks:

$$\sum_K |\tilde{c}_K| \lesssim_d 1.$$

Applying the triangle inequality and Lemma 3.4 in [16] yields:

$$\frac{\|(\widetilde{\mathfrak{P}}^{(2l)} \circ B)_{j_{I_2^c}}\|_{\mathcal{J}}}{\widetilde{\mathfrak{P}}^{(2l)}_{I_2}(j_{I_2})} = \Big\|\sum_K \tilde{c}_K \big(\mathbf{1}_{E'_K} \circ B_{j_{I_2^c}}\big)\Big\|_{\mathcal{J}} \lesssim_d \|B_{j_{I_2^c}}\|_{\mathcal{J}}.$$

This establishes (B.4) and completes the proof. ☐

LEMMA B.3 (Sliced partition norms versus the middle flattening)**.** *Let $F \subseteq [2d]$ with $|F| = q$, and fix $j_F \in [n]^F$. Let $\mathcal{K} = \{K_1, \dots, K_m\}$ be a partition of $F^c = [2d] \setminus F$. Then*

$$\|A_{j_F^c}\|_{\mathcal{K}} \le n^{d-(m+q)/2}\|A\|_{\mathsf{L},\mathsf{R}}. \tag{B.5}$$

Proof. For each block $K_s \in \mathcal{K}$, set

$$K_s^- := K_s \cap \mathsf{L}, \qquad K_s^+ := K_s \cap \mathsf{R}.$$

Let $\mathcal{K}'$ be the partition obtained from $\mathcal{K}$ by replacing every nonempty mixed block K_s by the two nonempty blocks K_s^- and K_s^+. Repeatedly applying Lemma B.1 yields

$$\|A_{j_F^c}\|_{\mathcal{K}} \le n^{\frac{1}{2}\sum_{s=1}^{m} \min\{|K_s^-|,|K_s^+|\}}\|A_{j_F^c}\|_{\mathcal{K}'}. \tag{B.6}$$

The partition $\mathcal{K}'$ refines the two-block partition

$$\{\mathsf{L} \setminus F,\, \mathsf{R} \setminus F\},$$

with empty blocks omitted. By monotonicity of partition norms under refinement, and by embedding the test vectors into the fixed coordinates j_F, we have

$$\|A_{j_F^c}\|_{\mathcal{K}'} \le \|A_{j_F^c}\|_{\mathsf{L}\setminus F,\, \mathsf{R}\setminus F} \le \|A\|_{\mathsf{L},\mathsf{R}}. \tag{B.7}$$

Finally,

$$\sum_{s=1}^{m} \min\{|K_s^-|, |K_s^+|\} \le \sum_{s=1}^{m}(|K_s| - 1) = |F^c| - m = 2d - q - m.$$

Combining this estimate with (B.6) and (B.7) proves (B.5). ☐

Lemma B.4. *Let $A = (A_{i_1,\dots,i_{2d}})$ be a $2d$-tensor. Fix $I = \{k_1,\dots,k_l\} \subseteq [d]$ with $k_1 < \cdots < k_l$, and fix $\varepsilon \in \{0,1\}^l$. Let $T_{I,\varepsilon}A$ be defined by* (B.1)*, and let $\rho_{I,\varepsilon}$ be defined by* (B.2)*.*

Let $I_2 \subseteq [2l]$ and set

$$F := \rho_{I,\varepsilon}(I_2) \subseteq [2d].$$

For a fixed j_{I_2}, denote by j_F^ρ the corresponding fixed coordinates of A, namely

$$i_{\rho_{I,\varepsilon}(s)} = j_s, \qquad s \in I_2.$$

If $\mathcal{J} = \{J_1,\dots,J_\kappa\} \in \Pi([2l] \setminus I_2)$, define

$$\mathcal{K} := \rho_{I,\varepsilon}(\mathcal{J}) \cup \big\{\{t, d+t\} : t \in [d] \setminus I\big\}.$$

Then $\mathcal{K}$ is a partition of $[2d] \setminus F$, and

$$\big\|(T_{I,\varepsilon}A)_{j_{I_2^c}}\big\|_{\mathcal{J}} \le n^{(d-l)/2} \big\|A_{(j_F^\rho)^c}\big\|_{\mathcal{K}}. \tag{B.8}$$

In particular, if $I_2 = \varnothing$, then

$$\|T_{I,\varepsilon}A\|_{\mathcal{J}} \le n^{(d-l)/2} \|A\|_{\rho_{I,\varepsilon}(\mathcal{J}),\, \{\{t,d+t\}: t\in[d]\setminus I\}}.$$

Proof. It suffices to prove (B.8). Let $x^{(s)} = (x^{(s)}_{j_{J_s}})$, $s = 1,\dots,\kappa$, be test tensors with Hilbert–Schmidt norm at most one. By the definition of $T_{I,\varepsilon}A$,

$$\begin{aligned}\sum_{j_{[2l]\setminus I_2}} (T_{I,\varepsilon}A)(j) \prod_{s=1}^{\kappa} x^{(s)}_{j_{J_s}} \\ = \sum_{i_{[2d]\setminus F}} A_i \prod_{s=1}^{\kappa} x^{(s)}_{i_{\rho_{I,\varepsilon}(J_s)}} \prod_{t\in[d]\setminus I} \mathbf{1}_{\{i_t = i_{d+t}\}},\end{aligned}$$

where the coordinates in F are fixed according to j_F^ρ. For every $t \in [d] \setminus I$, define

$$y^{(t)}_{a,b} := n^{-1/2} \mathbf{1}_{\{a=b\}}, \qquad a, b \in [n].$$

Then $\|y^{(t)}\|_2 = 1$. The last display equals

$$n^{(d-l)/2} \sum_{i_{[2d]\setminus F}} A_i \prod_{s=1}^{\kappa} x^{(s)}_{i_{\rho_{I,\varepsilon}(J_s)}} \prod_{t\in[d]\setminus I} y^{(t)}_{i_t, i_{d+t}}.$$

This is bounded by $n^{(d-l)/2}\|A_{(j_F^\rho)^c}\|_{\mathcal{K}}$ by the definition of the partition norm. Taking the supremum over all admissible test tensors $x^{(s)}$ proves (B.8). $\square$

Lemma B.5 (Uniform bounds for the unweighted partial traces)**.** *Let $I \subseteq [d]$, $|I| = l$, let $\varepsilon \in \{0,1\}^l$, and let $T_{I,\varepsilon}A$ be defined by* (B.1)*. Let $I_2 \subseteq [2l]$, set $q := |I_2|$, and let $\mathcal{J} \in \Pi([2l] \setminus I_2)$ with $\kappa := |\mathcal{J}|$. Then, uniformly in the fixed coordinates j_{I_2},*

$$\big\|(T_{I,\varepsilon}A)_{j_{I_2^c}}\big\|_{\mathcal{J}} \le n^{d-(\kappa+q)/2} \|A\|_{\mathsf{L},\mathsf{R}}. \tag{B.9}$$

Moreover, when $I_2 = \varnothing$,

$$\|T_{I,\varepsilon}A\|_{\mathcal{J}} \le n^{(d-l)/2} \|A\|_{[2d]}. \tag{B.10}$$

Proof. Let $\rho = \rho_{I,\varepsilon}$ and set $F := \rho(I_2)$. Let

$$\mathcal{K} := \rho(\mathcal{J}) \cup \big\{\{t, d+t\} : t \in [d] \setminus I\big\}.$$

Then $\mathcal{K}$ is a partition of F^c and

$$|\mathcal{K}| = \kappa + d - l.$$

By Lemma B.4,

$$\big\|(T_{I,\varepsilon}A)_{j_{I_2^c}}\big\|_{\mathcal{J}} \le n^{(d-l)/2} \|A_{(j_F^\rho)^c}\|_{\mathcal{K}}. \tag{B.11}$$

Applying Lemma B.3 with $m = |\mathcal{K}| = \kappa + d - l$ and $q = |F| = |I_2|$ gives

$$\|A_{(j_F^\rho)^c}\|_{\mathcal{K}} \le n^{d-(\kappa+d-l+q)/2} \|A\|_{\mathsf{L},\mathsf{R}}.$$

Multiplying by $n^{(d-l)/2}$ proves (B.9).

For (B.10), take $I_2 = \varnothing$ in (B.11). Then $F = \varnothing$ and $\mathcal{K}$ is a partition of $[2d]$. Since every partition norm is bounded by the Hilbert–Schmidt norm,

$$\|A\|_{\mathcal{K}} \le \|A\|_{[2d]},$$

and (B.10) follows. $\square$

Acknowledgment K. Wang was partially supported by the Hong Kong RGC grant GRF 16304222. The authors would like to thank Radosław Adamczak, Yiyun He and Yizhe Zhu for inspiring and valuable discussions.

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*, 26:1–36, 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] 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.

[6] C. Borell. The Brunn-Minkowski inequality in Gauss space. *Inventiones mathematicae*, 30(2):207–216, 1975.

[7] 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.

[8] P. Buterus and H. Sambale. Some notes on moment inequalities for heavy-tailed distributions. *arXiv preprint arXiv:2308.14410*, 2023.

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

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

[11] 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.

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

[13] G. Dai, Z. Su, V. Ulyanov, and H. Wang. On log-concave-tailed chaoses and the restricted isometry property. *Journal of Functional Analysis*, 289:111130, 2025.

[14] E. Giné, R. Latała, and J. Zinn. Exponential and moment inequalities for U-statistics. In *High Dimensional Probability II*, pages 13–38. Springer, 2000.

[15] E. 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.

[16] 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.

[17] F. Götze and A. Tikhomirov. On the largest and the smallest singular value of sparse rectangular random matrices. *Electronic Journal of Probability*, 28(27):18 pp, 2023.

[18] 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.

[19] Y. He, A. Knowles, and M. Marcozzi. Local law and complete eigenvector delocalization for supercritical Erdős–Rényi graphs. *The Annals of Probability*, 47(5):3278–3302, 2019.

[20] Y. He, K. Wang, and Y. Zhu. Sparse Hanson-Wright inequalities with applications. *Electronic Journal of Probability*, 31:1–49, 2026.

[21] 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.

[22] J. H. Kim and V. H. Vu. Concentration of multivariate polynomials and its applications. *Combinatorica*, 20(3):417–434, 2000.

[23] 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.

[24] 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.

[25] S. Kwapien. Decoupling inequalities for polynomial chaos. *The Annals of Probability*, 15(3):1062–1071, 1987.

[26] R. Latała. Estimation of moments of sums of independent real random variables. *The Annals of Probability*, 25(3):1502–1513, 1997.

[27] R. Latała. Tail and moment estimates for some types of chaos. *Studia mathematica*, 135:39–53, 1999.

[28] R. Latała. Estimation of moments and tails of gaussian chaoses. *The Annals of Probability*, 34(6):2315–2331, 2006.

[29] M. Ledoux and M. Talagrand. *Probability in Banach Space: Isoperimetry and Processes.* Springer, 1991.

[30] C. Louart and R. Couillet. Concentration of measure and generalized product of random vectors with an application to Hanson-Wright-like inequalities. *arXiv preprint arXiv:2102.08020*, 2021.

[31] E. Lubetzky and Y. Zhao. On the variational problem for upper tails in sparse random graphs. *Random Structures & Algorithms*, 50(3):420–436, 2017.

[32] S. Montgomery-Smith. The distribution of Rademacher sums. *Proceedings of the American Mathematical Society*, 109(2):517–522, 1990.

[33] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics*, LXII:1707–1739, 2009.

[34] M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 82(1-9), 2013.

[35] 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.

[36] 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.

[37] R. Vershynin. Concentration inequalities for random tensors. *Bernoulli*, 26(4):3139–3162, 2020.

[38] V. Vu and K. Wang. Random weighted projections, random quadratic forms and random eigenvectors. *Random Structures & Algorithms*, 47(4):792–821, 2015.

[39] S. Zhou. Sparse Hanson–Wright inequalities for subgaussian quadratic forms. *Bernoulli*, 25(3):1603–1639, 2019.

[40] S. Zhou. Concentration of measure bounds for matrix-variate data with missing values. *Bernoulli*, 30(1):198–226, 2024.