# SHARP INJECTIVE NORM BOUNDS FOR INHOMOGENEOUS GAUSSIAN TENSORS AND SPARSE EXTENSIONS

GUOZHENG DAI AND KE WANG

Abstract. We study the Euclidean injective norm of inhomogeneous random tensors with independent entries. For Gaussian tensors, we establish optimal two-sided bounds for the expected norm, with constants depending only on the tensor order. These bounds describe the interplay between variance accumulation along fibres and the contribution of the largest individual entries. In the matrix case, our result is consistent with the characterizations in [14]. We then extend the analysis to sparse Gaussian tensors with arbitrary amplitudes and activation probabilities. The expected norm is characterized by the expected largest absolute entry and the largest fibre standard deviation after truncating amplitudes at this scale. Our proofs combine geometric chaining with estimates for averaging over the random support.

## 1. Introduction and main results

Fix an integer $d \geq 2$ and positive integers $n_1, \dots, n_d$. Write $[n] = \{1, \dots, n\}$ and $\mathcal{I} = [n_1] \times \cdots \times [n_d]$. For a real tensor $U = (U_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathcal{I}}$, its *Euclidean injective norm* is defined by

$$\|U\|_{\mathrm{inj}} = \sup_{\|x^{(r)}\|_2 \leq 1,\, 1 \leq r \leq d} \left| \sum_{\boldsymbol{i} \in \mathcal{I}} U_{\boldsymbol{i}} \prod_{r=1}^{d} x^{(r)}_{i_r} \right|.$$

Here $x^{(r)} \in \mathbb{R}^{n_r}$ for each $r$. Thus $\|U\|_{\mathrm{inj}}$ is the supremum of the absolute value of the associated multilinear form over a product of Euclidean unit balls. When $d = 2$, it coincides with the usual matrix operator norm.

Estimating the injective norm of random tensors is important both in theory and in applications. This problem extends spectral-norm estimation for random matrices to higher-order random arrays [5] and also arises in the study of random hypergraphs [23], tensor principal component analysis with random noise [18], and tensor completion from randomly sampled entries [13], among other applications. These connections motivate the search for sharp bounds that capture how the distributions of the entries and their arrangement affect the expected norm. In this paper, we seek sharp two-sided bounds for the expected injective norm of random tensors with independent entries.

Tomioka and Suzuki [20] obtained injective-norm upper bounds of order $K\sqrt{(\sum_{r=1}^{d} n_r)\log(2d)}$ for $n_1 \times \cdots \times n_d$ tensors with independent centered sub-Gaussian entries having a common sub-Gaussian parameter $K$. For tensors with i.i.d. real or complex Gaussian entries, Dartois and McKenna [10] used the Kac–Rice formula to obtain high-probability upper bounds with refined leading constants. For tensors with independent Bernoulli entries, Zhou and Zhu [23, 24] developed concentration and regularization estimates allowing unequal success probabilities and deterministic weights. Lucca and Pesenti [17] established dimension-independent bounds for independent entries with sufficiently small probabilities of being nonzero, as well as Seginer-type two-sided estimates for i.i.d. centered entries under an equivalence assumption on their second and $(2+\varepsilon)$-th moments. More recently, Lin and Lu [15] established a Seginer-type fibre comparison for arbitrary centered integrable i.i.d. tensor entries.

2020 *Mathematics Subject Classification.* Primary 60B20, 60G15; Secondary 60G17, 46B09.

*Key words and phrases.* Generic chaining, inhomogeneous Gaussian tensors, injective norm, sparsity.

Another line of work studies random tensor series with independent Gaussian or, more generally, sub-Gaussian coefficients [2, 1]. These results allow correlations between entries and also yield upper bounds for the independent-entry model by taking the deterministic summands to be coordinate tensors. Specifically, for a deterministic real coefficient tensor $A = (a_{\boldsymbol{i}})_{\boldsymbol{i}\in\mathcal{I}}$, write

$$U = (a_{\boldsymbol{i}} g_{\boldsymbol{i}})_{\boldsymbol{i}\in\mathcal{I}} = \sum_{\boldsymbol{i}\in\mathcal{I}} g_{\boldsymbol{i}} (a_{\boldsymbol{i}} e_{i_1} \otimes \cdots \otimes e_{i_d}),$$

where the $g_{\boldsymbol{i}}$ are independent standard Gaussian variables. Let $\mathcal{F}_r$ denote the family of mode-$r$ fibres of $\mathcal{I}$, obtained by fixing all coordinates except $i_r$, and set

$$\sigma_r(A) = \max_{f\in\mathcal{F}_r} \left( \sum_{\boldsymbol{i}\in f} a_{\boldsymbol{i}}^2 \right)^{1/2}, \qquad a_* = \max_{\boldsymbol{i}\in\mathcal{I}} |a_{\boldsymbol{i}}|.$$

After a symmetric embedding, the independent-entry estimate of Bandeira et al. [2] gives

$$\begin{aligned} \mathbb{E}\,\|U\|_{\mathrm{inj}} \lesssim_d \left( \max_{1\le r\le d} \sigma_r(A) \right) \log\left( 2 + \sum_{r=1}^{d} n_r \right) \\ + \max_{\substack{J\subseteq[d]\\ |J|\ge 2}} a_*^{1-1/|J|} \max_{\boldsymbol{i}_{J^c}} \left( \sum_{\boldsymbol{i}_J} a_{\boldsymbol{i}}^2 \right)^{1/(2|J|)}, \end{aligned} \tag{1.1}$$

where $\boldsymbol{i}_J = (i_r)_{r\in J}$, the sums and maxima run over the corresponding coordinate ranges, and the implicit constant depends only on $d$. Thus, this specialization involves both fibre variances and variance sums over higher-dimensional slices. Aden-Ali [1] removed the logarithmic multiplier on the fibre term, while retaining the higher-dimensional slice terms.

Within the more restrictive independent-entry Gaussian model, Boedihardjo [5] obtained a bound that can be sharper than the direct specializations of the general tensor-series estimates of Bandeira et al. [2] and Aden-Ali [1]. His estimate involves only the fibre variances and the largest coefficient. When all mode dimensions are equal, $n_1 = \cdots = n_d = n \ge 2$, [5, Theorem 1.1] gives

$$\mathbb{E}\,\|U\|_{\mathrm{inj}} \le \sqrt{2d} \sum_{r=1}^{d} \sigma_r(A) + C d^3 (\log n)^2 a_*,$$

where $C$ is a universal constant. Here $\sigma_r(A)^2$ is the largest sum of entry variances along a mode-$r$ fibre. Thus, the leading term reflects how variances accumulate along fibres, while the correction term depends on the largest coefficient. This estimate also yields a bound of the form predicted by the Euclidean type-2 conjecture discussed by Lucca [16], with a polylogarithmic loss, for Gaussian tensor series with independent entries.

The gain over the direct series estimates, and the remaining gap to optimality, can both be seen in a diagonal example. If $A$ has unit coefficients on the diagonal $i_1 = \cdots = i_d$ and is zero elsewhere, then $\sigma_r(A) = a_* = 1$ for every $r$. For fixed $d$, the direct specializations of the bounds in [2, 1] contain a term of order $n^{1/(2d)}$, whereas Boedihardjo's upper bound is of order $(\log n)^2$. Nevertheless, a logarithmic gap remains in this example, since

$$\mathbb{E}\,\|U\|_{\mathrm{inj}} = \mathbb{E} \max_{1\le i\le n} |g_{i,\dots,i}| \asymp \sqrt{\log n}.$$

Thus, the correction $a_*(\log n)^2$ still overestimates the actual scale of entrywise extremes. This discrepancy motivates a finer characterization that captures both variance accumulation along fibres and the actual scale of entrywise extremes.

1.1. **Main results.** Our first main result gives a sharp two-sided characterization for the independent-entry Gaussian model. It retains the fibre scale and replaces the correction based on the largest coefficient by an entrywise extreme term determined by the decreasing

rearrangement of the amplitudes. We then extend this characterization to sparse Gaussian tensors with arbitrary amplitude and activation-probability profiles.

We return to the rectangular index set $\mathcal{I} = [n_1] \times \cdots \times [n_d]$, with $d \geq 2$ fixed. Recall that $\sigma_r(A)$ denotes the largest Euclidean norm of a mode-$r$ fibre of $A$. With $\mathcal{F} = \bigcup_{r=1}^{d} \mathcal{F}_r$, set

$$\sigma(A) = \max_{1 \leq r \leq d} \sigma_r(A) = \max_{f \in \mathcal{F}} \left( \sum_{\boldsymbol{i} \in f} a_{\boldsymbol{i}}^2 \right)^{1/2}.$$

We write $x \wedge y = \min\{x, y\}$. We write $X \lesssim_d Y$ if $X \leq C_d Y$ for a positive constant depending only on $d$, and $X \gtrsim_d Y$ if $Y \lesssim_d X$. We write $X \asymp_d Y$ when both inequalities hold. Omitted subscripts indicate absolute constants.

**Theorem 1.1** (Gaussian tensors with deterministic coefficients)**.** *Let $A = (a_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathcal{I}}$ be deterministic with $a_{\boldsymbol{i}} \geq 0$, and let $(g_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathcal{I}}$ be independent standard Gaussian variables. Let $a_1^* \geq \cdots \geq a_{|\mathcal{I}|}^*$ be the decreasing rearrangement of the entries of $A$. Then*

$$\mathbb{E} \, \|(a_{\boldsymbol{i}} g_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathcal{I}}\|_{\mathrm{inj}} \asymp_d \sigma(A) + \max_{1 \leq k \leq |\mathcal{I}|} a_k^* \sqrt{\log(k+1)}. \tag{1.2}$$

The restriction to nonnegative amplitudes entails no loss for Gaussian tensors: deterministic coefficient signs can be absorbed into the independent Gaussian variables.

*Remark* 1.2 (Comparison with earlier results). For each fixed order $d$, Theorem 1.1 gives a sharp two-sided refinement of the Gaussian estimate of Boedihardjo [5]. The fibre terms agree up to constants depending only on $d$, since

$$\sigma(A) \leq \sum_{r=1}^{d} \sigma_r(A) \leq d \, \sigma(A).$$

In the case $n_1 = \cdots = n_d = n \geq 2$, the global correction $a_1^* (\log n)^2$ is replaced by

$$\max_{1 \leq k \leq n^d} a_k^* \sqrt{\log(k+1)} \leq a_1^* \sqrt{d \log(n+1)}.$$

Thus, for fixed $d$, the worst-case entrywise correction improves from order $a_1^* (\log n)^2$ to order $a_1^* \sqrt{\log n}$, while retaining the full decreasing amplitude profile can yield further improvements.

The comparison constants in (1.2) depend on $d$, and we do not optimize this dependence. Boedihardjo's estimate, on the other hand, gives explicit dependence on the tensor order through the factors $\sqrt{2d}$ and $d^3$. Our refinement therefore concerns sharpness in the dimensions and coefficient profiles for fixed $d$.

When $d = 2$, Theorem 1.1 recovers, in an equivalent deterministic form, the sharp operator-norm characterization of Gaussian matrices due to Latała, van Handel, and Youssef [14, Corollary 4.1, with $p = \infty$].

We next consider sparse Gaussian tensors, motivated by models in which only a random subset of entries is observed or retained. Such random support structures arise in covariance estimation with incomplete observations, tensor completion and sparsification, randomized embeddings, and random hypergraphs, among other applications. For further background on these applications and related concentration problems, see [13, 23, 12, 7, 9, 8, 17] and the references therein. Allowing both the activation probabilities and the amplitudes to vary across the array accommodates nonuniform sparsity and heterogeneous entry magnitudes.

Let $A = (a_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathcal{I}}$ and $P = (p_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathcal{I}}$ be deterministic profiles with $a_{\boldsymbol{i}} \geq 0$ and $p_{\boldsymbol{i}} \in [0, 1]$. Let $\delta_{\boldsymbol{i}} \sim \mathrm{Bernoulli}(p_{\boldsymbol{i}})$ and $g_{\boldsymbol{i}} \sim N(0, 1)$, with all these variables mutually independent, and consider

$$U = (a_{\boldsymbol{i}} \delta_{\boldsymbol{i}} g_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathcal{I}}. \tag{1.3}$$

The amplitudes and activation probabilities may vary arbitrarily across the array.

The sparse characterization involves two deterministic quantities. The first is the expected largest absolute entry of the sparse Gaussian tensor in (1.3):

$$\mathfrak{e}(A,P)=\mathbb{E}_{\delta,g}\max_{\boldsymbol{i}\in\mathcal{I}}a_{\boldsymbol{i}}\delta_{\boldsymbol{i}}|g_{\boldsymbol{i}}|=\mathbb{E}\max_{\boldsymbol{i}\in\mathcal{I}}|U_{\boldsymbol{i}}|. \tag{1.4}$$

The second quantity measures variance accumulation along fibres. For $s\geq 0$, define

$$v_s(A,P)=\max_{f\in\mathcal{F}}\left(\sum_{\boldsymbol{i}\in f}p_{\boldsymbol{i}}(a_{\boldsymbol{i}}\wedge s)^2\right)^{1/2},\qquad \mathfrak{v}(A,P)=v_{\mathfrak{e}(A,P)}(A,P), \tag{1.5}$$

and put

$$\Psi_{\mathrm{sp}}(A,P)=\mathfrak{e}(A,P)+\mathfrak{v}(A,P). \tag{1.6}$$

**Theorem 1.3** (Sparse Gaussian tensors)**.** *For every choice of dimensions and deterministic profiles $A,P$ as above, the tensor $U$ defined in* (1.3) *satisfies*

$$\mathbb{E}\,\|U\|_{\mathrm{inj}}\asymp_d\Psi_{\mathrm{sp}}(A,P). \tag{1.7}$$

*Remark* 1.4 (Sub-Gaussian amplitudes). The upper bound in Theorem 1.3 remains valid when the Gaussian amplitudes are replaced by independent centered sub-Gaussian variables. More precisely, let $(\xi_{\boldsymbol{i}})_{\boldsymbol{i}\in\mathcal{I}}$ be independent of one another and of the Bernoulli masks, and suppose that

$$\mathbb{E}e^{t\xi_{\boldsymbol{i}}}\leq e^{K^2t^2/2},\qquad t\in\mathbb{R},\quad \boldsymbol{i}\in\mathcal{I},$$

for some $K>0$. Then the tensor $U^{(\xi)}=(a_{\boldsymbol{i}}\delta_{\boldsymbol{i}}\xi_{\boldsymbol{i}})_{\boldsymbol{i}\in\mathcal{I}}$ satisfies

$$\mathbb{E}\left\|U^{(\xi)}\right\|_{\mathrm{inj}}\leq C_dK\Psi_{\mathrm{sp}}(A,P).$$

Indeed, conditional on the masks, the associated multilinear process has sub-Gaussian increments controlled by $K$ times the canonical metric of the corresponding Gaussian process. The generic chaining upper bound and the Gaussian majorizing-measure theorem [19, 11] therefore give

$$\mathbb{E}_\xi\,\|(a_{\boldsymbol{i}}\delta_{\boldsymbol{i}}\xi_{\boldsymbol{i}})_{\boldsymbol{i}\in\mathcal{I}}\|_{\mathrm{inj}}\leq CK\,\mathbb{E}_g\,\|(a_{\boldsymbol{i}}\delta_{\boldsymbol{i}}g_{\boldsymbol{i}})_{\boldsymbol{i}\in\mathcal{I}}\|_{\mathrm{inj}}\,,$$

where $C$ is an absolute constant. Averaging over the masks and applying Theorem 1.3 yields the asserted upper bound.

Bounding the extreme scale in terms of the largest amplitude and expected support size gives the following simpler consequence of Theorem 1.3.

**Corollary 1.5** (A variance-plus-maximum upper bound)**.** *For the sparse Gaussian tensor $U$ in* (1.3)*, set*

$$V(A,P)=\max_{f\in\mathcal{F}}\Big(\sum_{\boldsymbol{i}\in f}p_{\boldsymbol{i}}a_{\boldsymbol{i}}^2\Big)^{1/2},\qquad a_*=\max_{\boldsymbol{i}\in\mathcal{I}}a_{\boldsymbol{i}},\qquad \Lambda=\sum_{\substack{\boldsymbol{i}\in\mathcal{I}\\ a_{\boldsymbol{i}}>0}}p_{\boldsymbol{i}}.$$

*Then*

$$\mathbb{E}\,\|U\|_{\mathrm{inj}}\lesssim_d V(A,P)+a_*\sqrt{\log(1+\Lambda)}. \tag{1.8}$$

Here $V(A,P)^2$ is the largest total variance along a fibre, and $\Lambda$ is the expected number of nonzero entries. The bound avoids truncation and the ordered amplitude profile, at the possible cost of sharpness. For an $n\times n$ Gaussian matrix with $P\equiv 1$, we have $\Lambda\leq n^2$, so the bound becomes $V(A,\mathbf{1})+a_*\sqrt{\log(n+1)}$ up to universal constants. This recovers the form of the Gaussian matrix bound of Bandeira and van Handel [3].

Theorem 1.3 also yields an upper bound for centered Bernoulli tensors through symmetrization and Gaussian comparison.

**Corollary 1.6** (Bernoulli tensors)**.** *Let $A = (a_i)_{i\in\mathcal{I}}$ be deterministic with $a_i \geq 0$, and let $\delta_i \sim \mathrm{Bernoulli}(p_i)$ be independent, where $p_i \in [0,1]$. Define*

$$U = (a_i\delta_i)_{i\in\mathcal{I}}, \qquad q_i = 2p_i(1-p_i), \qquad Q = (q_i)_{i\in\mathcal{I}}.$$

*Then*

$$\mathbb{E}\,\|U - \mathbb{E}U\|_{\mathrm{inj}} \leq C_d \Psi_{\mathrm{sp}}(A,Q), \tag{1.9}$$

*where $C_d$ depends only on $d$. Here $\Psi_{\mathrm{sp}}(A,Q)$ is defined by* (1.6) *with $P$ replaced by $Q$.*

*Remark* 1.7 (Worst-case loss of the Bernoulli bound). The bound in Corollary 1.6 can lose a factor of order $\sqrt{\log n}$ on $[n]^d$, and this is the worst possible loss for fixed $d$. More generally, with $N = |\mathcal{I}|$,

$$\Psi_{\mathrm{sp}}(A,Q) \lesssim_d \sqrt{\log(N+1)}\,\mathbb{E}\,\|U - \mathbb{E}U\|_{\mathrm{inj}}\,.$$

This follows from Theorem 1.3 and the Rademacher contraction principle, which bounds the cost of replacing signs by Gaussian variables by $\mathbb{E}\max_{i\in\mathcal{I}}|g_i| \lesssim \sqrt{\log(N+1)}$. Symmetrization changes the expected norm by at most a factor of two. On $[n]^d$, we have $N = n^d$.

To see that this order is attained, let $A$ have unit coefficients on a diagonal of length $n \geq 2$ and vanish elsewhere, with activation probability $1/2$ on the diagonal. Then

$$\|U - \mathbb{E}U\|_{\mathrm{inj}} = \frac{1}{2} \quad \text{almost surely}, \qquad \Psi_{\mathrm{sp}}(A,Q) \asymp \sqrt{\log n}.$$

Thus the factor $\sqrt{\log n}$ cannot be improved uniformly over all profiles for fixed tensor order.

Although the Bernoulli bound is not uniformly sharp, it can improve existing estimates for inhomogeneous amplitudes. Zhou and Zhu [24, Theorem 2.2] obtain optimal expectation bounds at every sparsity level for homogeneous tensors with $A \equiv 1$ and $p_i \equiv p \in (0, 1/2]$ on $[n]^d$. Their weighted extension also allows rectangular tensors and unequal activation probabilities at every sparsity level [24, Corollary 2.3]. To compare the expectation bounds in a common regime, consider tensors on $[n]^d$ with $p_i \equiv p \in (0,1/2]$ and $np \geq \log n$. Their expectation upper bound then has scale $a_*\sqrt{np}$, whereas

$$\Psi_{\mathrm{sp}}(A,Q) \lesssim_d V(A,Q) + a_*\sqrt{\log n} \lesssim_d a_*\sqrt{np}.$$

Our bound therefore improves this order whenever $\log n = o(np)$ and $V(A,Q) = o(a_*\sqrt{np})$.

1.2. **Proof strategy.** Our argument separates the geometry of the tensor norm from the averaging of the random support. We first prove Theorem 1.1 for deterministic coefficients, and then evaluate the random coefficient profiles obtained by conditioning on the Bernoulli masks.

For the deterministic-coefficient estimate, we use the contraction principle and uniform convexity estimates of van Handel [22], and adapt the geometric-midpoint localization scheme proposed in the candidate proof of Bednorz, Martynek, and Meller [4, Section 6]. The contribution here is to carry out this localization for a multilinear parametrization of arbitrary fixed tensor order while retaining the fibre variances and the ordered entrywise profile. We use one fibre seminorm for each mode to control the variation of the derivative. Their Gaussian averages are bounded by the two terms in Theorem 1.1, and the Hilbert–Schmidt norm of the derivative is controlled by $\sigma(A)$. These estimates allow square chaining in an auxiliary pseudometric to yield the required chaining bound in the canonical pseudometric. The transfer estimate is proved in full in Subsection 2.4.

For the sparse model, the key contribution is to identify a common scale that controls both the truncation error and the fluctuations of the truncated fibre sums. The extreme parameter $\mathfrak{e}(A,P)$ provides this scale: amplitudes above a constant multiple of it have a controlled total first-moment contribution, while truncation gives a second-moment bound for the random entrywise profile. Combined with symmetrization and an ordered maximum estimate, these bounds compare $\mathbb{E}_\delta\sigma(A\odot\delta)$ with $\mathfrak{v}(A,P)$ up to the additive term $\mathfrak{e}(A,P)$, as made precise in Proposition 3.5. Here $\odot$ denotes entrywise multiplication. The resulting

comparison removes the remaining expectation over the masks and explains the truncation appearing in the deterministic characterization.

## 2. Gaussian tensors with deterministic coefficients

This section proves Theorem 1.1. We retain the fibre parameters $\sigma_r(A)$ and $\sigma(A)$ introduced in Section 1. To simplify the notation in the proof, for the deterministic nonnegative coefficient tensor $A = (a_{\boldsymbol{i}})_{\boldsymbol{i}\in\mathcal{I}}$ we set

$$\beta(A) = \max_{1\le j\le|\mathcal{I}|} a_j^* \sqrt{\log(j+1)},$$

where $(a_j^*)$ is the decreasing rearrangement of the coefficients of $A$. Our goal is to prove

$$\mathbb{E}\,\|(a_{\boldsymbol{i}} g_{\boldsymbol{i}})\|_{\mathrm{inj}} \asymp_d \sigma(A) + \beta(A). \tag{2.1}$$

The lower bound is straightforward and follows by testing fibres and individual entries. The main difficulty lies in the upper bound. We first establish the profile and chaining estimates, then study the geometry of the multilinear parametrization and prove the required localization bound. The final subsection completes the proof of Theorem 1.1, including the lower bound. Throughout the proof, we assume $A \neq 0$, since the zero tensor is immediate.

### 2.1. Gaussian profiles and fibre seminorms.

The following classical estimate is recorded in [21, Lemma 4.5 and Remark 4.6]. For completeness and the reader's convenience, we include a short proof.

**Lemma 2.1** (Ordered Gaussian maximum)**.** *Let $b_1, \dots, b_M$ be nonnegative, and let $b_1^* \ge \cdots \ge b_M^*$ denote their decreasing rearrangement. For independent standard Gaussian variables $g_1, \dots, g_M$,*

$$\mathbb{E}\max_{j\le M} b_j|g_j| \asymp \max_{j\le M} b_j^* \sqrt{\log(j+1)}.$$

*Proof.* The claim is immediate if all coefficients vanish. By relabelling, we may assume $b_1 \ge \cdots \ge b_M$. The Gaussian tail bound and a union bound give, for $u \ge 2$,

$$\mathbb{P}\left\{\max_{j\le M} b_j|g_j| > u \max_{k\le M} b_k\sqrt{\log(k+1)}\right\} \le 2\sum_{j=1}^{M}(j+1)^{-u^2/2} \lesssim 2^{-u^2/2}.$$

Integration proves the upper bound. For the lower bound, for every $1 \le k \le M$,

$$\mathbb{E}\max_{j\le M} b_j|g_j| \ge b_k\,\mathbb{E}\max_{j\le k}|g_j| \gtrsim b_k\sqrt{\log(k+1)}.$$

The last inequality follows from independence and the Gaussian lower tail

$$\mathbb{P}\{|g| > t\} \gtrsim (1+t)^{-1}e^{-t^2/2}, \qquad t \ge 0.$$

Taking the maximum over $k$ completes the proof. □

Recall that $\mathcal{F}_r$ consists of the mode-$r$ fibres, obtained by fixing all coordinates except the $r$th. Define

$$N_r(h) = \max_{f\in\mathcal{F}_r}\left(\sum_{\boldsymbol{i}\in f} a_{\boldsymbol{i}}^2 h_{i_r}^2\right)^{1/2}, \qquad h \in \mathbb{R}^{n_r}. \tag{2.2}$$

These seminorms are unconditional: if $|h_i| \le |k_i|$ for every coordinate, then $N_r(h) \le N_r(k)$.

**Lemma 2.2** (Gaussian averages of the fibre seminorms)**.** *For a standard Gaussian vector $g^{(r)} \in \mathbb{R}^{n_r}$,*

$$\mathbb{E}N_r(g^{(r)}) \lesssim \sigma_r(A) + \beta(A).$$

*Proof.* Fibres on which all coefficients vanish may be omitted. For each remaining $f \in \mathcal{F}_r$, Jensen's inequality gives

$$\mathbb{E}\left(\sum_{\boldsymbol{i}\in f} a_{\boldsymbol{i}}^2 (g_{i_r}^{(r)})^2\right)^{1/2} \le \left(\sum_{\boldsymbol{i}\in f} a_{\boldsymbol{i}}^2\right)^{1/2} \le \sigma_r(A).$$

Moreover, the map

$$h \longmapsto \left(\sum_{\boldsymbol{i}\in f} a_{\boldsymbol{i}}^2 h_{i_r}^2\right)^{1/2}$$

is $\max_{\boldsymbol{i}\in f} a_{\boldsymbol{i}}$-Lipschitz in the Euclidean norm. Since the fibres in $\mathcal{F}_r$ are disjoint, their maximal coefficients occupy distinct positions in $A$. Their decreasing rearrangement is therefore bounded termwise by $(a_j^*)$. Gaussian concentration [6, Theorem 5.6] and a union bound consequently give, for $u \ge 2$,

$$\mathbb{P}\{N_r(g^{(r)}) > \sigma_r(A) + u\beta(A)\} \le \sum_{j\ge 1} (j+1)^{-u^2/2} \lesssim 2^{-u^2/2}.$$

Integration proves the claim. No independence between the fibre norms is required. □

2.2. **Chaining and entropy estimates.** We follow the notation for entropy numbers and chaining functionals in [22, Section 2]; see also [19] for background. For a nonempty subset $U$ of a metric space $(X, d_0)$, define

$$\operatorname{diam}_{d_0}(U) = \sup_{x,y\in U} d_0(x,y)$$

and, for each integer $j \ge 0$,

$$e_j(U, d_0) = \inf_{\substack{F\subseteq X \\ 1\le|F|<2^{2^j}}} \sup_{x\in U} \inf_{y\in F} d_0(x,y).$$

Thus $e_j(U, d_0)$ is the infimum of the radii of covers of $U$ by fewer than $2^{2^j}$ balls. The centres may lie in $X$; requiring them to lie in $U$ changes this quantity by at most a factor of two.

Let $S \subseteq X$ be nonempty. A finite partition $\mathcal{A}_j$ of $S$ is a finite collection of pairwise disjoint nonempty subsets whose union is $S$. Write $\mathcal{A}_j(x)$ for its unique member containing $x$. A sequence $(\mathcal{A}_j)_{j\ge 0}$ is *admissible* if

$$|\mathcal{A}_j| < 2^{2^j}, \qquad \mathcal{A}_{j+1}(x) \subseteq \mathcal{A}_j(x) \qquad (j \ge 0,\ x \in S).$$

In particular, $\mathcal{A}_0 = \{S\}$. For $1 \le p < \infty$, define

$$\gamma_{2,p}(S, d_0) = \inf_{\substack{(\mathcal{A}_j)_{j\ge 0} \\ \text{admissible}}} \sup_{x\in S} \left(\sum_{j\ge 0} 2^{jp/2} \operatorname{diam}_{d_0}(\mathcal{A}_j(x))^p\right)^{1/p},$$

where the infimum runs over all admissible sequences of partitions of $S$. We write $\gamma_2 = \gamma_{2,1}$ for the usual chaining functional; $\gamma_{2,2}$ uses a square sum over the levels. These definitions also apply to pseudometrics. When applying metric-space results, we choose one representative from each zero-distance class.

We use the notation

$$B_2^m = \{x \in \mathbb{R}^m : \|x\|_2 \le 1\}, \qquad B_{2,+}^m = B_2^m \cap [0,\infty)^m.$$

For linear maps between Euclidean spaces, $\|\cdot\|_{\mathrm{HS}}$ and $\|\cdot\|_{\mathrm{op}}$ denote the Hilbert–Schmidt and operator norms. When a seminorm $N$ appears in place of a distance in $e_j$ or $\gamma_{2,p}$, the associated pseudometric is $N(x-y)$.

We will use the following four estimates. The first controls stochastic suprema, the second converts local entropy estimates into a chaining bound, and the last two concern Euclidean balls and their linear images.

**Proposition 2.3** (Standard chaining estimates)**.** *The following estimates hold with absolute implicit constants.*

*(i) Let $(Z_x)_{x\in S}$ be a separable centred process with $Z_{x_0}=0$ almost surely for some $x_0\in S$. If*

$$\mathbb{E}\exp(t(Z_x-Z_y))\le\exp\left(\frac{t^2d_0(x,y)^2}{2}\right),\qquad t\in\mathbb{R},\quad x,y\in S,$$

*then*

$$\mathbb{E}\sup_{x\in S}|Z_x|\lesssim\gamma_2(S,d_0).$$

*(ii) Assume $\gamma_2(S,d_0)<\infty$, and let $\theta>0$ be a sufficiently small absolute constant. If nonnegative functions $s_j:S\to[0,\infty)$ satisfy*

$$e_j(U,d_0)\le\theta\operatorname{diam}_{d_0}(U)+\sup_{x\in U}s_j(x)$$

*for every nonempty $U\subseteq S$ and every $j\ge0$, then*

$$\gamma_2(S,d_0)\lesssim\sup_{x\in S}\sum_{j\ge0}2^{j/2}s_j(x).$$

*(iii) For every seminorm $N$ on $\mathbb{R}^m$ and a standard Gaussian vector $g\in\mathbb{R}^m$,*

$$\gamma_{2,2}(B_2^m,N)\lesssim\mathbb{E}N(g).$$

*(iv) For every linear map $L:\mathbb{R}^m\to\mathbb{R}^n$,*

$$\sum_{j\ge0}2^je_j(LB_2^m,\|\cdot\|_2)^2\lesssim\|L\|_{\mathrm{HS}}^2.$$

*Proof.* For part (i), the increment assumption allows us to apply the generic chaining upper bound [19, 11], which gives

$$\mathbb{E}\sup_{x\in S}|Z_x-Z_{x_0}|\lesssim\gamma_2(S,d_0).$$

The conclusion follows from $Z_{x_0}=0$ almost surely.

For part (ii), the contraction principle [22, Theorem 3.1], applied with $\alpha=2$ and $p=1$, yields

$$\gamma_2(S,d_0)\le C\theta\gamma_2(S,d_0)+C\sup_{x\in S}\sum_{j\ge0}2^{j/2}s_j(x)$$

for an absolute constant $C$. Choosing $\theta\le(2C)^{-1}$ and using $\gamma_2(S,d_0)<\infty$, we may absorb the first term on the right into the left-hand side to obtain

$$\gamma_2(S,d_0)\le2C\sup_{x\in S}\sum_{j\ge0}2^{j/2}s_j(x).$$

For part (iii), first suppose that $N$ is a norm. The Euclidean unit ball is 2-convex with an absolute constant, as the parallelogram identity gives

$$\left\|\frac{x+y}{2}\right\|_2\le1-\frac{1}{8}\|x-y\|_2^2,\qquad x,y\in B_2^m.$$

The uniform convexity estimate [22, Theorem 5.8], with $\alpha=q=2$, and the dual Sudakov inequality [19, Lemma 8.3.6] therefore give, respectively,

$$\gamma_{2,2}(B_2^m,N)\lesssim\sup_{j\ge0}2^{j/2}e_j(B_2^m,N)\lesssim\mathbb{E}N(g).$$

For a seminorm, apply this estimate to the norm $N+\varepsilon\|\cdot\|_2$. Monotonicity gives

$$\gamma_{2,2}(B_2^m,N)\le\gamma_{2,2}(B_2^m,N+\varepsilon\|\cdot\|_2)\lesssim\mathbb{E}N(g)+\varepsilon\mathbb{E}\|g\|_2.$$

The implicit constant is independent of $\varepsilon$, so letting $\varepsilon\downarrow0$ proves the claim.

For part (iv), the case $L=0$ is immediate. By the singular value decomposition, $LB_2^m$ is a possibly degenerate Euclidean ellipsoid whose semiaxes are the nonzero singular values of $L$.

The classical entropy estimate for Hilbert-space ellipsoids [19, Section 2.5], also used in [22, Section 7.3], bounds the weighted sum of squared entropy numbers by the sum of squared semiaxes. Since this latter sum equals $\|L\|_{\mathrm{HS}}^2$, we obtain

$$\sum_{j\geq 0} 2^j e_j(LB_2^m, \|\cdot\|_2)^2 \lesssim \|L\|_{\mathrm{HS}}^2 .$$

□

We next combine the square-chaining estimates for the mode balls on the parameter space used below. Identify $\mathbb{R}^{n_1}\times\cdots\times\mathbb{R}^{n_d}$ with $\mathbb{R}^{n_1+\cdots+n_d}$, equipped with

$$\|x\|_2^2 = \sum_{r=1}^{d} \left\|x^{(r)}\right\|_2^2, \qquad x = (x^{(1)},\dots,x^{(d)}).$$

Set

$$\mathcal{K} = B_{2,+}^{n_1}\times\cdots\times B_{2,+}^{n_d}, \qquad \rho(x,y) = \sqrt{d}\sum_{r=1}^{d} N_r(x^{(r)}-y^{(r)}).$$

Here $\rho$ is defined on the whole product space using the fibre seminorms in (2.2).

**Lemma 2.4** (Product square-chaining bound)**.** *For the preceding parameter set and pseudo-metric,*

$$\gamma_{2,2}(\mathcal{K},\rho) \lesssim_d \sigma(A) + \beta(A). \tag{2.3}$$

*Proof.* For each $r$, choose admissible partitions $(\mathcal{A}_j^{(r)})_{j\geq 0}$ of $B_2^{n_r}$, and set $s = \lceil \log_2 d\rceil + 1$. For $j \geq s$, take products of the cells of $\mathcal{A}_{j-s}^{(r)}$ restricted to $B_{2,+}^{n_r}$, omitting empty cells; for $j < s$, use the trivial partition $\{\mathcal{K}\}$. The resulting partitions $(\mathcal{A}_j)$ are nested and satisfy

$$|\mathcal{A}_j| < 2^{d2^{j-s}} \leq 2^{2^j}, \qquad j \geq s,$$

so they are admissible. For $x\in\mathcal{K}$ and $j\geq s$, their diameters satisfy

$$\operatorname{diam}_\rho(\mathcal{A}_j(x)) \leq \sqrt{d}\sum_{r=1}^{d} \operatorname{diam}_{N_r}(\mathcal{A}_{j-s}^{(r)}(x^{(r)})).$$

Minkowski's inequality therefore gives

$$\left(\sum_{j\geq s} 2^j \operatorname{diam}_\rho(\mathcal{A}_j(x))^2\right)^{1/2} \leq \sqrt{d}\,2^{s/2}\sum_{r=1}^{d}\left(\sum_{k\geq 0} 2^k \operatorname{diam}_{N_r}(\mathcal{A}_k^{(r)}(x^{(r)}))^2\right)^{1/2}.$$

The initial levels are controlled by

$$\operatorname{diam}_\rho(\mathcal{K}) \leq \sqrt{d}\sum_{r=1}^{d}\operatorname{diam}_{N_r}(B_2^{n_r}),$$

and these mode diameters already occur at level zero. Taking the supremum over $x$ and then the infima over the mode partitions yields

$$\gamma_{2,2}(\mathcal{K},\rho) \lesssim_d \sum_{r=1}^{d}\gamma_{2,2}(B_2^{n_r},N_r) \lesssim_d \sum_{r=1}^{d}\mathbb{E}N_r(g^{(r)}).$$

The last inequality follows from Proposition 2.3((iii)), and Lemma 2.2 completes the proof. □

2.3. **Geometry of the multilinear parametrization.** Define $\Phi : \mathbb{R}^{n_1+\cdots+n_d} \to \mathbb{R}^{\mathcal{I}}$ by

$$\Phi(x)_{\boldsymbol{i}} = a_{\boldsymbol{i}} \prod_{r=1}^{d} x_{i_r}^{(r)},$$

and equip $\mathcal{K}$ with the canonical pseudometric

$$D(x, y) = \|\Phi(x) - \Phi(y)\|_2$$

of the Gaussian process $\langle g, \Phi(x)\rangle$, where $g = (g_{\boldsymbol{i}})_{\boldsymbol{i}\in\mathcal{I}}$. We write $\mathrm{d}\Phi(x)$ for the derivative of $\Phi$ at $x$. The following estimates explain the role of the fibre seminorms: the fibre variances control the size of the derivative, while $\rho$ controls its variation and the error of linearization.

**Lemma 2.5** (Derivative estimates)**.** *For $x, y \in \mathcal{K}$,*

$$\|\mathrm{d}\Phi(x)\|_{\mathrm{HS}} \le \sqrt{d}\,\sigma(A), \tag{2.4}$$
$$\|\mathrm{d}\Phi(x) - \mathrm{d}\Phi(y)\|_{\mathrm{op}} \le \rho(x, y).$$

*Moreover, for $x, y, z \in \mathcal{K}$,*

$$\|\Phi(x) - \Phi(y) - \mathrm{d}\Phi(z)(x - y)\|_2 \le [\rho(x, z) + \rho(y, z)]\,\|x - y\|_2\,. \tag{2.5}$$

*Proof.* For $h = (h^{(1)}, \ldots, h^{(d)})$, differentiation gives

$$\big(\mathrm{d}\Phi(x)h\big)_{\boldsymbol{i}} = a_{\boldsymbol{i}} \sum_{r=1}^{d} h_{i_r}^{(r)} \prod_{s\neq r} x_{i_s}^{(s)}.$$

Summing the squared lengths of the columns of the derivative, and then summing over the mode-$r$ fibres, gives

$$\begin{aligned}\|\mathrm{d}\Phi(x)\|_{\mathrm{HS}}^2 &= \sum_{r=1}^{d} \sum_{\boldsymbol{i}\in\mathcal{I}} a_{\boldsymbol{i}}^2 \prod_{s\neq r} (x_{i_s}^{(s)})^2 \\ &\le \sum_{r=1}^{d} \sigma_r(A)^2 \prod_{s\neq r} \left\|x^{(s)}\right\|_2^2 \le d\sigma(A)^2.\end{aligned}$$

For the derivative difference, fix $r$ and telescope $\prod_{t\neq r} x_{i_t}^{(t)} - \prod_{t\neq r} y_{i_t}^{(t)}$ one mode at a time. In the term where mode $s \neq r$ changes, the remaining factors come from the corresponding modes of $x$ or $y$. Summing the squared coordinates first over the mode-$s$ fibres, using (2.2), bounds the norm of this term by

$$N_s(x^{(s)} - y^{(s)}) \left\|h^{(r)}\right\|_2,$$

since every remaining mode vector has Euclidean norm at most one. The triangle inequality and Cauchy–Schwarz now yield

$$\begin{aligned}\|(\mathrm{d}\Phi(x) - \mathrm{d}\Phi(y))h\|_2 &\le \sum_{r=1}^{d} \left\|h^{(r)}\right\|_2 \sum_{s\neq r} N_s(x^{(s)} - y^{(s)}) \\ &\le \sqrt{d}\,\|h\|_2 \sum_{s=1}^{d} N_s(x^{(s)} - y^{(s)}) = \rho(x, y)\,\|h\|_2\,.\end{aligned}$$

Finally, $\mathcal{K}$ is convex and

$$\Phi(x) - \Phi(y) - \mathrm{d}\Phi(z)(x - y) = \int_0^1 \big[\mathrm{d}\Phi((1 - t)y + tx) - \mathrm{d}\Phi(z)\big](x - y)\,\mathrm{d}t.$$

Since $\rho$ is induced by a seminorm,

$$\rho((1 - t)y + tx, z) \le (1 - t)\rho(y, z) + t\rho(x, z).$$

Integration proves (2.5). □

The localization argument also uses the coordinatewise geometric midpoint from [4, Section 6.1]. Its compatibility with $\Phi$ gives the following squared-norm identity and distance bounds.

**Lemma 2.6** (Geometric midpoint)**.** *For $x, y \in \mathcal{K}$, define*

$$m(x,y)_i^{(r)} = \sqrt{x_i^{(r)} y_i^{(r)}}.$$

*Then $m(x,y) \in \mathcal{K}$ and*

$$\|m(x,y)\|_2^2 = \frac{\|x\|_2^2 + \|y\|_2^2 - \|x-y\|_2^2}{2},$$

$$D(x, m(x,y)) \le D(x,y), \tag{2.6}$$

$$\rho(x, m(x,y)) \le \rho(x,y). \tag{2.7}$$

*Proof.* For each $r$, Cauchy–Schwarz gives

$$\left\|m(x,y)^{(r)}\right\|_2^2 = \left\langle x^{(r)}, y^{(r)} \right\rangle \le \left\|x^{(r)}\right\|_2 \left\|y^{(r)}\right\|_2 \le 1.$$

Thus $m(x,y) \in \mathcal{K}$. Summing over $r$ and applying polarization proves the squared-norm identity. Since the coefficients and coordinates are nonnegative,

$$\Phi(m(x,y))_{\boldsymbol{i}} = \sqrt{\Phi(x)_{\boldsymbol{i}} \Phi(y)_{\boldsymbol{i}}}.$$

The scalar inequality

$$|u - \sqrt{uv}| \le |u - v|, \qquad u, v \ge 0,$$

applied to the coordinates of $\Phi(x)$ and $\Phi(y)$ gives (2.6). Applying it to the mode coordinates and using the unconditionality of $N_r$ gives (2.7). □

2.4. **Localization and the canonical pseudometric.** We now combine the square-chaining estimate, the derivative bounds, and the geometric midpoint identity. The argument adapts the localization method of [4, Section 6] to $\Phi$.

**Proposition 2.7** (Chaining bound for the canonical pseudometric)**.** *With the preceding definitions,*

$$\gamma_2(\mathcal{K}, D) \lesssim_d \sigma(A) + \beta(A).$$

*Proof.* We apply the contraction principle in Proposition 2.3((ii)). To verify its entropy condition, we will cover each subset of $\mathcal{K}$ by first approximating its points through a regularized minimization problem and then covering the resulting set by linearization. The approximation errors and the costs of the linear covers will give the functions $s_j$. Fix a sufficiently small absolute $\theta > 0$, and let $b > 0$ be chosen at the end of the proof.

For $t > 0$ and $z \in \mathcal{K}$, set

$$J_t(z) = \min_{w \in \mathcal{K}} \{ \|w\|_2^2 + tD(z,w) + t^2 \rho(z,w)^2 \},$$

and choose a minimizer $\pi_t z$, which exists by compactness and continuity. The two distance penalties control the errors in the canonical pseudometric $D$ and in the auxiliary pseudometric $\rho$ used for linearization. Accordingly, put $t_j = b2^{j/2}$ for $j \ge -1$ and, for $j \ge 0$, write

$$\epsilon_j(z) = D(z, \pi_{t_j} z), \qquad \eta_j(z) = \rho(z, \pi_{t_j} z).$$

Testing $w = z$ gives $0 \le J_t(z) \le \|z\|_2^2 \le d$. Testing $\pi_{t_j} z$ in the minimization problem at scale $t_{j-1}$ gives

$$J_{t_j}(z) - J_{t_{j-1}}(z) \ge (1 - 2^{-1/2}) t_j \epsilon_j(z) + \frac{1}{2} t_j^2 \eta_j(z)^2.$$

The left-hand side telescopes, so

$$\sup_{z \in \mathcal{K}} \sum_{j \ge 0} [t_j \epsilon_j(z) + t_j^2 \eta_j(z)^2] \lesssim d. \tag{2.8}$$

To turn these approximations into covers, we need to control both the variation of $\mathrm{d}\Phi$ and the size of its linear images. For the first purpose, the linearization estimate (2.5) suggests working inside sets of controlled $\rho$-diameter. By (2.3), choose admissible partitions $(\mathcal{Q}_j)_{j\geq 0}$ of $\mathcal{K}$ such that

$$\sup_{z\in\mathcal{K}} \sum_{j\geq 0} 2^j \operatorname{diam}_\rho(\mathcal{Q}_j(z))^2 \lesssim_d [\sigma(A)+\beta(A)]^2. \tag{2.9}$$

For the second purpose, record the covering radii of the derivative images by setting

$$\lambda_j(z) = e_j\big(\mathrm{d}\Phi(z)B_2^{n_1+\cdots+n_d}, \|\cdot\|_2\big).$$

Proposition 2.3((iv)) and (2.4) yield

$$\sup_{z\in\mathcal{K}} \sum_{j\geq 0} 2^j \lambda_j(z)^2 \lesssim_d \sigma(A)^2. \tag{2.10}$$

Thus the partition diameters and the derivative covering radii both have the square-summability needed below.

We now construct the cover required by the contraction principle. Fix $j \geq 1$ and a nonempty $U \subseteq \mathcal{K}$. We use the partition $\mathcal{Q}_{j-1}$ and a level-$(j-1)$ cover inside each of its cells. This shift allocates the available points between the two operations: fewer than $2^{2^{j-1}}$ cells, each covered by fewer than $2^{2^{j-1}}$ points, give fewer than $2^{2^j}$ points in total, as required for $e_j(U,D)$. Fix a cell $Q \in \mathcal{Q}_{j-1}$ with $U_0 = U \cap Q \neq \varnothing$, and denote the Euclidean diameter of its approximating points by

$$h = \operatorname{diam}_2\big(\pi_{t_j}(U_0)\big).$$

The minimizing property of $\pi_{t_j}$ makes this diameter small enough for linearization to be useful. Indeed, take $z, z' \in U_0$, set $x = \pi_{t_j} z$ and $y = \pi_{t_j} z'$, and exchange the two pairs if necessary so that $\|x\|_2^2 \geq \|y\|_2^2$. Using the geometric midpoint from Lemma 2.6 as a competitor in the minimization defining $x$, we obtain

$$\begin{aligned}
\frac{1}{2}\|x-y\|_2^2 &\leq \|x\|_2^2 - \|m(x,y)\|_2^2 \\
&\leq t_j\big[D(z,m(x,y)) - D(z,x)\big] \\
&\qquad + t_j^2\big[\rho(z,m(x,y))^2 - \rho(z,x)^2\big] \\
&\leq t_j D(x,y) + t_j^2 \rho(z,m(x,y))^2.
\end{aligned}$$

Here the last inequality uses (2.6). Since $z, z' \in Q$, the triangle inequality and (2.7) give

$$\begin{aligned}
D(x,y) &\leq \operatorname{diam}_D(U) + 2\sup_{z\in U_0} \epsilon_j(z), \\
\rho(z,m(x,y)) &\leq \operatorname{diam}_\rho(Q) + 3\sup_{z\in U_0} \eta_j(z).
\end{aligned}$$

Taking the supremum over $z, z' \in U_0$ therefore yields

$$\begin{aligned}
h^2 \lesssim\ & t_j\left[\operatorname{diam}_D(U) + \sup_{z\in U_0}\epsilon_j(z)\right] \\
& + t_j^2\left[\operatorname{diam}_\rho(Q)^2 + \sup_{z\in U_0}\eta_j(z)^2\right].
\end{aligned} \tag{2.11}$$

Fix any $z_0 \in U_0$. The inclusion

$$\mathrm{d}\Phi(z_0)\big(\pi_{t_j}(U_0) - \pi_{t_j} z_0\big) \subseteq h\,\mathrm{d}\Phi(z_0)B_2^{n_1+\cdots+n_d}$$

allows us to cover the linear image by fewer than $2^{2^{j-1}}$ balls with radii arbitrarily close to $h\lambda_{j-1}(z_0)$. Assign each point of $U_0$ to a ball containing the linear image of its approximation,

and choose one original point from each nonempty group as a covering centre. If $x, y \in \pi_{t_j}(U_0)$ correspond to the same group, then, up to an arbitrarily small error,

$$\begin{aligned}\|\mathrm{d}\Phi(z_0)(x-y)\|_2 &\le 2h\lambda_{j-1}(z_0),\\ \rho(x,z_0),\ \rho(y,z_0) &\le \operatorname{diam}_\rho(Q) + \sup_{z\in U_0}\eta_j(z).\end{aligned}$$

Apply (2.5), use $\|x-y\|_2 \le h$, and add the two $D$-approximation errors to return to the original points. Letting the error in the covering radii tend to zero gives

$$\begin{aligned}e_{j-1}(U_0, D) \lesssim\ & \sup_{z\in U_0}\epsilon_j(z)\\ &+ h\left[\lambda_{j-1}(z_0) + \operatorname{diam}_\rho(Q) + \sup_{z\in U_0}\eta_j(z)\right].\end{aligned} \tag{2.12}$$

The centres lie in $U_0$. If $h = 0$, all approximating points coincide, and one point of $U_0$ covers it with radius at most $2\sup_{z\in U_0}\epsilon_j(z)$, so the same estimate holds.

Substitute (2.11) into (2.12) and apply Young's inequality, with coefficient $\theta$ for the term involving $\operatorname{diam}_D(U)$. For the fixed $\theta$, the resulting covering radius is at most

$$\begin{aligned}\theta\operatorname{diam}_D(U) + C_\theta\Big\{ & \sup_{z\in U_0}\epsilon_j(z)\\ &+ t_j\big[\lambda_{j-1}(z_0)^2 + \operatorname{diam}_\rho(Q)^2 + \sup_{z\in U_0}\eta_j(z)^2\big]\Big\}.\end{aligned}$$

This identifies the functions needed in the contraction principle: for $j \ge 1$, define

$$s_j(z) = C_\theta\left[\epsilon_j(z) + t_j\left(\lambda_{j-1}(z)^2 + \operatorname{diam}_\rho(\mathcal{Q}_{j-1}(z))^2 + \eta_j(z)^2\right)\right], \tag{2.13}$$

with $C_\theta$ enlarged if necessary. On $U_0$, the partition term is constant because $\mathcal{Q}_{j-1}(z) = Q$, and the derivative term at $z_0$ is bounded by its supremum over $U_0$. Each of the four nonnegative terms in the preceding radius bound is thus controlled by the supremum of the bracket in (2.13); their sum costs only an additional absolute factor. Consequently, $U_0$ has a cover of the stated cardinality with radius arbitrarily close to

$$\theta\operatorname{diam}_D(U) + \sup_{z\in U_0} s_j(z).$$

Combining these covers over the cells of $\mathcal{Q}_{j-1}$ gives fewer than $2^{2^j}$ centres in $U$, and hence

$$e_j(U, D) \le \theta\operatorname{diam}_D(U) + \sup_{z\in U} s_j(z).$$

For $j = 0$, the same inequality holds with $s_0(z) = \operatorname{diam}_D(\mathcal{K})$.

It remains to sum the covering costs and apply the contraction principle. Since $\sup_{z\in\mathcal{K}}\|\Phi(z)\|_2 \le \max_i a_i$, the initial level contributes at most $\operatorname{diam}_D(\mathcal{K}) \le 2\sigma(A)$. For the remaining levels, the identities

$$2^{j/2}\epsilon_j(z) = \frac{t_j\epsilon_j(z)}{b}, \qquad 2^{j/2}t_j\eta_j(z)^2 = \frac{t_j^2\eta_j(z)^2}{b}$$

show that (2.8) controls the two approximation errors by a constant multiple of $d/b$. The other two terms in (2.13) are controlled by (2.9) and (2.10), because $2^{j/2}t_j = b2^j$ and the shift from $j$ to $j-1$ changes each square sum only by a factor of two. Therefore

$$\sup_{z\in\mathcal{K}}\sum_{j\ge 0} 2^{j/2}s_j(z) \lesssim_d \sigma(A) + \frac{1}{b} + b[\sigma(A) + \beta(A)]^2.$$

For completeness, (2.4) and convexity of $\mathcal{K}$ give $D(x,y) \le \sqrt{d}\,\sigma(A)\,\|x-y\|_2$. Finite-dimensional Euclidean covering estimates then imply $\gamma_2(\mathcal{K}, D) < \infty$, as required by the contraction principle. If $D$ is only a pseudometric, apply that principle to a set containing one representative of each zero-distance class, restrict the $\rho$-partitions to this set, and pull

back the resulting $D$-partitions to $\mathcal{K}$. The minimizers $\pi_t z$ remain defined over the original compact set $\mathcal{K}$, so all the preceding estimates still apply. Proposition 2.3((ii)) now gives

$$\gamma_2(\mathcal{K}, D) \lesssim_d \sigma(A) + \frac{1}{b} + b[\sigma(A) + \beta(A)]^2.$$

Choosing $b = [\sigma(A) + \beta(A)]^{-1}$ proves the claim. □

### 2.5. **Completion of the proof.**

*Proof of Theorem 1.1.* The process $\langle g, \Phi(x)\rangle$ is continuous on $\mathcal{K}$, has canonical pseudometric $D$, and vanishes at $x = 0$. Proposition 2.3((i)) and Proposition 2.7 give

$$\mathbb{E} \sup_{x\in\mathcal{K}} |\, \langle g, \Phi(x)\rangle \,| \lesssim \gamma_2(\mathcal{K}, D) \lesssim_d \sigma(A) + \beta(A).$$

For arbitrary $x^{(r)} \in B_2^{n_r}$, write $x^{(r)} = x_+^{(r)} - x_-^{(r)}$, with both parts in $B_{2,+}^{n_r}$. Expanding the multilinear form gives

$$\|(a_i g_i)\|_{\mathrm{inj}} \le 2^d \sup_{x\in\mathcal{K}} |\, \langle g, \Phi(x)\rangle \,|,$$

which proves the upper bound in (2.1).

For the lower bound, fixing all but one mode at coordinate vectors gives

$$\|(a_i g_i)\|_{\mathrm{inj}} \ge \left(\sum_{i\in f} a_i^2 g_i^2\right)^{1/2}, \qquad f \in \mathcal{F}.$$

Jensen's inequality applied to the Euclidean norm of $(a_i|g_i|)_{i\in f}$ yields

$$\mathbb{E}\left(\sum_{i\in f} a_i^2 g_i^2\right)^{1/2} \ge \left(\sum_{i\in f} (a_i \mathbb{E}|g_i|)^2\right)^{1/2} = \sqrt{\frac{2}{\pi}}\left(\sum_{i\in f} a_i^2\right)^{1/2}.$$

Maximizing over $f$ gives the lower bound in terms of $\sigma(A)$. Testing individual entries and using Lemma 2.1 also gives

$$\mathbb{E}\,\|(a_i g_i)\|_{\mathrm{inj}} \ge \mathbb{E} \max_{i\in\mathcal{I}} a_i|g_i| \gtrsim \beta(A).$$

Combining the two estimates proves the lower bound in (2.1). □

## 3. Sparse Gaussian tensors

In this section, we prove Theorem 1.3 and derive Corollaries 1.5 and 1.6.

For any nonnegative coefficient tensor $B = (b_i)_{i\in\mathcal{I}}$, we use the fibre parameter $\sigma(B)$ and the entrywise parameter $\beta(B)$ introduced in Sections 1 and 2, respectively. The decreasing rearrangement in $\beta(B)$ includes zero entries. Writing $A \odot \delta = (a_i\delta_i)_{i\in\mathcal{I}}$, conditioning on the masks and applying Theorem 1.1 reduces the proof of Theorem 1.3 to the estimate

$$\mathbb{E}_\delta \sigma(A \odot \delta) + \mathbb{E}_\delta \beta(A \odot \delta) \asymp_d \mathfrak{e}(A, P) + \mathfrak{v}(A, P).$$

We first show that the entrywise term $\mathbb{E}_\delta\beta(A \odot \delta)$ is comparable to $\mathfrak{e}(A, P)$ and establish the scalar truncation estimates. We then average the fibre parameter in Proposition 3.5. The final subsection combines these estimates with the conditional Gaussian bound and proves the two corollaries.

3.1. **The sparse entrywise maximum.** We now give an integral expression equivalent to the expected maximum in (1.4) and relate it to the ordered entrywise profile after masking.

**Lemma 3.1** (The sparse entrywise maximum)**.** *For every choice of profiles $A, P$,*

$$\mathbb{E}_\delta \beta(A \odot \delta) \asymp \mathfrak{e}(A, P).$$

*Moreover, define*

$$I(A,P) = \int_0^\infty \left[ 1 \wedge \sum_{\substack{i \in \mathcal{I} \\ a_i > 0}} p_i \exp\left(-\frac{t^2}{a_i^2}\right) \right] \mathrm{d}t, \tag{3.1}$$

*where zero amplitudes are omitted from the sum. Then*

$$\mathfrak{e}(A,P) \asymp I(A,P), \qquad \mathfrak{e}(A,P) \le \sqrt{2}\, I(A,P).$$

*Proof.* Conditioning on the masks and applying Lemma 2.1 gives the first comparison. Let $g$ be a standard Gaussian variable. Independence and integration of the tail give

$$\mathfrak{e}(A,P) = \int_0^\infty \left[ 1 - \prod_{i:\, a_i > 0} (1 - p_i \mathbb{P}\{|g| > t/a_i\}) \right] \mathrm{d}t.$$

For any finite family $z_i \in [0,1]$,

$$(1 - \mathrm{e}^{-1}) \left( 1 \wedge \sum_i z_i \right) \le 1 - \prod_i (1 - z_i) \le 1 \wedge \sum_i z_i.$$

The lower bound follows from $\prod_i (1 - z_i) \le \exp(-\sum_i z_i)$, and the upper bound is the union bound. The Gaussian tail satisfies

$$\exp(-u^2) \lesssim \mathbb{P}\{|g| > u\} \le \exp(-u^2/2), \qquad u \ge 0.$$

The lower estimate follows by integrating the density over $[u, u + (u+1)^{-1}]$. For the upper estimate, shift the density by $u$ and use $\exp(-(s+u)^2/2) \le \exp(-u^2/2)\exp(-s^2/2)$ for $s \ge 0$. Substituting these bounds into the tail integral gives $\mathfrak{e}(A,P) \gtrsim I(A,P)$ and

$$\mathfrak{e}(A,P) \le \int_0^\infty \left[ 1 \wedge \sum_{i:\, a_i > 0} p_i \exp\left(-\frac{t^2}{2a_i^2}\right) \right] \mathrm{d}t = \sqrt{2}\, I(A,P).$$

This proves the remaining assertions. □

We use $\mathfrak{e}(A,P)$ directly as the truncation scale. The estimates below involving $A, P$ are immediate when $\mathfrak{e}(A,P) = 0$, since every entry with positive amplitude then has activation probability zero. In their proofs, we therefore assume $\mathfrak{e}(A,P) > 0$.

**Lemma 3.2** (Contribution of large amplitudes)**.** *For every choice of profiles $A, P$,*

$$\sum_{i:\, a_i > 4\mathfrak{e}(A,P)} p_i a_i \le 4\mathfrak{e}(A,P).$$

*Proof.* Conditional on the masks, test a position at which $a_i \delta_i$ is largest. Using $\mathbb{E}|g| = \sqrt{2/\pi}$ and (1.4) gives

$$\begin{aligned} \mathbb{E}_\delta \max_{i \in \mathcal{I}} a_i \delta_i &\le \sqrt{\pi/2}\, \mathbb{E}_{\delta,g} \max_{i \in \mathcal{I}} a_i \delta_i |g_i| \\ &= \sqrt{\pi/2}\, \mathfrak{e}(A,P) \le 2\mathfrak{e}(A,P). \end{aligned}$$

By Markov's inequality, the probability that an entry with amplitude greater than $4\mathfrak{e}(A,P)$ is activated is at most $1/2$. Hence

$$\prod_{i:\, a_i > 4\mathfrak{e}(A,P)} (1 - p_i) \ge \frac{1}{2}.$$

The events that exactly one of these entries is activated are disjoint. Therefore

$$\mathbb{E}_\delta \max_{i\in\mathcal{I}} a_i\delta_i \geq \sum_{i:\, a_i>4\mathfrak{e}(A,P)} a_i p_i \prod_{\substack{j:\, a_j>4\mathfrak{e}(A,P)\\ j\neq i}} (1-p_j)$$
$$\geq \frac{1}{2} \sum_{i:\, a_i>4\mathfrak{e}(A,P)} a_i p_i.$$

Combining the two bounds proves the claim. □

The next estimate controls the second moment of the entrywise profile after truncation. It will be used to bound the fluctuations of the fibre sums.

**Lemma 3.3** (Second moment after truncation)**.** *Let $B = (b_i)_{i\in\mathcal{I}}$, where $b_i = a_i \wedge 4\mathfrak{e}(A,P)$. Then*

$$\mathbb{E}_\delta \beta(B\odot\delta)^2 \lesssim \mathfrak{e}(A,P)^2.$$

*Proof.* Put $W = \max_{i\in\mathcal{I}} b_i\delta_i|g_i|$. Since $b_i \leq a_i$, the definition (1.4) gives $\mathbb{E}W \leq \mathfrak{e}(A,P)$. Markov's inequality and independence therefore imply

$$\prod_{i:\, b_i>0} (1 - p_i\mathbb{P}\{|g| > 4\mathfrak{e}(A,P)/b_i\}) \geq \frac{1}{2},$$

where $g$ is a standard Gaussian variable. Using $1-x \leq \mathrm{e}^{-x}$, we obtain

$$\sum_{i:\, b_i>0} p_i\mathbb{P}\{|g| > 4\mathfrak{e}(A,P)/b_i\} \leq \log 2.$$

For $v \geq u \geq 0$, shifting the Gaussian density gives

$$\mathbb{P}\{|g| > v\} \leq \exp\left(-\frac{v^2-u^2}{2}\right)\mathbb{P}\{|g| > u\}.$$

Since $b_i \leq 4\mathfrak{e}(A,P)$, the preceding two estimates and a union bound yield

$$\mathbb{P}\{W > t\} \leq (\log 2)\exp\left(-\frac{t^2 - 16\mathfrak{e}(A,P)^2}{32\mathfrak{e}(A,P)^2}\right), \qquad t \geq 4\mathfrak{e}(A,P).$$

Integration gives $\mathbb{E}W^2 \lesssim \mathfrak{e}(A,P)^2$. Conditional on the masks, Lemma 2.1 and Jensen's inequality give

$$\beta(B\odot\delta)^2 \lesssim (\mathbb{E}_g W)^2 \leq \mathbb{E}_g W^2.$$

Averaging over the masks proves the result. □

3.2. **Averaging the fibre profile.** To control the fluctuations of the fibre sums, we first establish an estimate for their symmetrized counterparts. The argument uses the fact that the fibres in each $\mathcal{F}_r$ are pairwise disjoint.

**Lemma 3.4** (Signed sums along fibres)**.** *For every deterministic nonnegative coefficient tensor $B$ and independent Rademacher variables $(\varepsilon_i)$,*

$$\mathbb{E}_\varepsilon \max_{f\in\mathcal{F}} \left|\sum_{i\in f} \varepsilon_i b_i^2\right| \lesssim_d \beta(B)\sigma(B).$$

*Proof.* The case $B = 0$ is immediate. Fix a mode $r$. For each $f \in \mathcal{F}_r$, the Rademacher sum has sub-Gaussian variance proxy

$$\sum_{i\in f} b_i^4 \leq \left(\max_{i\in f} b_i\right)^2 \sigma(B)^2.$$

Choose a largest coefficient in each fibre. Since the fibres in $\mathcal{F}_r$ are disjoint, these coefficients occupy distinct positions. Their $j$th largest value is therefore at most $b_j^*$. The sub-Gaussian tail bound and a union bound give, for $t \geq 2$,

$$\begin{aligned}\mathbb{P}\left\{\max_{f\in\mathcal{F}_r}\left|\sum_{i\in f}\varepsilon_i b_i^2\right| > t\beta(B)\sigma(B)\right\} &\leq 2\sum_{\substack{f\in\mathcal{F}_r\\ \max_{i\in f} b_i>0}} \exp\left(-\frac{t^2\beta(B)^2}{2(\max_{i\in f} b_i)^2}\right)\\ &\leq 2\sum_{j\geq 1}(j+1)^{-t^2/2} \lesssim 2^{-t^2/2}.\end{aligned}$$

Integrating this bound gives the desired estimate for a fixed mode, with an absolute comparison constant. Since $\mathcal{F} = \bigcup_{r=1}^d \mathcal{F}_r$, summing over the modes proves the stated bound. □

We can now average the fibre profile. The extreme term $\mathfrak{e}(A,P)$ accounts for both the truncation error and the fluctuations of the truncated fibre sums.

**Proposition 3.5** (Mask averaging)**.** *For every choice of profiles $A, P$,*

$$\mathfrak{e}(A,P) + \mathbb{E}_\delta\sigma(A\odot\delta) \asymp_d \mathfrak{e}(A,P) + \mathfrak{v}(A,P).$$

*Proof.* Set $b_i = a_i \wedge 4\mathfrak{e}(A,P)$ and $B = (b_i)_{i\in\mathcal{I}}$. By the triangle inequality and the bound of the maximum fibre norm by the entrywise $\ell_1$ norm, Lemma 3.2 gives

$$\mathbb{E}_\delta\sigma(A\odot\delta) \leq \mathbb{E}_\delta\sigma(B\odot\delta) + 4\mathfrak{e}(A,P).$$

It remains to control the truncated tensor. Recalling the definition of $v_s(A,P)$ in (1.5), symmetrization yields

$$\begin{aligned}\mathbb{E}_\delta\sigma(B\odot\delta)^2 &\leq v_{4\mathfrak{e}(A,P)}(A,P)^2 + \mathbb{E}_\delta\max_{f\in\mathcal{F}}\left|\sum_{i\in f} b_i^2(\delta_i - p_i)\right|\\ &\leq v_{4\mathfrak{e}(A,P)}(A,P)^2 + 2\mathbb{E}_{\delta,\varepsilon}\max_{f\in\mathcal{F}}\left|\sum_{i\in f}\varepsilon_i b_i^2\delta_i\right|.\end{aligned}$$

For the second inequality, introduce an independent copy of the masks, apply conditional Jensen's inequality, insert independent signs in the differences, and use the triangle inequality. Conditional on the masks, Lemma 3.4 applies to $B\odot\delta$. Together with Lemma 3.3 and Cauchy–Schwarz, it gives

$$\begin{aligned}\mathbb{E}_\delta\sigma(B\odot\delta)^2 &\lesssim_d v_{4\mathfrak{e}(A,P)}(A,P)^2 + \mathbb{E}_\delta\big[\beta(B\odot\delta)\sigma(B\odot\delta)\big]\\ &\lesssim_d v_{4\mathfrak{e}(A,P)}(A,P)^2 + \mathfrak{e}(A,P)\big[\mathbb{E}_\delta\sigma(B\odot\delta)^2\big]^{1/2}.\end{aligned}$$

Solving this quadratic inequality and using Jensen's inequality, we obtain

$$\begin{aligned}\mathbb{E}_\delta\sigma(B\odot\delta) &\leq \big[\mathbb{E}_\delta\sigma(B\odot\delta)^2\big]^{1/2}\\ &\lesssim_d \mathfrak{e}(A,P) + v_{4\mathfrak{e}(A,P)}(A,P) \lesssim_d \mathfrak{e}(A,P) + \mathfrak{v}(A,P).\end{aligned}$$

The last step uses $a \wedge 4\mathfrak{e}(A,P) \leq 4(a\wedge\mathfrak{e}(A,P))$. Adding the truncation error proves the upper bound.

For the reverse bound, choose a fibre $f$ attaining the maximum in the definition of $\mathfrak{v}(A,P)$, and put

$$Z = \sum_{i\in f}(a_i \wedge \mathfrak{e}(A,P))^2\delta_i.$$

Independence gives

$$\mathbb{E}_\delta Z = \mathfrak{v}(A,P)^2,$$

and

$$\mathrm{Var}(Z) = \sum_{i \in f} p_i(1-p_i)(a_i \wedge \mathfrak{e}(A,P))^4 \le \mathfrak{e}(A,P)^2 \mathfrak{v}(A,P)^2.$$

If $\mathfrak{v}(A,P) \ge \mathfrak{e}(A,P)$, then $\mathbb{E}_\delta Z^2 \le 2\mathfrak{v}(A,P)^4$. The Paley–Zygmund inequality therefore yields

$$\mathbb{P}\{Z \ge \mathfrak{v}(A,P)^2/2\} \ge \frac{1}{8}, \qquad \mathbb{E}_\delta \sigma(A \odot \delta) \ge \mathbb{E}_\delta \sqrt{Z} \gtrsim \mathfrak{v}(A,P).$$

If $\mathfrak{v}(A,P) < \mathfrak{e}(A,P)$, the extreme term already controls $\mathfrak{v}(A,P)$. This proves the reverse bound. □

3.3. **Completion of the proofs.** The scalar and fibre estimates now allow us to average the conditional Gaussian bound.

*Proof of Theorem 1.3.* If $\mathfrak{e}(A,P) = 0$, every masked entry vanishes almost surely, and the conclusion is immediate. Otherwise, condition on the masks and apply Theorem 1.1 to the coefficient tensor $A \odot \delta$. Its comparison constants depend only on $d$ and are uniform over all realizations of the masks. Averaging gives

$$\begin{aligned} \mathbb{E}\,\|U\|_{\mathrm{inj}} &\asymp_d \mathbb{E}_\delta \sigma(A \odot \delta) + \mathbb{E}_\delta \beta(A \odot \delta) \\ &\asymp_d \mathbb{E}_\delta \sigma(A \odot \delta) + \mathfrak{e}(A,P) \\ &\asymp_d \mathfrak{e}(A,P) + \mathfrak{v}(A,P). \end{aligned}$$

The second comparison follows from Lemma 3.1, and the last from Proposition 3.5. All expectations are finite, since the index set $\mathcal{I}$ is finite. This proves both bounds in (1.7). □

*Proof of Corollary 1.5.* Let $N = \sum_{i:\, a_i > 0} \delta_i$, so that $\mathbb{E}N = \Lambda$. Conditional on the masks, Lemma 2.1 gives

$$\mathbb{E}_g \max_{i \in \mathcal{I}} a_i \delta_i |g_i| \lesssim a_* \sqrt{\log(1+N)},$$

with both sides zero when $N = 0$. Since $x \mapsto \sqrt{\log(1+x)}$ is concave on $[0,\infty)$, Jensen's inequality yields

$$\mathfrak{e}(A,P) \lesssim a_* \mathbb{E}\sqrt{\log(1+N)} \le a_* \sqrt{\log(1+\Lambda)}.$$

Combining this with $\mathfrak{v}(A,P) \le V(A,P)$ and Theorem 1.3 proves the claim. □

We now deduce the Bernoulli estimate by symmetrization.

*Proof of Corollary 1.6.* Let $\delta'$ be an independent copy of $\delta$, and recall $q_i = 2p_i(1-p_i)$ and $Q = (q_i)$. The differences have the joint distribution

$$(\delta_i - \delta_i')_{i \in \mathcal{I}} \overset{d}{=} (\varepsilon_i \eta_i)_{i \in \mathcal{I}},$$

where the $\varepsilon_i$ are independent Rademacher variables, the $\eta_i$ are independent Bernoulli variables with parameters $q_i$, and the two families are independent. Symmetrization and the Rademacher-to-Gaussian comparison give

$$\begin{aligned} \mathbb{E}\,\|A \odot (\delta - P)\|_{\mathrm{inj}} &\le \mathbb{E}\,\|A \odot \varepsilon \odot \eta\|_{\mathrm{inj}} \\ &\le \sqrt{\pi/2}\,\mathbb{E}\,\|A \odot g \odot \eta\|_{\mathrm{inj}}\,. \end{aligned}$$

For the first inequality, use conditional Jensen's inequality with respect to $\delta'$. For the second, condition on $\eta$ and the signs, and apply Jensen's inequality to the independent Gaussian magnitudes, whose means are $\sqrt{2/\pi}$. The Gaussian signs are independent of these magnitudes and have the same distribution as $\varepsilon$. Theorem 1.3, applied with probability profile $Q$, now gives (1.9). □

**Acknowledgement.** Ke Wang is supported by Hong Kong RGC grant GRF 16305526.

**Declaration of AI use.** ChatGPT Astra was used to assist with mathematical discussions, proof review, and language editing during the preparation of this paper. The authors take full responsibility for the content and correctness of the paper.

## References


[1] I. Aden-Ali, *On the injective norm of sums of random tensors and the moments of Gaussian chaoses*, Preprint, arXiv:2503.10580 (2025).

[2] A. S. Bandeira, S. Gopi, H. Jiang, K. Lucca, and T. Rothvoss, *Tensor concentration inequalities: A geometric approach*, in Proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC '25), ACM, 2025, 822–832. doi:10.1145/3717823.3718188.

[3] A. S. Bandeira and R. van Handel, *Sharp nonasymptotic bounds on the norm of random matrices with independent entries*, Ann. Probab. **44** (2016), no. 4, 2479–2506. doi:10.1214/15-AOP1025.

[4] W. Bednorz, R. Martynek, and R. Meller, *A candidate proof of the sharp Latała and Strzelecka Gaussian matrix bound*, Preprint, arXiv:2609.22927v1 (2026).

[5] M. T. Boedihardjo, *Injective norm of random tensors with independent entries*, Preprint, arXiv:2412.21193v2 (2025).

[6] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, Oxford, 2013. doi:10.1093/acprof:oso/9780199535255.001.0001.

[7] G. Dai, Y. He, K. Wang, and Y. Zhu, *A note on the improved sparse Hanson–Wright inequalities*, Preprint, arXiv:2505.20799 (2025).

[8] G. Dai, Y. He, K. Wang, and Y. Zhu, *Concentration of bounded sparse chaoses and sparse Khatri–Rao embeddings*, Preprint, arXiv:2609.28875 (2026).

[9] G. Dai and K. Wang, *Sparsity-adaptive concentration inequalities for random polynomials*, Preprint, arXiv:2606.24090 (2026).

[10] S. Dartois and B. McKenna, *Injective norm of real and complex random tensors I: From spin glasses to geometric entanglement*, Preprint, arXiv:2404.03627 (2024).

[11] S. Dirksen, *Tail bounds via generic chaining*, Electron. J. Probab. **20** (2015), paper no. 53, 1–29. doi:10.1214/EJP.v20-3760.

[12] Y. He, K. Wang, and Y. Zhu, *Sparse Hanson–Wright inequalities with applications*, Electron. J. Probab. **31** (2026), 1–49. doi:10.1214/26-EJP1493.

[13] P. Jain and S. Oh, *Provable tensor factorization with missing data*, Advances in Neural Information Processing Systems **27** (2014), 1431–1439. arXiv:1406.2784.

[14] R. Latała, R. van Handel, and P. Youssef, *The dimension-free structure of nonhomogeneous random matrices*, Invent. Math. **214** (2018), no. 3, 1031–1080. doi:10.1007/s00222-018-0817-x.

[15] Z. Lin and H. Lu, *Optimal fiber comparisons for random tensors*, Preprint, arXiv:2610.39389v1 (2026).

[16] K. Lucca, *Tensor concentration inequalities (problem 16)*, Randomstrasse 101, April 30, 2025; updated May 5, 2025. `https://randomstrasse101.math.ethz.ch/posts/tensor-concentration/`. Accessed October 6, 2026.

[17] K. Lucca and L. Pesenti, *Norm bounds for sparse random tensors and spectral gap of random hypergraphs*, Preprint, arXiv:2607.07308v1 (2026).

[18] A. Montanari and E. Richard, *A statistical model for tensor PCA*, Advances in Neural Information Processing Systems **27** (2014), 2897–2905. arXiv:1411.1076.

[19] M. Talagrand, *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*, Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, vol. 60, Springer, Heidelberg, 2014. doi:10.1007/978-3-642-54075-2.

[20] R. Tomioka and T. Suzuki, *Spectral norm of random tensors*, Technical report, arXiv:1407.1870 (2014).

[21] R. van Handel, *Structured random matrices*, in E. Carlen, M. Madiman, and E. M. Werner (eds.), *Convexity and Concentration*, The IMA Volumes in Mathematics and its Applications, vol. 161, Springer, New York, 2017, 107–156. doi:10.1007/978-1-4939-7005-6_4.

[22] R. van Handel, *Chaining, interpolation, and convexity II: The contraction principle*, Ann. Probab. **46** (2018), no. 3, 1764–1805. doi:10.1214/17-AOP1214.

[23] Z. Zhou and Y. Zhu, *Sparse random tensors: Concentration, regularization and applications*, Electron. J. Stat. **15** (2021), no. 1, 2483–2516. doi:10.1214/21-EJS1838.

[24] Z. Zhou and Y. Zhu, *Sharp spectral norm concentration and regularization of sparse random tensors*, Preprint, arXiv:2609.20520v2 (2026).

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

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