CONCENTRATION INEQUALITIES FOR POLYNOMIALS IN α-SUB-EXPONENTIAL RANDOM VARIABLES

FRIEDRICH GÖTZE[1], HOLGER SAMBALE[1] AND ARTHUR SINULIS[1]

ABSTRACT. In this work we derive multi-level concentration inequalities for polynomial functions in independent random variables with a α-sub-exponential tail decay. A particularly interesting case is given by quadratic forms $f(X_1, \ldots, X_n) = \langle X, AX \rangle$, for which we prove Hanson–Wright-type inequalities with explicit dependence on various norms of the matrix A. A consequence of these inequalities is a two-level concentration inequality for quadratic forms in α-sub-exponential random variables, such as quadratic Poisson chaos.

We provide various applications of these inequalities. Among these are generalizations the results given by Rudelson–Vershynin from sub-Gaussian to α-sub-exponential random variables, i.e. concentration of the Euclidean norm of the linear image of a random vector, small ball probability estimates and concentration inequalities for the distance between a random vector and a fixed subspace. Moreover, we obtain concentration inequalities for the excess loss in a fixed design linear regression and the norm of a randomly projected random vector.

1. INTRODUCTION

Let $X_1, \ldots, X_n$ be independent random variables and let $f : \mathbb{R}^n \to \mathbb{R}$ be a measurable function. One of the main and rather classical questions of probability theory consists in finding good estimates on the fluctuations of $f(X_1, \ldots, X_n)$ around a deterministic value (e. g. its expectation or median), i. e. to determine a function $h : [0, \infty) \to [0, 1]$ such that

$$(1.1) \qquad \mathbb{P}\Big(|f(X_1, \ldots, X_n) - \mathbb{E} f(X_1, \ldots, X_n)| \geq t\Big) \leq h(t).$$

Of course, h should take into account both the information given by f as well as $X_1, \ldots, X_n$. Perhaps one of the most well-known *concentration inequalities* is the tail decay of the Gaussian distribution: if $X_1, \ldots, X_n$ are independent and are distributed as a standard a standard normal distribution $\mathcal{N}(0, 1)$, and $f(X_1, \ldots, X_n) = n^{-1/2} \sum_{i=1}^n X_i$, then $f(X_1, \ldots, X_n) \sim \mathcal{N}(0, 1)$ and

$$(1.2) \qquad \mathbb{P}\Big(|f(X_1, \ldots, X_n) - \mathbb{E} f(X_1, \ldots, X_n)| \geq t\Big) \leq 2 \exp\Big(-\frac{t^2}{2}\Big).$$

Using the entropy method, it is possible to show that the estimate (1.2) remains true for any Lipschitz function f (see e. g. [Led01, Section 5]). On the other hand, if f

is a polynomial of degree 2, then the tails of $f(X_1, \ldots, X_n)$ are heavier. Indeed, the *Hanson–Wright inequality* states that for a quadratic form in independent, standard Gaussian random variables $X_1, \ldots, X_n$ we have

$$(1.3) \quad \mathbb{P}\Big(\Big| \sum_{i,j=1}^{n} a_{ij} X_i X_j - \text{trace}(A) \Big| \geq t \Big) \leq 2 \exp\Big(-\frac{1}{C} \min\Big(\frac{t^2}{\|A\|_{\text{HS}}}, \frac{t}{\|A\|_{\text{op}}} \Big) \Big).$$

Here, $\|A\|_{\text{op}}$ is the operator norm and $\|A\|_{\text{HS}}$ the Hilbert–Schmidt norm (also called Frobenius norm) of A respectively. For a proof see [RV13]. Thus the tails of the quadratic form decay like $\exp(-t)$ for large t. There are inequalities similar to (1.3) for the multilinear chaos in Gaussian random variables proven in [Lat06] (and in fact, a lower bound using the same quantities as well), and in [AW15] for polynomials in sub-Gaussian random variables. However, a key component is that the individual random variables X_i have a sub-Gaussian tail decay, i. e. $\mathbb{P}(|X_i| \geq t) \leq c \exp(-Ct^2)$ for some constants c, C.

In recent works [BGS18], [GSS18b], [GSS18a] we have studied similar concentration inequalities for bounded functions f of either independent or weakly dependent random variables. There, the situation is clearly different, since the distribution of $f(X_1, \ldots, X_n)$ has a compact support, and is thus sub-Gaussian, and the challenge is to give an estimate depending on different quantities derived from f and X. However, there are many situations of interest where boundedness does not hold, such as quadratic forms in unbounded random variables as in (1.3). Here it seems reasonable to focus on certain classes of functions for which the tail behavior can directly be traced back to the tails of the random variables under consideration. Therefore, in this note we restrict ourselves to polynomial functions.

In the following results, the setup is as follows. We consider independent random variables $X_1, \ldots, X_n$ which have α-*sub-exponential tail decay.* By this we mean that there exists two constants c, C and a parameter $\alpha > 0$ such that for all $i = 1, \ldots, n$ and $t \geq 0$

$$(1.4) \qquad\qquad \mathbb{P}\left(|X_i| \geq t \right) \leq c \exp\left(-\frac{t^\alpha}{C} \right).$$

There are many interesting choices of random variables X_i of this type, such as bounded random variables (for any $\alpha > 0$), random variables with a sub-Gaussian (for $\alpha = 2$) or sub-exponential distribution ($\alpha = 1$) such as Poisson random variables, or "fatter" tails such as Weibull random variables with shape parameter $\alpha \in (0, 1]$.

We reformulate condition (1.4) in terms of so-called *exponential Orlicz norms*, but we emphasize that these two concepts are equivalent. For any random variable X on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $\alpha > 0$ define the (quasi-)norm

$$(1.5) \qquad\qquad \|X\|_{\psi_\alpha} := \inf \left\{ t > 0 \colon \mathbb{E} \exp\left(\frac{|X|^\alpha}{t^\alpha} \right) \leq 2 \right\},$$

adhering to the standard definition $\inf \emptyset = \infty$. Strictly speaking, this is a norm for $\alpha \geq 1$ only, since otherwise the triangle inequality does not hold. Nevertheless, the above expression makes sense for any $\alpha > 0$, and we choose to call it a norm in these cases as well. For some properties of the Orlicz norms in the case $\alpha \in (0, 1]$, see Appendix A. In this note we concentrate on values $\alpha = 2/q$ for some $q \in \mathbb{N}$, but also prove results for the case $\alpha \in (0, 1]$. Throughout this work, we denote by C an

absolute constant and by $C_{l_1,\ldots,l_k}$ a constant that only depends on some parameters $l_1, \ldots, l_k$.

For illustration, we start with a simplified version of some of our results which may already be sufficient for application purposes. The first result is a concentration inequality which may be considered as a generalization of the Hanson–Wright inequality (1.3) to quadratic forms in random variables with α-sub-exponential tail decay.

Proposition 1.1. *Let $X_1, \ldots, X_n$ be independent random variables satisfying $\mathbb{E}\, X_i = 0, \mathbb{E}\, X_i^2 = \sigma_i^2, \|X_i\|_{\Psi_\alpha} \le M$ for some $\alpha \in (0,1] \cup \{2\}$, and A be a symmetric $n \times n$ matrix. For any $t > 0$ we have*

$$\mathbb{P}\Big(\big|\sum_{i,j} a_{ij} X_i X_j - \sum_{i=1}^n \sigma_i^2 a_{ii}\big| \ge t\Big) \le 2\exp\Big(-\frac{1}{C}\min\Big(\frac{t^2}{M^4\|A\|_{\mathrm{HS}}^2}, \Big(\frac{t}{M^2\|A\|_{\mathrm{op}}}\Big)^{\frac{\alpha}{2}}\Big)\Big).$$

As we will see in Proposition 1.5, the tail decay $\exp(-t^{\alpha/2}\|A\|_{\mathrm{op}}^{-\alpha/2})$ (for large t) can be sharpened by replacing the operator norm by a smaller norm. Actually, the technical result contains up to four different regimes instead of two as above.

The next theorem provides tail estimates for polynomials in independent random variables. Note that this is not a generalization of Proposition 1.1 due to the use of the Hilbert–Schmidt instead of the operator norms.

Theorem 1.2. *Let $X_1, \ldots, X_n$ be independent random variables satisfying $\|X_i\|_{\Psi_\alpha} \le M$ for some $\alpha \in (0,1] \cup \{2\}$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be a polynomial of total degree $D \in \mathbb{N}$. Then for all $t > 0$*

$$(1.6) \quad \mathbb{P}(|f(X) - \mathbb{E} f(X)| \ge t) \le 2\exp\Big(-\frac{1}{C_{D,\alpha}}\min_{1\le d\le D}\Big(\frac{t}{M^d\|\mathbb{E} f^{(d)}(X)\|_{\mathrm{HS}}}\Big)^{\alpha/d}\Big).$$

In particular, if $\|\mathbb{E} f^{(d)}(X)\|_{\mathrm{HS}} \le 1$ for $d = 1, \ldots, D$, then

$$\mathbb{E}\exp\Big(\frac{C_{D,\alpha}}{M^\alpha}|f(X)|^{\frac{\alpha}{D}}\Big) \le 2,$$

or equivalently

$$\|f(X)\|_{\Psi_{\frac{\alpha}{D}}} \le C_{d,\alpha} M^D.$$

Intuitively, Theorem 1.2 states that a polynomial in random variables with tail decay as in (1.4) also exhibits α-sub-exponential tail decay whenever the Hilbert–Schmidt norms are not too large. Moreover, the tail decay is "as expected", i.e. one just needs to account for the total degree D by taking the D-th root.

One particularly interesting case is when the functional under consideration is a d-th order chaos. That is, given a d-tensor $A = (a_{i_1\ldots i_d})$ which we assume to be symmetric, i.e. $a_{i_1\ldots i_d} = a_{i_{\sigma(1)}\ldots i_{\sigma(d)}}$ for any permutation $\sigma \in \mathcal{S}_d$, we consider the polynomial

$$(1.7) \quad f_{d,A}(X) := f_{d,A}(X_1, \ldots, X_n) := \sum_{i_1,\ldots,i_d} a_{i_1\ldots i_d}(X_{i_1} - \mathbb{E} X_{i_1})\cdots(X_{i_d} - \mathbb{E} X_{i_d}).$$

Additionally, we often assume that A has vanishing generalized diagonal in the sense that $a_{i_1\ldots i_d} = 0$ whenever $i_1, \ldots, i_d$ are not pairwise different. In this situation, Theorem 1.2 reads as follows:

Corollary 1.3. *Let $X_1, \ldots, X_n$ be independent random variables with $\|X_i\|_{\Psi_\alpha} \leq M$ for some $\alpha \in (0,1] \cup \{2\}$ and let A be a symmetric d-tensor with vanishing generalized diagonal such that $\|A\|_{\mathrm{HS}} \leq 1$. Then*

$$\mathbb{E} \exp \left(\frac{C_{d,\alpha}}{M^\alpha} |f_{d,A}(X)|^{\frac{\alpha}{d}} \right) \leq 2.$$

As in Theorem 1.2, the conclusion is equivalent to a $\Psi_{\alpha/d}$-norm estimate.

1.1. **Main results.** In comparison to the aforementioned results, our main concentration inequalities provide more refined tail estimates. To this end, we need a family of tensor-product matrix norms $\|A\|_{\mathcal{J}}$ for a d-tensor A and a partition $\mathcal{J} \in P_{qd}$ of $\{1, \ldots, qd\}$. For the exact definitions, we refer to (3.5). Using these norms, we may formulate our first result for chaos-type functionals. Note that we focus on the case $\alpha = 2/q$ for some $q \in \mathbb{N}$ only, which is sufficient for many applications, like products or powers of sub-Gaussian or sub-exponential random variables. The general case $\alpha \in (0,1]$ will be treated later.

Theorem 1.4. *Let $X_1, \ldots, X_n$ be a set of independent random variables satisfying $\|X_i\|_{\psi_{2/q}} \leq M$ for some $q \in \mathbb{N}$ and $M > 0$, and let A be a symmetric d-tensor with vanishing diagonal. Consider $f_{d,A}(X)$ as in (1.7). Then, for any $t > 0$,*

$$(1.8) \qquad \mathbb{P}(|f_{d,A}(X)| \geq t) \leq 2 \exp \left(-\frac{1}{C_{d,q}} \min_{\mathcal{J} \in P_{qd}} \left(\frac{t}{M^d \|A\|_{\mathcal{J}}} \right)^{2/|\mathcal{J}|} \right).$$

To give an elementary example, consider the case $d = 1$ and $q = 2$. Here, $A = a = (a_1, \ldots, a_n)$ is a vector, and $f_{1,A}(X) = \sum_{i=1}^n a_i(X_i - \mathbb{E} X_i)$ is just a linear functional of random variables with sub-exponential tails ($\|X_i\|_{\psi_1} \leq M$). It easily follows from the definition that $\|A\|_{\{1,2\}} = |a|$ (i.e. the Euclidean norm of a) and $\|A\|_{\{\{1\},\{2\}\}} = \max_i |a_i|$. As a consequence, for any $t > 0$

$$\mathbb{P}\left(\left| \sum_{i=1}^n a_i(X_i - \mathbb{E} X_i) \right| \geq t \right) \leq 2 \exp \left(-\frac{1}{C} \min \left(\frac{t^2}{M^2 |a|^2}, \frac{t}{M \max_i |a_i|} \right) \right).$$

Hence, up to constants, we get back a classical result for the tails of a linear form in random variables with sub-exponential tails. For more general functions f and similar results under a Poincaré-type inequality, we refer to [BL97] (the first order case) and [GS18] (the higher order case).

Moreover, Theorem 1.4 can be used to give Hanson–Wright-type bounds for quadratic forms in sub-exponential random variables. Here we provide a sharpened version of Proposition 1.1. Let $\langle x, y \rangle$ be the standard scalar product in $\mathbb{R}^n$.

Proposition 1.5. *Let $q \in \mathbb{N}$, $A = (a_{ij})$ be a symmetric $n \times n$ matrix and let $X_1, \ldots, X_n$ be a set of independent, centered random variables with $\|X_i\|_{\Psi_{2/q}} \leq M$ and $\mathbb{E} X_i^2 = \sigma_i^2$. For any $t > 0$*

$$\mathbb{P}\left(\left| \sum_{i,j} a_{ij} X_i X_j - \sum_{i=1}^n \sigma_i^2 a_{ii} \right| \geq t \right) \leq 2 \exp \left(-\frac{1}{C} \eta(A, q, t/M^2) \right),$$

where

$$\eta(A, q, t) := \min \left(\frac{t^2}{\|A\|_{\mathrm{HS}}^2}, \frac{t}{\|A\|_{\mathrm{op}}}, \left(\frac{t}{\max_{i=1,\ldots,n} \|(a_{ij})_j\|_2} \right)^{\frac{2}{q+1}}, \left(\frac{t}{\|A\|_\infty} \right)^{\frac{1}{q}} \right).$$

Consequently, for any $x > 0$ we have with probability at least $1 - 2\exp(-x/C)$

$$|\langle X, AX\rangle - \mathbb{E}\langle X, AX\rangle| \le M^2 \max\left(\sqrt{x}\|A\|_{\mathrm{HS}}, x\|A\|_{\mathrm{op}}, x^{\frac{q+1}{2}} \max_{i=1,\ldots,n} \|(a_{ij})_j\|_2, x^q\|A\|_\infty\right).$$

It is possible to replace $2/q$ by a general $\alpha \in (0,1] \cup \{2\}$ (see Section 6). In this case, we have to replace $2/(q+1)$ by $2\alpha/(2+\alpha)$ and $1/q$ by $\alpha/2$.

Remark. Note that in comparison to the Hanson–Wright inequality (1.3) and Proposition 1.1, the more refined version contains two additional terms. The respective norms $\max_{i=1,\ldots,n}\|(a_{ij})_j\|_2$ and $\|A\|_\infty$ can no longer be written in terms of the eigenvalues of A (in contrast to $\|A\|_{\mathrm{HS}}$ and $\|A\|_{\mathrm{op}}$). Indeed, as we see later, we have $\max_{i=1,\ldots,n}\|(a_{ij})_j\|_2 = \|A\|_{2\to\infty}$, and $\|A\|_\infty = \max_{i,j}|\langle e_i, Ae_j\rangle|$ for the standard basis $(e_i)_i$ of $\mathbb{R}^n$. Moreover, the norms might have a very different scaling in n. For example, if $e = (1,\ldots,1)$ and $A = ee^T - \mathrm{Id}$, then $\|A\|_{\mathrm{HS}} \sim \|A\|_{\mathrm{op}} \sim n$, $\max_i\|(a_{ij})_j\|_2 \sim n^{1/2}$ and $\|A\|_\infty = 1$.

Finally, let us state the result for general polynomials in random variables with bounded Orlicz norms. To fix some notation, if $f\colon \mathbb{R}^n \to \mathbb{R}$ is a function in $\mathcal{C}^D(\mathbb{R}^n)$, for $d \le D$ we denote by $f^{(d)}$ the (symmetric) d-tensor of its d-th order partial derivatives.

Theorem 1.6. *Let $X_1, \ldots, X_n$ be a set of independent random variables satisfying $\|X_i\|_{\psi_{2/q}} \le M$ for some $q \in \mathbb{N}$ and $M > 0$. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a polynomial of total degree $D \in \mathbb{N}$. Then, for any $t > 0$,*

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \ge t) \le 2\exp\left(-\frac{1}{C_{D,q}} \min_{1\le d\le D} \min_{\mathcal{J}\in P_{qd}} \left(\frac{t}{M^d\|\mathbb{E}f^{(d)}(X)\|_{\mathcal{J}}}\right)^{\frac{2}{|\mathcal{J}|}}\right).$$

Note that if $f(X) = f_{D,A}(X)$ as in (1.7), only the D-th order tensor gives a contribution, i.e. we retrieve Theorem 1.4. We discuss Theorems 1.4 and 1.6 and compare them to known results in Subsection 1.2. A variant of Theorem 1.6 for polynomials in independent random variables with $\|X_i\|_{\psi_\alpha} \le 1$ for any $\alpha \in (0,1]$ with be derived in Section 6.

Remark. With the help of these inequalities, it is possible to prove many results on concentration of linear and quadratic forms in independent random variables scattered throughout the literature. For example, [NSU17, Lemma A.6] is an immediate consequence of Theorem 1.4 (combined with Lemma A.1 for $f(X, X') = \sum_{i=1}^n a_i X_i X_i'$). In a similar way, one can deduce [Yan+17, Lemma C.4] by applying Theorem 1.4 to the random variable $Z_i := X_i Y_i$, whenever (X_i, Y_i) is a vector with sub-exponential marginal distributions. More generally, one can consider a linear form (or higher order polynomial chaoses) in a product of k random variables $X_1, \ldots, X_k$ with sub-exponential tails, for which Lemma A.1 provides estimates for the $\Psi_{\frac{1}{k}}$ norm.

Lastly, the results in [EYY12, Appendix B] can be sharpened for $\alpha \in (0,1] \cup \{2\}$ by a more general version of Proposition 1.5, using the same arguments as in [RV13, Section 3] to treat complex-valued matrices.

1.2. **Related work.** Inequalities for the L^p-norms of polynomial chaos have been established in various works. From these L^p norm inequalities one can quite easily derive concentration inequalities. For a thorough discussion on inequalities involving linear forms in independent random variables we refer to [PG99, Chapter 1].

Starting with linear forms, there have been generalizations to certain classes of random variables as well as multilinear forms of higher degree (also called polynomial chaoses). Among these are the two classes of random variables with either log-convex or log-concave tails (i. e. $t \mapsto -\log \mathbb{P}(|X| \geq t)$ is convex respectively concave). Two-sided L^p norm estimates for the log-convex case were derived in [HMO97] for linear forms and in [KL15] for chaoses of all orders. On the other hand, for measures with log-concave tails similar two-sided estimates have been derived in [GK95; Lat96; Lat99; LŁ03; AL12] under different conditions. Moreover, two-sided estimates for non-negative random variables have been derived in [Mel16] and for chaos of order two in symmetric random variables satisfying the inequality $\|X\|_{2p} \leq A\|X\|_p$ in [Mel17].

Our approach is closer to the work of Adamczak and Wolff, [AW15], where the case of polynomials in sub-Gaussian random variables has been treated. Lastly, let us mention the two results [EYY12, Lemma B.2, Lemma B.3] and [VW15, Corollary 1.6], where concentration inequalities for quadratic forms in independent random variables with α-sub-exponential tails have been proven.

To be able to compare our results to the results listed above, let us discuss their conditions. Firstly, the conditions of a bounded Orlicz norm and log-convex or log-concave tails cannot be compared in general. It is known that random variables with log-convex tails satisfy $\|X\|_{\Psi_1} < \infty$. On the other hand, the tail function of any discrete random variable X is a step function (for example, if X has the geometric distribution, then $-\log \mathbb{P}(X \geq t) = \lfloor x \rfloor \log(1/(1-p)))$, which is neither log-convex nor log-concave but can still have a finite Ψ_α norm for some α. For example, a Poisson-distributed random variable X satisfies $\|X\|_{\Psi_1} < \infty$.

The condition $\|X\|_{2p} \leq \alpha\|X\|_p$ for all $p \geq 1$ and some $\alpha > 1$ used in the works of Meller implies the existence of the $\Psi_{\tilde{\alpha}}$-norm for $\tilde{\alpha} := (\log_2 \alpha)^{-1}$. Especially in the case $\alpha = 2^d$ this yields the existence of the $\Psi_{1/d}$ norm. However, we want to stress that the results in [AL12; KL15; Mel16; Mel17] are two-sided and require very different tools.

Moreover, the two works of Schudy and Sviridenko [SS11; SS12] contain concentration inequalities for polynomials in so-called *moment bounded random* variables. Therein, a random variable Z is called moment bounded with parameter $L > 0$, if for all $i \geq 1$ $\mathbb{E}|Z|^i \leq iL \mathbb{E}|Z|^{i-1}$. Actually, using Stirling's formula, it is easy to see that moment-boundedness implies $\|Z\|_{\Psi_1} < \infty$, but it is not clear whether the converse implication also holds. However, there is no inequality of the form $L \leq C\|X\|_{\Psi_1}$, as can be seen by $X \sim \mathrm{Ber}(p)$.

Considering quadratic forms in random variables X which are moment bounded and centered, one can easily see that (apart from the constants) the bound in Proposition 1.5 is sharper than the corresponding inequality in [SS12, Theorem 1.1]. Since for log-convex distributions there are two-sided estimates, Proposition 1.5 is sharp in this class. Apart from quadratic forms, due to the different conditions and quantities, it is difficult to compare [SS12] and Theorem 1.6 in general.

1.3. **Outline.** In Section 2 we formulate and prove several applications which can be deduced from the main results. Section 3 contains the proof for the concentration inequalities for multilinear forms (Theorem 1.4). Thereafter, we provide the proof of Proposition 1.5 in Section 4 and of Theorem 1.6 in Section 5. Section 6 is devoted to some extensions of the main results for random variables with finite Orlicz-norms for

any $\alpha \in (0,1]$. Lastly, we finish this note by collecting some elementary properties of the Orlicz-norms in the Appendix A.

2. APPLICATIONS

In the following, we provide some applications of our main results. In particular, all the results in this section follow from either Proposition 1.1 or 1.5. For any random variables $X_1, \ldots, X_n$ we write $X = (X_1, \ldots, X_n)$.

2.1. **Concentration of the Euclidean norm of a vector with independent components.**
As a start, Proposition 1.1 can be used to give concentration properties of the Euclidean norm of a linear transformation of X consisting of independent, normalized random variables with sub-exponential tails. We give two different forms thereof. The first form is inspired by the results in [RV13] for sub-Gaussian random variables.

Proposition 2.1. *Let $X_1, \ldots, X_n$ be independent random variables satisfying $\mathbb{E} X_i = 0, \mathbb{E} X_i^2 = 1, \|X_i\|_{\Psi_\alpha} \leq M$ for some $\alpha \in (0,1] \cup \{2\}$ and let $B \neq 0$ be an $m \times n$ matrix. For any $c > 0$ and any $t \geq c\|B\|_{\mathrm{HS}}$ we have*

$$(2.1) \qquad \mathbb{P}\left(\left| \|BX\|_2 - \|B\|_{\mathrm{HS}} \right| \geq t \right) \leq 2\exp\left(-\frac{\min(c^{2-\alpha}, 1)}{CM^4\|B\|_{\mathrm{op}}^\alpha} t^\alpha \right).$$

Note that in the case $\alpha = 2$ the constant is not present on the right hand side and thus we can choose any $t > 0$, which is exactly [RV13, Theorem 2.1]. In the general case, we need to restrict t to be of the order $\|B\|_{\mathrm{HS}}$.

The assumption of unit variance can be weakened, with some minor modifications, i.e. $\|B\|_{\mathrm{HS}}$ has to be replaced by $(\sum_{i=1}^n \sigma_i^2 \sum_{j=1}^n b_{ij}^2)^{1/2}$ and the constant C will depend on $\min_{i=1,\ldots,n} \sigma_i^2$. We omit the details.

Proof. First off, note that it suffices to prove the inequality for a matrix B such that $\|B\|_{\mathrm{HS}} = 1$ and $t \geq c$, since the general case follows by considering $\tilde{B} := B\|B\|_{\mathrm{HS}}^{-1}$.

Let us apply Proposition 1.1 to the matrix $A := B^T B$. An easy calculation shows that $\mathrm{trace}(A) = \mathrm{trace}(B^T B) = \|B\|_{\mathrm{HS}}^2 = 1$, so that we have

$$\mathbb{P}\left(\left| \|BX\|_2^2 - 1 \right| \geq t \right) \leq 2\exp\left(-\frac{1}{CM^4} \min\left(\frac{t^2}{\|B\|_{\mathrm{op}}^2}, \left(\frac{t}{\|B\|_{\mathrm{op}}^2} \right)^{\frac{\alpha}{2}} \right) \right)$$

$$= 2\exp\left(-\frac{1}{CM^4\|B\|_{\mathrm{op}}^\alpha} \min\left(\frac{t^{2-\alpha}}{\|B\|_{\mathrm{op}}^{2-\alpha}} t^\alpha, t^{\frac{\alpha}{2}} \right) \right)$$

$$(2.2)$$

$$\leq 2\exp\left(-\frac{\min(c^{2-\alpha} t^\alpha, t^{\frac{\alpha}{2}})}{CM^4\|B\|_{\mathrm{op}}^\alpha} \right)$$

$$\leq 2\exp\left(-\frac{\min(c^{2-\alpha}, 1)}{CM^4\|B\|_{\mathrm{op}}^\alpha} \min(t^\alpha, t^{\frac{\alpha}{2}}) \right).$$

Here, in the first step we have used the estimates $\|A\|_{\mathrm{HS}}^2 \leq \|B\|_{\mathrm{op}}^2\|B\|_{\mathrm{HS}}^2 = \|B\|_{\mathrm{op}}^2$ and $\|A\|_{\mathrm{op}} \leq \|B\|_{\mathrm{op}}^2$ as well as the fact that by Lemma A.2, $\mathbb{E} X_i^2 = 1$ for any i implies $M \geq C_\alpha > 0$. The second inequality follows from $t \geq c \geq c\|B\|_{\mathrm{op}}$ and the third inequality is a consequence of $\min(c^{2-\alpha} t^\alpha, t^{\frac{\alpha}{2}}) \geq \min(c^{2-\alpha}, 1)\min(t^\alpha, t^{\frac{\alpha}{2}})$.

Now, as in [RV13], we use the inequality $|z - 1| \leq \min(|z^2 - 1|, |z^2 - 1|^{1/2})$, giving for any $t > 0$

$$(2.3) \qquad \mathbb{P}\left(|\|BX\|_2 - 1| \geq t\right) \leq \mathbb{P}\left(|\|BX\|_2^2 - 1| \geq \max(t, t^2)\right).$$

Hence, a combination of (2.2), (2.3) and $\min(\max(r, r^2), \max(r^{1/2}, r)) = r$ yields for $t > c$

$$(2.4) \qquad \mathbb{P}\left(|\|BX\|_2 - 1| \geq t\right) \leq 2\exp\left(-\frac{\min(c^{2-\alpha}, 1)}{CM^4\|B\|_{\mathrm{op}}^\alpha} t^\alpha\right).$$

$\square$

The next corollary provides an alternative estimate for $\|BX\|_2$:

Corollary 2.2. *Let $X_1, \ldots, X_n$ be independent, centered random variables satisfying $\|X_i\|_{\Psi_1} \leq M$ and $\mathbb{E} X_i^2 = \sigma_i^2$. For an $n \times n$ matrix B with real entries let $A = B^T B = (a_{ij})$. Then, for any $x > 0$, with probability at least $1 - 2\exp(-x/C)$ we have*

$$\|BX\|_2^2 \leq \sum_{i=1}^n \sigma_i^2 \sum_{j=1}^n b_{ji}^2 + M^2 \max\left(\sqrt{x}\|A\|_{\mathrm{HS}}, x\|A\|_{\mathrm{op}}, x^{3/2} \max_{i=1,\ldots,n} \|(a_{ij})_j\|_2, x^2\|A\|_\infty\right).$$

Corollary 2.2 can be compared to various bounds on the norms of $\|BX\|_2$ in the case that X is a sub-Gaussian vector (see for example [HKZ12] or [Ada15]). For sub-Gaussian vectors with sub-Gaussian constant 1, we have with probability at least $1 - \exp(-x)$

$$\|BX\|_2^2 \leq \mathrm{trace}(B^T B) + 2\|B^T B\|_{\mathrm{HS}}\sqrt{x} + 2\|B^T B\|_{\mathrm{op}} x,$$

so that we have similar terms corresponding to $\sqrt{x}$ and x, whereas in the sub-exponential case we need two additional terms to account for the heavier tails of its components.

Proof. Define the quadratic form

$$Z := \|BX\|_2^2 = \langle BX, BX \rangle = \langle X, B^T B X \rangle = \langle X, AX \rangle.$$

Using Proposition 1.5 with the matrix A gives with probability $1 - 2\exp(-x/C)$

$$|Z - \mathbb{E} Z| \leq \max\left(\sqrt{x}\|A\|_{\mathrm{HS}}, x\|A\|_{\mathrm{op}}, x^{3/2} \max_{i=1,\ldots,n} \|A_{i\cdot}\|_2, x^2\|A\|_\infty\right).$$

From these inequalities and $|x| \geq x$ the claim easily follows by taking the square root. Note that $\mathbb{E} Z = \mathbb{E}\langle X, AX \rangle = \sum_{i=1}^n \sigma_i^2 \sum_{j=1}^n b_{ji}^2$. $\square$

2.2. **Projections of a random vector and distance to a fixed subspace.** It is possible to apply Proposition 1.5 to any matrix A associated to an orthogonal projection. In these cases, the norms can be explicitly calculated. Moreover, these norms do not depend on the structure of the subspace onto which one projects, but merely on its dimension. This leads to the following application, where we replace a fixed projection by a random one.

Corollary 2.3. *Let $X_1, \ldots, X_n$ be independent random variables satisfying $\mathbb{E} X_i = 0, \mathbb{E} X_i^2 = \sigma_i^2$ and $\|X_i\|_{\Psi_1} \leq M$. Furthermore, let $m < n$ and P be the (random) orthogonal projection onto an m-dimensional subspace of $\mathbb{R}^n$, distributed according*

to the Haar measure on the Grassmanian manifold $G_{m,n}$. For any $x > 0$, with probability at least $1 - 2\exp(-x/C)$, we have

$$(2.5) \qquad \left| \|PX\|_2^2 - \frac{m}{n} \sum_{i=1}^n \sigma_i^2 \right| \leq M^2 \max\left(\sqrt{xm}, x^2 \right).$$

Proof of Corollary 2.3. This is an application of Proposition 1.5. To see

$$\mathbb{E}\|PX\|_2^2 = \frac{m}{n} \sum_{i=1}^n \sigma_i^2,$$

we use [Ver18, Lemma 5.3.2] conditionally on X, i.e. we have

$$\mathbb{E}\|PX\|_2^2 = \mathbb{E}\,\mathbb{E}\left(\|PX\|_2^2 \mid X \right) = \frac{m}{n} \mathbb{E}\|X\|_2^2 = \frac{m}{n} \sum_{i=1}^n \sigma_i^2.$$

Moreover, for any projection P onto an m-dimensional subspace, one can see that $\|P\|_{\mathrm{HS}}^2 = \sum_{i=1}^n \lambda_i(P)^2 = m$. Moreover, it is clear that $\|P\|_\infty \leq \max_{i=1,\dots,n}\|(p_{ij})_j\|_2 \leq \|P\|_{2\to 2} = 1$. $\qquad\square$

A very similar result which follows from Proposition 2.1 is the following variant of [RV13, Corollary 3.1]. We use the notation $d(X, E) = \inf_{e \in E} d(X, e)$ for the distance between an element X and a subset E of a metric space (M, d).

Corollary 2.4. *Let $X_1, \dots, X_n$ be independent random variables satisfying $\mathbb{E}\,X_i = 0, \mathbb{E}\,X_i^2 = 1$ and $\|X_i\|_{\Psi_\alpha} \leq M$ for some $\alpha \in (0, 1] \cup \{2\}$, and let E be a subspace of $\mathbb{R}^n$ of dimension d. For any $t \geq \sqrt{n - d}$ we have*

$$\mathbb{P}\left(|d(X, E) - \sqrt{n - d}| \geq t \right) \leq 2\exp\left(-\frac{t^\alpha}{CM^4} \right).$$

Proof. This follows exactly as in [RV13, Corollary 3.1] by using Proposition 2.1. $\qquad\square$

2.3. **Spectral bound for a product of a fixed and a random matrix.** We can also extend the second application in [RV13] to any α-sub-exponential random vector as follows.

Proposition 2.5. *Let B be a fixed $m \times N$ matrix and let G be a $N \times n$ random matrix with independent entries satisfying $\mathbb{E}\,g_{ij} = 0, \mathbb{E}\,g_{ij}^2 = 1$ and $\|g_{ij}\|_{\Psi_\alpha} \leq M$ for some $\alpha \in (0, 1]$. For any $u, v \geq 1$ with probability at least $1 - 2\exp(-u^\alpha r(B)^\alpha - v^\alpha n)$ we have*

$$\|BG\|_{\mathrm{op}} \leq 4C_\alpha M^{4/\alpha} \left(u\|B\|_{\mathrm{HS}} + vn^{1/\alpha}\|B\|_{\mathrm{op}} \right).$$

Proof. We mimic the proof of [RV13, Theorem 3.2]. For any fixed $x \in S^{n-1}$ consider the linear operator $T : \mathbb{R}^{Nn} \to \mathbb{R}^m$ given by $T(G) = BGx$, and (by abuse of notation) write T for the matrix corresponding to this linear map in the standard basis. Using Proposition 2.1 applied to the matrix T we have

$$\mathbb{P}\left(\big| \|BGx\|_2 - \|T\|_{\mathrm{HS}} \big| \geq t \right) \leq 2\exp\left(-\frac{t^\alpha}{CM^4\|T\|_{\mathrm{op}}^\alpha} \right).$$

Now, since $\|T\|_{\mathrm{HS}} = \|B\|_{\mathrm{HS}}$ and $\|T\|_{\mathrm{op}} \leq \|B\|_{\mathrm{op}}$, this yields for any $t \geq \|B\|_{\mathrm{HS}}$

$$\mathbb{P}\left(\|BGx\|_2 > \|B\|_{\mathrm{HS}} + t \right) \leq 2\exp\left(-\frac{t^\alpha}{CM^4\|B\|_{\mathrm{op}}^\alpha} \right).$$

If we define $t = (2CM^4)^{1/\alpha}\Big(u\|B\|_{\mathrm{HS}} + (\log(5)+1)^{1/\alpha}vn^{1/\alpha}\|B\|_{\mathrm{op}}\Big)$ for arbitrary $u, v \geq 1$ and use the inequality $2(r+s)^\alpha \geq (r^\alpha + s^\alpha)$ valid for all $r, s \geq 0$, we obtain

$$\mathbb{P}\Big(\|BGx\|_2 > \|B\|_{\mathrm{HS}} + t\Big) \leq 2\exp\Big(-\frac{u^\alpha\|B\|_{\mathrm{HS}}^\alpha + v^\alpha n(\log(5)+1)\|B\|_{\mathrm{op}}^\alpha}{\|B\|_{\mathrm{op}}^\alpha}\Big)$$

$$\leq 2\exp\Big(-u^\alpha r(B)^\alpha - v^\alpha n - v^\alpha n\log(5)\Big)$$

$$\leq 5^{-n}2\exp\Big(-u^\alpha r(B)^\alpha - v^\alpha n\Big).$$

The last step is again a covering argument as in [RV13]. Choose a $1/2$-covering $\mathcal{N}$ (satisfying $|\mathcal{N}| \leq 5^n$, see [Ver12, Lemma 5.2]) of the unit sphere in $\mathbb{R}^n$, and note that a union bound gives

$$\mathbb{P}\Big(\bigcap_{x \in \mathcal{N}}\|BGx\|_2 \leq \|B\|_{\mathrm{HS}} + t\Big) \geq 1 - \sum_{x \in \mathcal{N}}\mathbb{P}\Big(\|BGx\|_2 > \|B\|_{\mathrm{HS}} + t\Big)$$

$$\geq 1 - 2\exp\Big(-u^\alpha r(B)^\alpha - v^\alpha n\Big).$$

Lemma 5.3 in [Ver12, Lemma 5.3] yields

$$\|BG\|_{\mathrm{op}} \leq 2\max_{x \in \mathcal{N}}\|BGx\|_2 \leq 2(\|B\|_{\mathrm{HS}} + t),$$

from which the assertion easily follows by upper bounding and simplifying the expression $2\|B\|_{\mathrm{HS}} + 2t$. $\qquad\square$

2.4. **Special cases.** It is possible to apply all results to random variables having a Poisson distribution, i.e. $X_i \sim \mathrm{Poi}(\lambda_i)$ for some $\lambda_i \in (0, \infty)$. By using the moment generating function of the Poisson distribution, it is easily seen that

$$\|X_i\|_{\Psi_1} = \frac{1}{\log\Big(\log(2)\lambda_i^{-1} + 1\Big)} =: g(\lambda_i).$$

The function g is increasing and satisfies $g(x) \sim \log(1/x)$ (for $x \to 0$) and $g(x) \sim x/\log(2)$ (for $x \to \infty$). More generally, if the random variable $|X|$ has a moment generating function $\phi_{|X|}$ in a neighborhood of 0, it can be used to explicitly calculate the Ψ_1-norm. Indeed, we have $\mathbb{E}\exp(|X|/t) = \phi_{|X|}(t^{-1})$, and so $\|X\|_{\Psi_1} = 1/\phi_{|X|}^{-1}(2)$.

Thus, as a special case of Proposition 1.5, we obtain the following corollary.

Corollary 2.6. *Let* $X_i \sim \mathrm{Poi}(\lambda_i)$, $B := g(\max_{i=1,\ldots,n}\lambda_i)$ *and* $A = (a_{ij})$ *be a symmetric* $n \times n$ *matrix. We have for any* $t > 0$

$$\mathbb{P}\Big(\Big|\sum_{i,j}a_{ij}X_iX_j - \sum_{i=1}^n a_{ii}\lambda_i\Big| \geq B^2 t\Big)$$

$$\leq 2\exp\Big(-\frac{1}{C}\min\Big(\frac{t^2}{\|A\|_{\mathrm{HS}}^2}, \frac{t}{\|A\|_{\mathrm{op}}}, \Big(\frac{t}{\max_i\|(a_{ij})_j\|_2}\Big)^{\frac{2}{3}}, \Big(\frac{t}{\|A\|_\infty}\Big)^{\frac{1}{2}}\Big)\Big)$$

$$\leq 2\exp\Big(-\frac{1}{C}\min\Big(\frac{t^2}{\|A\|_{\mathrm{HS}}^2}, \Big(\frac{t}{\|A\|_{\mathrm{op}}}\Big)^{\frac{1}{2}}\Big)\Big).$$

For Poisson chaos of arbitrary order $d \in \mathbb{N}$, one may derive similar results by evaluating Theorem 1.4 or Corollary 6.1 (both for $\alpha = 1$). Note though that already for $d = 1$, we lose a logarithmic factor in the exponent. However, we are not aware of any more refined fluctuation estimates for $d \geq 2$.

Another interesting example of a sub-exponential random variable arises in stochastic geometry. If $K \subseteq \mathbb{R}^n$ is an isotropic, convex body and X is distributed according to the cone measure on K, then $\|\langle X, \theta \rangle\|_{\Psi_1} \leq c$ for some constant c and any $\theta \in S^{n-1}$. For the details and the proof we refer to [PTT18, Lemma 5.1].

2.5. Concentration properties for fixed design linear regression. It is possible to extend the example of the fixed design linear regression in [HKZ12] to the situation of a sub-exponential noise (instead of sub-Gaussian).

To this end, let $y_1, \ldots, y_n \in \mathbb{R}^d$ be fixed vectors (commonly called the *design vectors*), $Y = (y_1, \ldots, y_n)$ (the $d \times n$ *design matrix*) and assume that the $d \times d$ matrix $\Sigma = n^{-1} \sum_{i=1}^n y_i y_i^T$ is invertible; in this case, define $B := n^{-1} \Sigma^{-1/2} Y \in M(d \times n)$. Let $X_1, \ldots, X_n$ be independent random variables with $\|X_i\|_{\Psi_1} \leq M$ and define

$$\beta := n^{-1} \sum_{i=1}^n \mathbb{E} X_i \Sigma^{-1} y_i$$

$$\hat{\beta}(X) := n^{-1} \sum_{i=1}^n X_i \Sigma^{-1} y_i.$$

β is the coefficient vector of the least expected squared error and $\hat{\beta}(X)$ is its ordinary least squares estimator (given the observation X). The quality of the estimator $\hat{\beta}$ can be judged by the excess loss

$$(2.6) \qquad R(X) = \|\Sigma^{1/2}(\hat{\beta}(X) - \beta)\|^2 = \sum_{i,j} a_{ij}(X_i - \mathbb{E} X_i)(X_j - \mathbb{E} X_j),$$

where $A = (a_{ij}) = B^T B = n^{-2} Y^T \Sigma^{-1} Y$, as can be shown by elementary calculations. Observe that this is a quadratic form in X_i with coefficients depending on the vectors y_i. Thus, Proposition 1.5 yields the following corollary.

Corollary 2.7. *In the above setting, for any $x > 0$ the inequality*

$$|R(X) - \mathbb{E} R(X)| \leq 4M^2 \max\left(\sqrt{x}\|A\|_{\mathrm{HS}}, x\|A\|_{\mathrm{op}}, x^{3/2} \max_{i=1,\ldots,n} \|(a_{ij})_j\|_2, x^2\|A\|_\infty\right)$$

holds with probability at least $1 - 2\exp(-x/C)$.

Thus the concentration properties of $R(X)$ around its mean depends on the four different norms of the matrix A. The factor 4 appears due to the necessary centering of the X_i.

2.6. Central limit theorems for quadratic forms and random edge weights.
In this section, our aim is to quantify central limit theorems for quadratic forms $Q(X) = Q_A(X) = \sum_{i,j} a_{ij} X_i X_j$ in sub-exponential random variables $X_1, \ldots, X_n$ using concentration of measure results. Typically, the first step is finding conditions such that $Q(X)$ can be approximated by a linear form $L(X)$. This reduces the problem to finding conditions such that $L(X)$ is asymptotically normal (e. g. using the Lyapunov central limit theorem).

The weak convergence of quadratic forms to a normal distribution is classical, and we refer to [Jon87] and [GT99], [Cha08] for general statements (and rates of convergence), as well as [PG99] for general statements on central limit theorems for U-statistics.

Let us first consider the task of approximating $Q(X)$ by a linear form $L(X)$. To this end, assume that A is symmetric with vanishing diagonal and $\mathbb{E} X_i \neq 0$ for some i. Then, we may decompose

$$Q(X) = \sum_{i,j} a_{ij}(X_i - \mathbb{E} X_i)(X_j - \mathbb{E} X_j) + 2\sum_{i=1}^{n}\Big(\sum_{j=1}^{n} a_{ij}\mathbb{E} X_j\Big)(X_i - \mathbb{E} X_i) + \mathbb{E} Q(X)$$

(this is in fact the Hoeffding decomposition of $Q(X)$), and we therefore define

$$L(X) := \sum_{i=1}^{n}\Big(\sum_{j=1}^{n} a_{ij}\mathbb{E} X_j\Big)(X_i - \mathbb{E} X_i) =: \sum_{i=1}^{n} c_{A,i}(X_i - \mathbb{E} X_i).$$

Obviously, $\mathrm{Var}(L(X)) = \sum_{i=1}^{n} c_{A,i}^2 \mathrm{Var}(X_i)$. Thus, under the condition

$$(2.7) \qquad \lim_{n\to\infty} \frac{\sum_{i=1}^{n} c_{A,i}^2 \mathrm{Var}(X_i)}{\|A\|_{\mathrm{HS}}^2} = \infty,$$

the asymptotic behavior of the properly normalized quadratic form is dominated by the linear term. Under additional assumptions of the tail behavior of the X_i, the approximation can also be quantified.

Lemma 2.8. *Let $X = (X_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables with $\|X_i\|_{\Psi_1} \leq M$ for a constant $M > 0$ and assume $\mathbb{E} X = (\mathbb{E} X_n)_{n\in\mathbb{N}} \neq 0$, $\mathrm{Var}(X) := (\mathrm{Var}(X_n))_{n\in\mathbb{N}} \neq 0$. Furthermore, let $A = A^{(n)}$ be a sequence of symmetric matrices satisfying (2.7). Then for any $t > 0$*

$$\mathbb{P}\left(|Q(X) - \mathbb{E} Q(X) - 2L(X)| \geq t\right) \leq 2\exp\left(-\min\left(\frac{t^2\mathrm{Var}(L(X))^2}{\|A\|_{\mathrm{HS}}^2}, \frac{t^{1/2}\mathrm{Var}(L(X))^{1/2}}{\|A\|_{\mathrm{HS}}^{1/2}}\right)\right).$$

Proof. Rewrite the Hoeffding decomposition of Q as

$$(2.8) \quad \langle A^{(n)}(X - \mathbb{E} X), X - \mathbb{E} X\rangle = \langle A^{(n)}X, X\rangle - \mathbb{E}\langle A^{(n)}X, X\rangle - 2\langle X, A\mathbb{E} X\rangle + 2\mathbb{E}\langle X, A\mathbb{E} X\rangle,$$

and recall $c_n = \mathrm{Var}(L(X))$. An application of Theorem 1.4 yields

$$(2.9) \quad \mathbb{P}\left(|\langle A^{(n)}(X - \mathbb{E} X), (X - \mathbb{E} X)\rangle| \geq c_n t\right) \leq 2\exp\left(-\frac{1}{C_d}\min\left(\Big(\frac{c_n t}{\|A\|_{\mathrm{HS}}}\Big)^2, \Big(\frac{c_n t}{\|A\|_{\mathrm{HS}}}\Big)^{1/2}\right)\right).$$

$\square$

In the case that the X_i are also identically distributed, (2.7) is equivalent to $\|A^{(n)}\|_{\mathrm{HS}}^2 = o(\sum_{i=1}^{n}(\sum_{j=1}^{n} A_{ij}^{(n)})^2)$. For example it is satisfied for $A^{(n)} = ee^T - \mathrm{Id}$, where $e = (1,\dots,1)^T \in \mathbb{R}^n$.

We may apply these results to sequences of graphs. Here we always assume that the X_i are identically distributed. For each n, let $G_n = (V_n, E_n)$ be some undirected graph on n nodes (which we may consider as a kind of "base graph"). If $A = A^{(n)}$ denotes its adjacency matrix, then (2.7) can be rewritten as

$$(2.10) \qquad \frac{\sum_{v\in V_n} \deg(v)^2}{2|E_n|} \to \infty.$$

Sequences of graphs satisfying (2.10) are the complete graph, the complete bipartite graph $G_n = K_{m_1(n),m_2(n)}$ for parameters $m_1(n), m_2(n)$ satisfying $m_1(n)+m_2(n) \to \infty$ and d_n-regular graphs for $d_n \to \infty$.

The example of the n-stars shows that (2.10) is not sufficient for a central limit theorem of the quadratic form. Indeed, in this case we have $Q(X) = X_1 \sum_{i=2}^{N} X_i$, where 1 is the vertex with degree $(n-1)$. As is easily seen, $Q(X) = 0$ on $\{X_1 = 0\}$, and thus if X are Bernoulli distributed, the distribution has an atom which does not vanish for $n \to \infty$.

Finally, let us provide an example of a sequence of graphs for which a central limit theorem can be shown by imposing additional conditions. Here we assume that the random variables X_i are non-negative. In this case, they can be used to define edge weights $w_n(X) : E_n \to \mathbb{R}_+$ by $w_n(\{i,j\})(X) = X_i X_j$. Also let $W_n(X) := \sum_{e \in E_n} w_n(e)(X)$ be the total edge weight. Note that $W_n(X) = \langle AX, X \rangle$ for the adjacency matrix A of G.

Proposition 2.9. *Let X be a non-negative random variable with $\|X\|_{\Psi_1} \leq M$ and $\mathbb{E}X = \lambda > 0, \mathrm{Var}(X) = \sigma^2 > 0$, and let $(X_n)_{n \in \mathbb{N}}$ be a sequence of independent copies. Consider a sequence $G_n = (V_n, E_n)$ of graphs with $V_n = \{1, \dots, n\}$ such that (2.10) and*

$$(2.11) \qquad \frac{\left(\sum_{v \in V_n} \deg(v)^3\right)^2}{\left(\sum_{v \in V_n} \deg(v)^2\right)^3} \to 0$$

hold. Then, for the total edge weight $W_n(X)$, we have

$$\frac{W_n(X) - \mathbb{E}W_n(X)}{2\lambda\sigma\left(\sum_{v \in V_n} \deg(v)^2\right)^{1/2}} \Rightarrow \mathcal{N}(0,1).$$

Note that $W_n(X)$ is neither a sum of independent random variables, nor can be it written as a sum of an m-dependent sequence, since $w(e)(X)$ and $w(f)(X)$ are dependent whenever $e \cap f \neq \emptyset$.

In the case that $X \sim \mathrm{Ber}(p)$, the quantity $W_n(X)$ has a nice interpretation. If we interpret $X_v = 0$ as a failed vertex in the "base graph" G_n, $W_n(X)$ is the number of edges in the subgraph that is induced by the (random) vertex set $\{v \in V_n : X_v = 1\}$.

Proof. Consider the linear approximation given in Lemma 2.8

$$L(X) := \left(\sum_i c_i(A^{(n)})^2\right)^{-1/2} \sum_{i=1}^{n} c_i(A^{(n)})X_i.$$

It is also easy to see that condition (2.11) implies Lyapunov's condition with $\delta = 1$. Consequently, by Lindeberg's central limit theorem

$$L(X) \Rightarrow \mathcal{N}(0,1).$$

The claim now easily follows by combining Lemma 2.8 and Slutsky's theorem. $\square$

It should be possible to extend the result to any sequence of random graphs satisfying (2.10) and (2.11) by conditioning. Moreover, with appropriately modified conditions, by a more refined analysis it is possible to vary the sub-exponential constant M with n. We omit the details.

3. The multilinear case: Proof of Theorem 1.4

To begin with, let us introduce some notation. Define $[n] := \{1, \ldots, n\}$, and let $\mathbf{i} = (i_1, \ldots, i_d) \in [n]^d$ be a multiindex. For any subset $C \subseteq [d]$ with cardinality $|C| > 1$, we may introduce the "generalized diagonal" of $[n]^d$ with respect to C by

$$(3.1) \qquad \{\mathbf{i} \in [n]^d : i_k = i_l \text{ for all } k, l \in C\}.$$

This notion of generalized diagonals naturally extends to d-tensors $A = (a_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$ (obviously, the generalized diagonal of A with respect to C is the set of coefficients $a_{\mathbf{i}}$ such that $\mathbf{i}$ lies on the generalized diagonal of $[n]^d$ with respect to C). If $d = 2$ and $C = \{1, 2\}$, this gives back the usual notion of the diagonal of an $n \times n$ matrix. Moreover, write

$$[n]^{\underline{d}} := \{\mathbf{i} \in [n]^d : i_1, \ldots, i_d \text{ are pairwise different}\}.$$

If A, B are d-tensors, we define $\langle A, B \rangle = \sum_{\mathbf{i} \in [n]^d} a_{\mathbf{i}} b_{\mathbf{i}}$. Given a set of d vectors $v^1, \ldots, v^d \in \mathbb{R}^n$, we write $v^1 \otimes \ldots \otimes v^d$ for the outer product

$$(v^1 \otimes \ldots \otimes v^d)_{i_1 \ldots i_d} := \prod_{j=1}^{d} v_{i_j}^j.$$

In fact, $v^1 \otimes \ldots \otimes v^d$ is a d-tensor. In particular, we may regard A as a multilinear form by setting $A(v^1, \ldots, v^d) := \langle A, v^1 \otimes \ldots \otimes v^d \rangle$ for any $v^1, \ldots, v^d \in \mathbb{R}^n$.

The latter idea may be generalized by noting that any partition $\mathcal{J} = \{J_1, \ldots, J_k\}$ of $[d]$ induces a partition of the space of d-tensors as follows. Identify the space of all d-tensors with $\mathbb{R}^{n^d}$ and decompose

$$(3.2) \qquad \mathbb{R}^{n^d} \cong \bigotimes_{i=1}^{k} \mathbb{R}^{n^{J_i}} \cong \bigotimes_{i=1}^{k} \bigotimes_{j \in J_i} \mathbb{R}^n.$$

For any $x = x^{(1)} \otimes \ldots \otimes x^{(k)}$, the identification with a d-tensor is given by $x_{J_1, \ldots, J_d} = \prod_{l=1}^{k} x_{J_{I_l}}^{(l)}$. For example, for $d = 4$ and $\mathcal{I} = \{\{1, 4\}, \{2, 3\}\}$ we have two matrices x, y and $x_{J_1, J_2, J_3, J_4} = x_{J_1 J_4} y_{J_2 J_3}$. Using this representation, any d-tensor A can be trivially identified with a linear functional on $\mathbb{R}^{n^d}$ via the standard scalar product, i. e.

$$Ax = A\left(x^{(1)} \otimes \ldots \otimes x^{(k)}\right) = \langle A, x^{(1)} \otimes \ldots \otimes x^{(k)} \rangle = \sum_{\mathbf{i} \in [n]^d} a_{\mathbf{i}} \prod_{l=1}^{d} x_{\mathbf{i}_{J_l}}^{(l)}.$$

These identifications give rise to a family of tensor-product matrix norms: for any partition $\mathcal{J} \in P_d$, define a norm on the space (3.2) by

$$\|x\|_{\mathcal{J}} := \|x^{(1)} \otimes \ldots \otimes x^{(k)}\|_{\mathcal{J}} := \max_{i=1,\ldots,k} \|x^{(i)}\|_2.$$

Now, we may define $\|A\|_{\mathcal{J}}$ as the the operator norm with respect to $\|\cdot\|_{\mathcal{J}}$:

$$(3.3) \qquad \|A\|_{\mathcal{J}} = \sup_{\|x\|_{\mathcal{J}} \le 1} |Ax|.$$

This family of tensor norms agrees with the definitions in [Lat06] and [AW15] (among others).

Next we extend these definitions to a family of norms $\|A\|_{\mathcal{J}}$ where A is a d-tensor but $\mathcal{J} \in P_{qd}$ for some $q \in \mathbb{N}$. To this end, we first embed A into the space of qd-tensors. Indeed, denote by $e_q(A)$ the qd-tensor given by
(3.4)
$$(e_q(A))_{\mathbf{i}} := \begin{cases} a_{i_1 i_{q+1} i_{2q+1} \ldots i_{(k-1)q+1}} & \text{if } i_{kq+j} = i_{kq+1} \ \forall k = 0, \ldots, d-1 \ \forall j = 2, \ldots, q \\ 0 & \text{else.} \end{cases}$$

In other words, we divide $\mathbf{i} \in [n]^{qd}$ into d consecutive blocks with q indices in each block $(i_1, \ldots, i_q), (i_{q+1}, \ldots, i_{2q}), \ldots$ and only consider such indices for which all elements of these blocks take the same value. In fact, this is an intersection of d "generalized diagonals". Now we set

(3.5) $$\|A\|_{\mathcal{J}} := \|e_q(A)\|_{\mathcal{J}}.$$

For $q = 1$, this definition trivially agrees with (3.3).

Remark 3.1. The norms (3.5) are monotone with respect to the underlying partition in the following sense. For any two partitions $\mathcal{I} = \{I_1, \ldots, I_\mu\}$ and $\mathcal{J} = \{J_1, \ldots, J_\nu\}$ of $[qd]$, we say that $\mathcal{I}$ is finer than $\mathcal{J}$ (and write $\mathcal{I} \preccurlyeq \mathcal{J}$) if for any $j = 1, \ldots, \mu$ there is a $k \in \{1, \ldots, \nu\}$ such that $I_j \subseteq J_k$. If $\mathcal{I} \preccurlyeq \mathcal{J}$, we have $\|A\|_{\mathcal{I}} \le \|A\|_{\mathcal{J}}$. In particular, we always have

(3.6) $$\|A\|_{\{\{1\},\ldots,\{qd\}\}} \le \|A\|_{\mathcal{J}} \le \|A\|_{\{1,\ldots,qd\}}.$$

In view of (3.6), the two "extreme" norms corresponding to the coarsest and the finest partition of $[qd]$ deserve special attention. Firstly, it is elementary that

(3.7) $$\|A\|_{\{1,\ldots,qd\}} = \|e_q(A)\|_{\mathrm{HS}} = \|A\|_{\mathrm{HS}} = \Big(\sum_{\mathbf{i} \in [n]^d} a_{\mathbf{i}}^2 \Big)^{1/2}.$$

Here, $\|\cdot\|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt norm. Secondly, we have

$$\|A\|_{\{\{1\},\ldots,\{qd\}\}} = \|e_q(A)\|_{\mathrm{op}} = \begin{cases} \|A\|_{\mathrm{op}} & q = 1 \\ \max_{i,j} |a_{ij}| & q \ge 2 \end{cases},$$

see Lemma 4.1.

To prove Theorem 1.4, we furthermore need some auxiliary results. The first one compares the moments of sums of random variables with finite Orlicz norms to moments of Gaussian polynomials and the second one provides the estimates for multilinear forms in Gaussian random variables.

Lemma 3.2 ([AW15], Lemma 5.4). *For any positive integer k and any $p \ge 2$, if $Y_1, \ldots, Y_n$ are independent symmetric random variables with $\|Y_i\|_{\psi_{2/k}} \le M$, then*

$$\Big\| \sum_{i=1}^n a_i Y_i \Big\|_p \le C_k M \Big\| \sum_{i=1}^n a_i g_{i1} \cdots g_{ik} \Big\|_p,$$

where g_{ij} are i.i.d. $\mathcal{N}(0,1)$ variables.

Theorem 3.3 ([Lat06], Theorem 1). *Let $A = (a_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$ be a d-tensor, and let $G_1, \ldots, G_d$ be i.i.d. standard Gaussian random variables in $\mathbb{R}^n$. Then, for every $p \ge 2$,*

$$C_d^{-1} \sum_{\mathcal{J} \in P_d} p^{|\mathcal{J}|/2} \|A\|_{\mathcal{J}} \le \|\langle A, G_1 \otimes \ldots \otimes G_d \rangle\|_p \le C_d \sum_{\mathcal{J} \in P_d} p^{|\mathcal{J}|/2} \|A\|_{\mathcal{J}}.$$

In the proof of Theorem 1.4, we actually show L^p-estimates for $f_{d,A}(X)$. The following proposition provides the link to concentration inequalities. It was originally proven by Adamczak in [Ada06] and [AW15], while at this point we cite it in the form given in [SS18], with a small modification to adjust the constant in front of the exponential.

Proposition 3.4. *Assume that a random variable Z satisfies for every $p \geq 2$*

$$\|Z - \mathbb{E}Z\|_p \leq \sum_{k=1}^{d} (C_k p)^{k/2}$$

for some constants $C_1, \ldots, C_d \geq 0$. Let $L := |\{l \colon C_l > 0\}|$ and $r := \min\{l \in \{1, \ldots, d\} \colon C_l > 0\}$. Then, for any $t > 0$

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq 2 \exp\left(-\frac{\log(2)}{2(Le)^{2/r}} \min_{k=1,\ldots,d} \left\{\frac{t^{2/k}}{C_k}\right\}\right).$$

Now we are able to prove Theorem 1.4.

Proof of Theorem 1.4. For simplicity, we always write $f(X) := f_{d,A}(X)$. Moreover, without loss of generality, we may assume the X_i to be centered.

Let $X^{(1)}, \ldots, X^{(d)}$ be independent copies of the random vector X. Take a set of i.i.d. Rademacher variables $(\varepsilon_i^{(j)})$, $i \leq n$, $j \leq d$, which are independent of the $(X^{(j)})_j$. By standard decoupling and symmetrization inequalities (see [PG99, Theorem 3.1.1] and [PG99, Lemma 1.2.6]),

$$\|f(X)\|_p \leq C_d \Big\| \sum_{\mathbf{i} \in [n]^{\underline{d}}} a_{i_1,\ldots,i_d} X_{i_1}^{(1)} \cdots X_{i_d}^{(d)} \Big\|_p \leq C_d \Big\| \sum_{\mathbf{i} \in [n]^{\underline{d}}} a_{i_1,\ldots,i_d} \varepsilon_{i_1}^{(1)} X_{i_1}^{(1)} \cdots \varepsilon_{i_d}^{(d)} X_{i_d}^{(d)} \Big\|_p.$$

An iteration of Lemma 3.2 together with $\|X_i\|_{\psi_{2/q}} \leq M$ hence leads to

$$\|f(X)\|_p \leq C_d M^d \Big\| \sum_{\mathbf{i} \in [n]^{\underline{d}}} a_{i_1,\ldots,i_d} (g_{i_1,1}^{(1)} \cdots g_{i_1,q}^{(1)}) \cdots (g_{i_d,1}^{(d)} \cdots g_{i_d,q}^{(d)}) \Big\|_p.$$

Here, $(g_{i,k}^{(j)})$ is an array of i.i.d. standard Gaussian random variables. Rewriting (recall (3.4)) and applying Theorem 3.3 yields

$$\|f(X)\|_p \leq C_d M^d \|\langle e_q(A), \otimes_{j=1}^{d} \otimes_{k=1}^{q} (g_{i,k}^{(j)})_{i \leq n}\rangle\|_p \leq C_d M^d \sum_{\mathcal{J} \in P_{qd}} p^{|\mathcal{J}|/2} \|A\|_{\mathcal{J}}.$$

The proof is now easily completed by applying Proposition 3.4. $\qquad\square$

4. Hanson–Wright-type inequality: Proof of Proposition 1.5

The main task in the proof of Proposition 1.5 is explicitly calculating the norms.

Lemma 4.1. *For any d-tensor A and $q \geq 2$ we have*

$$\|A\|_{\{\{1\},\ldots,\{qd\}\}} = \|A\|_\infty = \max_{i_1,\ldots,i_d} |a_{i_1,\ldots,i_d}|.$$

Proof. Write $\mathcal{J} = \{\{1\}, \ldots, \{qd\}\}$. We have

$$\|A\|_{\mathcal{J}} = \sup \left\{ \Big| \sum_{i_1 \ldots i_{qd}} (e_q(A))_{i_1,\ldots,i_{qd}} x_{i_1}^1 \cdots x_{i_{qd}}^{qd} \Big| : |x^j| \leq 1 \text{ for all } j = 1, \ldots, qd \right\}$$

$$= \sup \left\{ \Big| \sum_{i_1,\ldots,i_d} a_{i_1,\ldots,i_d} x_{i_1}^1 \cdots x_{i_1}^q x_{i_2}^{q+1} \cdots x_{i_1}^{2q} \cdots x_{i_d}^{(d-1)q+1} \cdots x_{i_d}^{qd} \Big| : |x^j| \leq 1 \right\}$$

$$\leq \|A\|_\infty \sup\Big\{ \sum_{i_1,\ldots,i_d} |x_{i_1}^1 x_{i_1}^2| \cdots |x_{i_d}^{(d-1)q+1} x_{i_d}^{(d-1)q+2}| : |x^j| \leq 1 \Big\}$$

$$\leq \|A\|_\infty.$$

In the third step, we have iteratively used that for x^j with $|x^j| \leq 1$ we also have $|x_i^j| \leq 1$, and applied the Cauchy–Schwarz inequality d times.

To obtain the lower bound, let $l_1, \ldots, l_d$ be the index which achieves the maximum. Let $x^1 = \ldots = x^q = \delta_{l_1}$, $x^{q+1} = \ldots = x^{2q} = \delta_{l_2}$ and so on, so that

$$\|A\|_{\mathcal{J}} \geq |a_{l_1 \cdots l_d}| = \|A\|_\infty.$$

$\square$

The following easy observation helps in calculating the norms $\|\cdot\|_{\mathcal{J}}$. For any partition $\mathcal{J} = \{J_1, \ldots, J_k\} \in P_{[qd]}$ we write $\tilde{\mathcal{J}} = \{\tilde{J}_1, \ldots, \tilde{J}_k\}$ for

$$(4.1) \qquad \tilde{J}_j = \{i \in \{1, \ldots, d\} : J_j \cap \{q(i-1)+1, \ldots, qi\} \neq \emptyset\}.$$

That is, the sets $\tilde{J}_j$ indicate which of the d q-blocks intersect J_j. Note that $\cup_j \tilde{J}_j = [d]$, but $\tilde{\mathcal{J}}$ need not be a partition of $[d]$. In fact, some sets I may even appear more than once (with a slight abuse of notation, we choose to keep the set notation in this case anyway). Note that Remark 3.1 extends from partitions to decompositions (all definitions remain valid, even in case of some sets appearing multiple times). Nevertheless, we have by definition

$$(4.2) \qquad \|A\|_{\mathcal{J}} = \|A\|_{\tilde{\mathcal{J}}} := \sup\Big\{ \sum_{i_1,\ldots,i_d} a_{i_1 \ldots i_d} \prod_{j=1}^{k} x_{\mathbf{i}_{\tilde{J}_j}}^{(j)} : \|x_{\mathbf{i}_{\tilde{J}_j}}^{(j)}\|_2 \leq 1 \Big\},$$

i.e. the norm does not depend on $\mathcal{J}$, but on its "projection" $\tilde{\mathcal{J}}$. We will use this observation in the next lemma to calculate the norms $\|A\|_{\mathcal{J}}$ for quadratic forms (i.e. $d = 2$) and any $q \geq 2$.

Lemma 4.2. *Let A be a symmetric matrix, $q \geq 2$ and $\mathcal{J}$ be a partition of $[2q]$.*

   (1) If $\tilde{\mathcal{J}}$ contains $\{1, 2\}$ two or more times, then $\|A\|_{\mathcal{J}} = \|A\|_\infty$.
   (2) If $\tilde{\mathcal{J}}$ contains $\{1, 2\}$ and $\{1\}$ and $\{2\}$, then $\|A\|_{\mathcal{J}} = \|A\|_\infty$.
   (3) If $\tilde{\mathcal{J}} = \{\{1, 2\}, \{1\}, \ldots, \{1\}\}$ or $\tilde{\mathcal{J}} = \{\{1, 2\}, \{2\}, \ldots, \{2\}\}$, then $\|A\|_{\mathcal{J}} = \max_i \|A_{i\cdot}\|_2$.
   (4) If $\tilde{\mathcal{J}}$ comprises l times $\{1\}$ and k times $\{2\}$ for $k \geq 2, l \geq 2$, then $\|A\|_{\mathcal{J}} = \|A\|_\infty$. On the other hand, if $l = 1, k \geq 2$ or $k = 1, l \geq 2$ we have $\|A\|_{\mathcal{J}} = \max_i \|A_{i\cdot}\|_2$.
   (5) If $\tilde{\mathcal{J}} = \{\{1\}, \{2\}\}$, then $\|A\|_{\mathcal{J}} = \|A\|_{\mathrm{op}}$.
   (6) We have $\|A\|_{\{[qd]\}} = \|A\|_{\mathrm{HS}}$.

Proof. To see (1), write $\tilde{\mathcal{J}} = \{\tilde{J}_1, \ldots, \tilde{J}_k\}$, use the triangle inequality and the fact that $\|x\|_\infty \leq \|x\|_{\mathrm{HS}}$ for any tensor x:

$$\|A\|_{\mathcal{J}} = \sup\Big\{ \sum_{i,j} a_{ij} \prod_{k=1}^{l} x_{\mathbf{i}_{\tilde{J}_k}}^{(k)} \Big\} \leq \|A\|_\infty \sup\Big\{ \sum_{i,j} |x_{ij}||y_{ij}| \Big\} \leq \|A\|_\infty,$$

where the supremum is taken over all unit vectors $x^{(k)}$. The lower bound follows from (3.6) and Lemma 4.1.

(2) follows immediately from $\tilde{\mathcal{J}} \preccurlyeq \{\{1,2\},\{1,2\}\}$.

(3) follows from the triangle and Cauchy–Schwarz inequality:

$$\|A\|_{\mathcal{J}} \leq \sup\Big\{\sum_i |\prod_{k=1}^l y_i^k| |\sum_j a_{ij}x_{ij}|\Big\} \leq \sup\Big\{\sum_i |\prod_{k=1}^l y_i^k| \|(a_{ij})_j\|_2 \|x_{i\cdot}\|_2\Big\}$$

$$\leq \max_i \|(a_{ij})_j\|_2 \sup\Big\{|\prod_{k=1}^l y_i^k| \|x_{i\cdot}\|_2\Big\} \leq \max_i \|(a_{ij})_j\|_2.$$

The lower bound is obtained by choosing $y^1,\ldots,y^l$ as a Dirac delta on the row for which $\max_i \|A_{i\cdot}\|$ is attained.

To see (4), note that the case $k \geq 2, l \geq 2$ is very similar to the second part. If $l = 1, k \geq 2$ or $k = 1, l \geq 2$, similar arguments as in the third part give for any $x, y^1, \ldots, y^l$ with norm at most one

$$|\sum_{i,j} a_{ij}x_i \prod_k y_j^k| \leq \sum_j |\prod_k y_j^k| |\sum_i a_{ij}x_i| \leq \sum_j |\prod_k y_j^k| \|(a_{ij})_j\|_2| \leq \max_i \|(a_{ij})_j\|_2.$$

The lower bound again follows by choosing suitable Dirac deltas.

(5) and (6) are obvious from the definitions. $\qquad\square$

Actually, we have the equality

$$\max_{i=1,\ldots,n} \|(a_{ij})_j\|_2 = \|A\|_{2\to\infty},$$

where $\|A\|_{p\to q} := \sup\big\{\|Ax\|_q : \|x\|_p \leq 1\big\}$. For the proof, see [CTP17, Proposition 6.1]. Especially this yields $\max_{i=1,\ldots,n} \|(a_{ij})_j\|_2 \leq \|A\|_{\mathrm{op}}$.

We are now ready to prove Proposition 1.5. Throughout the rest of this section, for a matrix A let us denote by A^{od} its off-diagonal and by A^{d} the diagonal part.

Proof of Proposition 1.5. Lemma 4.2 shows that we only need to consider the four norms $\|A\|_{\mathrm{HS}}, \|A\|_{\mathrm{op}}, \max_{i=1,\ldots,n} \|(a_{ij})_j\|_2$ and $\|A\|_\infty$. It is easy to see that $\|A\|_{\mathrm{HS}} \geq \|A\|_{\mathrm{op}} \geq \max_i \|(a_{ij})_j\|_2 \geq \|A\|_\infty$. Thus, we need to determine which partitions give rise to which norms.

The only partition producing the Hilbert–Schmidt norm is $\mathcal{J}_1 = \{[qd]\}$, with $|\mathcal{J}_1| = 1$. The operator norm appears for the decomposition $\mathcal{J}_2 = \{\{1,\ldots,q\},\{q+1,\ldots,2q\}\}$ with $|\mathcal{J}_2| = 2$. Moreover, it is easy to see that all partitions $\mathcal{J}_3$ of $[2q]$ giving rise to $\max_{i=1,\ldots,n} \|(a_{ij})_j\|_2$ satisfy $|\mathcal{J}_3| \in \{2,\ldots,q+1\}$. Finally, for all $k = 2,\ldots,2q$ there are partitions $\mathcal{J}_4$ such that $\|A\|_{\mathcal{J}_4} = \|A\|_\infty$.

Hence for a diagonal-free matrix A we have by simply plugging in the norms calculated in Lemmas 4.1 and 4.2 into Theorem 1.4

$$(4.3) \qquad \mathbb{P}\Big(\Big|\sum_{i,j} a_{ij}(X_iX_j - \mathbb{E}\,X_i\,\mathbb{E}\,X_j)\Big| \geq t\Big) \leq 2\exp\Big(-\frac{1}{C}\eta(A,q,t/M^2)\Big),$$

where

$$\eta(A,q,t) = \min\Big(\frac{t^2}{\|A\|_{\mathrm{HS}}^2}, \frac{t}{\|A\|_{\mathrm{op}}}, \min_{l=2,\ldots,q+1}\Big(\frac{t}{\max_i \|(a_{ij})_j\|_2}\Big)^{\frac{2}{l}}, \min_{l=2,\ldots,2q}\Big(\frac{t}{\|A\|_\infty}\Big)^{\frac{2}{l}}\Big)$$

$$= \min\Big(\frac{t^2}{\|A\|_{\mathrm{HS}}^2}, \frac{t}{\|A\|_{\mathrm{op}}}, \Big(\frac{t}{\max_i \|(a_{ij})_j\|_2}\Big)^{\frac{2}{q+1}}, \Big(\frac{t}{\|A\|_\infty}\Big)^{\frac{1}{q}}\Big).$$

In the last two terms, we can choose the largest l since we can assume that $\frac{t}{\|A\|_{\mathcal{J}}} \geq 1$ for any partition $\mathcal{J}$, as the minimum is achieved in $\frac{t^2}{\|A\|_{\mathrm{HS}}^2}$ otherwise.

For matrices with non-vanishing diagonal, we divide the quadratic form into an off-diagonal and a purely diagonal part, i.e.

$$\sum_{i,j} a_{ij} X_i X_j = \sum_{i,j} a_{ij}^{\mathrm{od}} X_i X_j + \sum_{i=1}^{n} a_{ii}^{\mathrm{d}} X_i^2.$$

For brevity, let us define $P(t) := \mathbb{P}\Big(\big| \sum_{i,j} a_{ij} X_i X_j - \sum_{i=1}^{n} \sigma_i^2 a_{ii} \big| \geq t \Big)$. Use the above decomposition and the subadditivity to obtain

$$P(t) \leq \mathbb{P}\Big(\big| \sum_{i,j} a_{ij}^{\mathrm{od}} X_i X_j \big| \geq t/2 \Big) + \mathbb{P}\Big(\big| \sum_{i=1}^{n} a_{ii}^{\mathrm{d}} (X_i^2 - \sigma_i^2) \big| \geq t/2 \Big) =: p_1(t) + p_2(t).$$

Equation (4.3) can be used to upper bound $p_1(t)$ as

$$(4.4) \qquad p_1(t) \leq 2 \exp\left(-\frac{1}{C_2} \eta(A^{od}, q, t/M^2) \right).$$

The diagonal term can be treated by applying Theorem 1.4 for $d = 1$, $q = 4$ and $a = (A_{ii})_{i=1,\dots,n}$. Moreover, it is easy to see that we have $\|a\|_{\{1,2,3,4\}} = \sum_i (a_{ii}^{\mathrm{d}})^2$ (cf. (3.7)) and $\|a\|_{\mathcal{J}} = \|A^{\mathrm{d}}\|_\infty$ for any other decomposition $\mathcal{J}$. Consequently,

$$(4.5) \quad p_2(t) \leq 2 \exp\left(-\frac{1}{C_1} \min\left(\frac{t^2}{\|A^{\mathrm{d}}\|_{\mathrm{HS}}^2}, \frac{t}{\|A^{\mathrm{d}}\|_\infty}, \left(\frac{t}{\|A^{\mathrm{d}}\|_\infty}\right)^{2/3}, \left(\frac{t}{\|A^{\mathrm{d}}\|_\infty}\right)^{1/2} \right) \right)$$

$$(4.6) \qquad = 2 \exp\left(-\frac{1}{C_1} \eta_{1,A^{\mathrm{d}}}(t) \right).$$

Thus, by combining (4.4) and (4.5) we have

$$P(t) \leq 4 \exp\left(-C \min(\eta(A^{\mathrm{od}}, q, t), \eta_{1,A^{\mathrm{d}}}(t) \right).$$

Now it remains to lower bound the minimum by grouping the terms according to the different powers of t. This gives

$$p(t) \leq 4 \exp\left(-\frac{1}{C} \tilde{\eta}(A, q, t/M) \right),$$

where

$$\tilde{\eta}(A, q, t) := \min\left(\frac{t^2}{\|A\|_{\mathrm{HS}}^2}, \frac{t}{\max(\|A^{\mathrm{od}}\|_{\mathrm{op}}, \|A^{\mathrm{d}}\|_\infty)}, \left(\frac{t}{\max_{i=1,\dots,n} \|(a_{ij})_j\|_2} \right)^{2/(q+1)}, \left(\frac{t}{\|A\|_\infty} \right)^{1/q} \right).$$

Lastly, from the characterization $\|A\|_{\mathrm{op}} := \sup_{x \in S^{n-1}} |\langle x, Ax \rangle|$ it can be easily seen that the inequalities $\|A^{\mathrm{d}}\|_\infty \leq \|A\|_{\mathrm{op}}$ and $\|A^{\mathrm{od}}\|_{\mathrm{op}} \leq 2\|A\|_{\mathrm{op}}$ hold, and the constant 4 can be changed to 2 by adjusting the constant in the exponent. $\qquad\square$

5. THE POLYNOMIAL CASE: PROOF OF THEOREM 1.6

Let us now treat the case of general polynomials $f(X)$ of total degree $D \in \mathbb{N}$. Before we start, we need to discuss some more properties of the norms $\|A\|_{\mathcal{J}}$. To this end, recall the Hadamard product of two d-tensors A, B given by $A \circ B := (a_{\mathbf{i}} b_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$ (pointwise multiplication). If we interpret a d-tensor as a function $[n]^d \to \mathbb{R}$, we may

define "indicator matrices" 1_C for a set $C \subseteq [n]^d$ by setting $1_C = (a_{\mathbf{i}})_{\mathbf{i}}$ with $a_{\mathbf{i}} = 1$ if $\mathbf{i} \in C$ and $a_{\mathbf{i}} = 0$ otherwise. If $|\mathcal{J}| > 1$, we do not have

$$\|A \circ 1_C\|_{\mathcal{J}} \leq \|A\|_{\mathcal{J}} \tag{5.1}$$

in general. However, [AW15, Lemma 5.2] shows a number of situations in which such an inequality does hold.

Lemma 5.1. *Let $A = (a_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$ be a d-tensor.*
  (1) *If $C = \{\mathbf{i} : i_{k_1} = j_1, \ldots, i_{k_l} = j_l\}$ for some $1 \leq k_1 < \ldots < k_l \leq d$ ("generalized row"), then (5.1) holds.*
  (2) *If $C = \{\mathbf{i} : i_k = i_l \ \forall k, l \in K\}$ for some $K \subset [d]$ ("generalized diagonal"), then (5.1) holds.*
  (3) *If $C_1, C_2 \subset [n]^d$ are such that (5.1) holds, then so is $C_1 \cap C_2$.*

There is a further situation in which a version of (5.1) holds. Indeed, for any partition $\mathcal{K} = \{K_1, \ldots, K_a\} \in P_d$ of $[d]$ we define

$$L(\mathcal{K}) = \{\mathbf{i} \in [n]^d : i_k = i_l \Leftrightarrow \exists j : k, l \in K_j\}. \tag{5.2}$$

That is, $L(\mathcal{K})$ is the set of those indices for which the partition into level sets is equal to $\mathcal{K}$.

Lemma 5.2. *Let $\mathcal{J} \in P_{qd}$, $\mathcal{K} \in P_d$ and A be a d-tensor. Then,*

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

Proof. This is a generalization of [AW15, Corollary 5.3] which corresponds to the case $q = 1$. First note that by definition,

$$\|A \circ 1_{L(\mathcal{K})}\|_{\mathcal{J}} = \|e_q(A \circ 1_{L(\mathcal{K})})\|_{\mathcal{J}} = \|e_q(A) \circ e_q(1_{L(\mathcal{K})})\|_{\mathcal{J}}.$$

Therefore, it suffices to prove that for any qd-tensor B,

$$\|B \circ e_q(1_{L(\mathcal{K})})\|_{\mathcal{J}} \leq 2^{|\mathcal{K}|(|\mathcal{K}|-1)/2} \|B\|_{\mathcal{J}}.$$

To see this, observe that $e_q(1_{L(\mathcal{K})})$ is the indicator matrix of a set C which can be written as an intersection of $|\mathcal{K}|$ generalized diagonals (with the cardinality of the underlying sets of indices in (3.1) always being an integer multiple of q) and $|\mathcal{K}|(|\mathcal{K}|-1)/2$ sets of the form $\{\mathbf{i} : i_{kq+1} \neq i_{lq+1}\}$ for $k < l$. Recall that

$$\|B \circ 1_{\{i_{kq+1} \neq i_{lq+1}\}}\|_{\mathcal{J}} = \|B - B \circ 1_{\{i_{kq+1} = i_{lq+1}\}}\|_{\mathcal{J}} \leq 2\|B\|_{\mathcal{J}},$$

using Lemma 5.1 (2) in the last step. As a consequence, the claim follows by applying Lemma 5.1 (2) again and a generalization of Lemma 5.1 (3). $\qquad \square$

Finally, it remains to note that [AW15, Lemma 5.1] can be generalized as follows.

Lemma 5.3. *Let A be a d-tensor, and let $v_1, \ldots, v_d \in \mathbb{R}^n$ be any vectors. Then, for any partition $\mathcal{J} \in P_{qd}$, $\|A \circ \otimes_{i=1}^d v_i\|_{\mathcal{J}} \leq \|A\|_{\mathcal{J}} \prod_{i=1}^d \|v_i\|_\infty$.*

Proof. Recall equations (4.1) and (4.2). We have

$$\|A \circ \otimes_{i=1}^d v_i\|_{\mathcal{J}} = \sup\left\{ \sum_{i_1,\ldots,i_{qd}} (e_q(A))_{i_1 \ldots i_{qd}} (e_q(\otimes_{i=1}^d v_i))_{i_1 \ldots i_{qd}} \prod_{j=1}^k x_{\mathbf{i}_{J_j}}^{(j)} : \|x_{\mathbf{i}_{J_j}}^{(j)}\|_2 \leq 1 \right\}$$

$$= \sup\left\{ \sum_{i_1,\ldots,i_d} a_{i_1 \ldots i_d} v_1^{i_1} \cdots v_d^{i_d} \prod_{j=1}^k x_{\mathbf{i}_{\tilde{J}_j}}^{(j)} : \|x_{\mathbf{i}_{\tilde{J}_j}}^{(j)}\|_2 \leq 1 \right\}$$

$$\leq \sup\Big\{ \sum_{i_1,\dots,i_d} a_{i_1\dots i_d} \prod_{j=1}^{k} x_{\mathbf{i}_{\tilde{J}_i}}^{(j)} : \|x_{\mathbf{i}_{\tilde{J}_j}}^{(j)}\|_2 \leq 1 \Big\} \prod_{i=1}^{d} \|v_i\|_\infty$$

$$= \sup\Big\{ \sum_{i_1,\dots,i_{qd}} (e_q(A))_{i_1\dots i_{qd}} \prod_{j=1}^{k} x_{\mathbf{i}_{J_j}}^{(j)} : \|x_{\mathbf{i}_{J_j}}^{(j)}\|_2 \leq 1 \Big\} \prod_{i=1}^{d} \|v_i\|_\infty$$

$$= \|A\|_{\mathcal{J}} \prod_{i=1}^{d} \|v_i\|_\infty.$$

To see the third step, for each v_l we choose a set $\mathcal{J}_j$ such that $l \in \mathcal{J}_j$ and then define vectors $\tilde{x}_{\mathbf{i}_{\tilde{J}_j}}^{(j)}$ by multiplying $x_{\mathbf{i}_{\tilde{J}_j}}^{(j)}$ by the components of the vectors v_l which were attributed to $\mathcal{J}_j$. In particular, this leads to $\|\tilde{x}_{\mathbf{i}_{\tilde{J}_j}}^{(j)}\|_2 \leq \prod_l \|v_l\|_\infty \|x_{\mathbf{i}_{\tilde{J}_j}}^{(j)}\|_2$, where the product is taken over all the vectors v_l which were attributed to $x_{\mathbf{i}_{\tilde{J}_j}}^{(j)}$. $\qquad \square$

Before we begin with the proof of the concentration results for general polynomials, let us give some definitions. Boldfaced letters will always represent a vector (mostly a multiindex with integer components), and for any vector $\mathbf{i}$ let $|\mathbf{i}| := \sum_j i_j$. For the sake of brevity we define

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

Given two vectors $\mathbf{i}, \mathbf{k}$ of equal size, we write $\mathbf{k} \leq \mathbf{l}$ if $k_j \leq l_j$ for all j, and $\mathbf{k} < \mathbf{l}$ if $\mathbf{k} \leq \mathbf{l}$ and there is at least one index such that $k_j < l_j$. Lastly, by $f \lesssim g$ we mean an inequality of the form $f \leq C_{D,q} g$.

Proof of Theorem 1.6. We assume $M = 1$. For the general case, given random variables $X_1,\dots,X_n$ with $\|X_i\|_{\Psi_{2/q}} \leq M$, define $Y_i := M^{-1} X_i$. The polynomial $f = f(X)$ can be written as a polynomial $\tilde{f} = \tilde{f}(Y)$ by appropriately modifying the coefficients, i.e. multiplying each monomial by M^r, where r is its total degree. Now it remains to see that $\partial_{i_1\dots i_j} \tilde{f}(Y) = M^j \partial_{i_1\dots i_j} f(X)$.

Step 1. First, we reduce the problem to generalizations of chaos-type functionals (1.7). Indeed, by sorting according to the total grade, f may be represented as

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

where the constants satisfy $c_{(i_1,k_1),\dots,(i_\nu,k_\nu)}^{(d)} = c_{(i_{\pi_1},k_{\pi_1}),\dots,(i_{\pi_\nu},k_{\pi_\nu})}^{(d)}$ for any permutation $\pi \in \mathcal{S}_\nu$. As in [AW15], by rearranging and making use of the independence of $X_1,\dots,X_n$, this leads to the estimate

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

where

$$a_{\mathbf{i}}^{\mathbf{k}} = \sum_{m=\nu}^{D} \sum_{\substack{k_{\nu+1},\dots,k_m>0 \\ k_1+\dots+k_m \leq D}} \sum_{\substack{i_{\nu+1},\dots,i_m \\ (i_1,\dots,i_m) \in [n]^{\underline{m}}}} \binom{m}{\nu} c_{(i_1,k_1),\dots,(i_m,k_m)}^{(k_1+\dots+k_m)} \prod_{\alpha=1}^{m} \mathbb{E}X_{i_\alpha}^{k_{i_\alpha}}.$$

Step 2. Note that $\|X_i^k\|_{\psi_{2/(qk)}} = \|X_i\|_{\psi_{2/q}}^k \leq 1$. Thus, slightly modifying the proof of Theorem 1.4 (in particular, also using Lemma 3.2 for the non-linear terms), we obtain the estimate

$$\|f(X) - \mathbb{E}f(X)\|_p \lesssim \sum_{d=1}^{D} \sum_{\nu=1}^{d} \sum_{\mathbf{k} \in I_{\nu,d}} \Big\| \sum_{\mathbf{i} \in [n]^{\underline{\nu}}} a_{\mathbf{i}}^{\mathbf{k}} (g_{i_1,1}^{(1)} \cdots g_{i_1,qk_1}^{(1)}) \cdots (g_{i_\nu,1}^{(\nu)} \cdots g_{i_\nu,qk_\nu}^{(\nu)}) \Big\|_p.$$

Here, $(g_{i,k}^{(j)})$ is an array of i.i.d. standard Gaussian random variables.

Moreover, the family $(a_{\mathbf{i}}^{\mathbf{k}})_{\nu \in \{1,\ldots,d\}, k \in I_{\nu,d}, i \in [n]^{\underline{\nu}}}$ gives rise to a d-tensor A_d as follows. Given any index $\mathbf{i} = (i_1, \ldots, i_d)$ there is a unique number $r \in \{1, \ldots, d\}$ of distinct elements $j_1, \ldots, j_r$ with each j_l appearing exactly k_l times in $\mathbf{i}$. Consequently, we set $a_{i_1 \ldots i_d} := a_{j_1,\ldots,j_r}^{(l_1,\ldots,l_r)}$, and $A_d = (a_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$. Note that this is well-defined due to the symmetry assumption.

For any $\mathbf{k} \in I_{\nu,d}$ denote by $\mathcal{K}(\mathbf{k}) = \mathcal{K}(k_1, \ldots, k_\nu) \in P_d$ the partition which is defined by splitting the set $\{1, \ldots, d\}$ into consecutive intervals of length $k_1, \ldots, k_\nu$. In other words, $\mathcal{K}(\mathbf{k}) = \{K_1, \ldots, K_\nu\}$ with $K_l = \{\sum_{i=1}^{l-1} k_i + 1, \sum_{i=1}^{l-1} k_i + 2, \ldots, \sum_{i=1}^{l} k_i\}$, $l = 1, \ldots, \nu$. Now, recalling the definitions of e_q (3.4) and of $L(\mathcal{K})$ (5.2), by rewriting and applying Lemma 5.1 we obtain

$$\|f(X) - \mathbb{E}f(X)\|_p \lesssim \sum_{d=1}^{D} \sum_{\nu=1}^{d} \sum_{\mathbf{k} \in I_{\nu,d}} \|\langle e_q(A_d \circ 1_{L(\mathcal{K}(\mathbf{k}))}), \otimes_{j=1}^{\nu} \otimes_{k=1}^{qk_j} (g_{i,k}^{(j)})_{i \leq n} \rangle\|_p$$

(5.3)
$$\lesssim \sum_{d=1}^{D} \sum_{\nu=1}^{d} \sum_{\mathbf{k} \in I_{\nu,d}} \sum_{\mathcal{J} \in P_{qd}} p^{|\mathcal{J}|/2} \|A_d \circ 1_{L(\mathcal{K}(k_1,\ldots,k_\nu))}\|_{\mathcal{J}}$$

$$\lesssim \sum_{d=1}^{D} \sum_{\mathcal{J} \in P_{qd}} p^{|\mathcal{J}|/2} \|A_d\|_{\mathcal{J}}.$$

Step 3. Next, we replace $\|A_d\|_{\mathcal{J}}$ by $\|\mathbb{E}f^{(d)}(X)\|_{\mathcal{J}}$. To this end, first note that for $\mathbf{i} \in [n]^d$ with distinct indices $j_1, \ldots, j_\nu$ which are taken $l_1, \ldots, l_\nu$ times, we have

$$\mathbb{E}\frac{\partial^d f}{\partial x_{i_1} \ldots \partial x_{i_d}}(X) = \sum_{\mathbf{k}:\mathbf{k} \geq \mathbf{l}} \sum_{m=\nu}^{D} \sum_{\substack{k_{\nu+1},\ldots,k_m > 0 \\ k_1 + \ldots + k_m \leq D}} \sum_{\substack{j_{\nu+1},\ldots,j_m \\ (j_1,\ldots,j_m) \in [n]^{\underline{m}}}}$$

$$\left(\binom{m}{\nu} \nu! c_{(j_1,k_1),\ldots,(j_m,k_m)}^{(k_1+\ldots+k_m)} \prod_{\alpha=1}^{\nu} \mathbb{E}X_{j_\alpha}^{k_\alpha - l_\alpha} \prod_{\alpha=\nu+1}^{m} \mathbb{E}X_{j_\alpha}^{k_\alpha} \prod_{\alpha=1}^{\nu} \frac{k_\alpha!}{(k_\alpha - l_\alpha)!} \right)$$

$$= \nu! l_1! \cdots l_\nu! a_{i_1,\ldots,i_d} + R_{\mathbf{i}}^{(d)},$$

where the "remainder term" $R_{\mathbf{i}}^{(d)}$ corresponds to the set of indices $\mathbf{k}$ satisfying $\mathbf{k} > \mathbf{l}$. If $d = D$, we clearly have $R_{\mathbf{i}}^{(d)} = 0$, and therefore

(5.4)
$$\mathbb{E}\frac{\partial^D f}{\partial x_{i_1} \ldots \partial x_{i_D}}(X) = \nu! l_1! \cdots l_\nu! a_{i_1 \ldots i_D} = \nu! |I_1|! \cdots |I_\nu|! a_{i_1 \ldots i_D},$$

where $\mathcal{I} = \{I_1, \ldots, I_\nu\}$ is the partition given by the level sets of the index $\mathbf{i}$. It follows that for any partition $\mathcal{J} \in P_{qD}$,

$$\|A_D\|_{\mathcal{J}} \leq \sum_{\mathcal{K} \in P_D} \|A_D \circ 1_{L(\mathcal{K})}\|_{\mathcal{J}} \leq \sum_{\mathcal{K} \in P_D} \|\mathbb{E}f^{(D)}(X) \circ 1_{L(\mathcal{K})}\|_{\mathcal{J}} \lesssim \|\mathbb{E}f^{(D)}(X)\|_{\mathcal{J}},$$

using the partition of unity $1 = \sum_{\mathcal{K} \in P_D} 1_{L(\mathcal{K})}$ and the triangle inequality in the first, equation (5.4) in the second and Lemma 5.2 in the last step.

The proof is now completed by induction. More precisely, in the next step will show that for any $d \in \{1, \ldots, D-1\}$ and any partitions $\mathcal{I} = \{I_1, \ldots, I_\mu\} \in P_d$, $\mathcal{J} = \{J_1, \ldots, J_\nu\} \in P_{qd}$,

$$(5.5) \qquad \|R^{(d)} \circ 1_{L(\mathcal{I})}\|_{\mathcal{J}} \lesssim \sum_{k=d+1}^{D} \sum_{\substack{\mathcal{K} \in P_{qk} \\ |\mathcal{K}| = |\mathcal{J}|}} \|A_k\|_{\mathcal{K}}.$$

Having (5.5) at hand, it follows by reverse induction and Lemma 5.2 that

$$\sum_{d=1}^{D} \sum_{\mathcal{J} \in P_{qd}} p^{|\mathcal{J}|/2} \|A_d\|_{\mathcal{J}} \lesssim \sum_{d=1}^{D} \sum_{\mathcal{J} \in P_{qd}} p^{|\mathcal{J}|/2} \|\mathbb{E} f^{(d)}(X)\|_{\mathcal{J}}.$$

Plugging this into (5.3) and applying Proposition 3.4 finishes the proof.

Step 4: To show (5.5), let us analyze the "remainder tensors" $R^{(d)}$ in more detail. To this end, fix $d \in \{1, \ldots, D-1\}$ and partitions $\mathcal{I} = \{I_1, \ldots, I_\nu\} \in P_d$, $\mathcal{J} = \{J_1, \ldots, J_\mu\} \in P_{qd}$, and let $\mathbf{l}$ be the vector with $l_\alpha := |I_\alpha|$ (note that this implies $|\mathbf{l}| = d$). For any $\mathbf{k} \in I_{\nu, \leq D}$ with $\mathbf{k} > \mathbf{l}$, we define a d-tensor $S_{\mathcal{I}}^{(d,\mathbf{k})} = (s_{\mathbf{i}}^{(d,k_1,\ldots,k_\nu)})_{\mathbf{i} \in [n]^d} = (s_{\mathbf{i}}^{(d)})_{\mathbf{i} \in [n]^d}$ as follows:

$$s_{\mathbf{i}}^{(d)} = 1_{\mathbf{i} \in L(\mathcal{I})} \sum_{\substack{m=\nu \\ k_{\nu+1}, \ldots, k_m > 0 \\ k_1 + \ldots + k_m \leq D}}^{D} \sum_{\substack{j_{\nu+1}, \ldots, j_m \\ (j_1, \ldots, j_m) \in [n]^{\underline{m}}}} \binom{m}{\nu} c_{(j_1,k_1),\ldots,(j_m,k_m)}^{(k_1+\ldots+k_m)} \prod_{\alpha=1}^{\nu} \mathbb{E} X_{j_\alpha}^{k_\alpha - l_\alpha} \prod_{\alpha=\nu+1}^{m} \mathbb{E} X_{j_\alpha}^{k_\alpha}$$

Here, we denote by j_α the value of $\mathbf{i}$ on the level set I_α. Clearly,

$$R^{(d)} \circ 1_{L(\mathcal{I})} = \sum_{\substack{\mathbf{k} \in I_{\nu, \leq D} \\ \mathbf{k} > \mathbf{l}}} \nu! \frac{k_1}{(k_1 - l_1)!} \cdots \frac{k_\nu}{(k_\nu - l_\nu)!} S_{\mathcal{I}}^{(d,\mathbf{k})}.$$

Therefore, it remains to prove that there is a partition $\mathcal{K} \in P_{q|\mathbf{k}|}$ with $|\mathcal{K}| = |\mathcal{J}|$ such that

$$(5.6) \qquad \|S_{\mathcal{I}}^{(d,\mathbf{k})}\|_{\mathcal{J}} \lesssim \|A_{|\mathbf{k}|}\|_{\mathcal{K}}.$$

The tensor will be given by an appropriate embedding of the d-tensor $S_{\mathcal{I}}^{(d,\mathbf{k})}$. To this end, choose any partition $\tilde{\mathcal{I}} = \{\tilde{I}_1, \ldots, \tilde{I}_\nu\} \in P_{|\mathbf{k}|}$ with $|\tilde{I}_\alpha| = k_\alpha$ and $I_\alpha \subset \tilde{I}_\alpha$ for all α. Embedding the d-tensor $S_{\mathcal{I}}^{(d,\mathbf{k})}$ into the space of $|\mathbf{k}|$-tensors is done by defining a new tensor $\tilde{S}^{|\mathbf{k}|} = (\tilde{s}_{\mathbf{i}}^{|\mathbf{k}|})_{\mathbf{i}}$ given by

$$(5.7) \qquad \tilde{s}_{\mathbf{i}}^{|\mathbf{k}|} = s_{\mathbf{i}_{[d]}}^{(d)} 1_{\mathbf{i} \in L(\tilde{\mathcal{I}})}.$$

We choose the partition $\mathcal{K} = \{K_1, \ldots, K_\mu\}$ defined in the following way: for any j, we have $J_j \subset K_j$, so that it remains to assign the elements $r \in \{qd+1, \ldots, q|\mathbf{k}|\}$ to the sets K_j. Write $r = \eta q + m$ for some $\eta \in \{d, \ldots, |\mathbf{k}| - 1\}$ and $m \in \{1, \ldots, q\}$. Since $\tilde{\mathcal{I}}$ is a partition of $|\mathbf{k}|$, there is a unique $j \in \{1, \ldots, \nu\}$ such that $\eta + 1 \in \tilde{I}_j$. Take the smallest element t in $\tilde{I}_j$ – since $I_j \subset \tilde{I}_j$, we have $t \in [d]$ – and add r to the same set as $\pi(r) := (t-1)q + m$.

We claim that

$$(5.8) \qquad \|S_{\mathcal{I}}^{(d,|\mathbf{k}|)}\|_{\mathcal{J}} \leq \|\tilde{S}^{|\mathbf{k}|}\|_{\mathcal{K}}.$$

To see this, let $x^{(\beta)} = (x^{(\beta)}_{\mathbf{i}_{J_\beta}})$, $\beta = 1, \ldots, \mu$, be a collection of vectors satisfying $\|x^{(\beta)}\|_2 \leq 1$. This gives rise to a further collection of unit vectors $y^{(\beta)} = (y^{(\beta)}_{\mathbf{i}_{K_\beta}})$, $\beta = 1, \ldots, \mu$, defined by

$$y^{(\beta)}_{\mathbf{i}_{K_\beta}} = x^{(\beta)}_{\mathbf{i}_{K_\beta \cap [qd]}} \prod_{r \in K_\beta \setminus [qd]} 1_{i_r = i_{\pi(r)}}$$

(recall the definition of $\pi(r)$ given in the paragraph above). Now, it follows that

$$\sum_{|\mathbf{i}_{[d]}| \leq n} s^{(d)}_{\mathbf{i}_{[d]}} \prod_{\beta=1}^{\mu} x^{(\beta)}_{(e_q(\mathbf{i}))_{J_\beta}} = \sum_{|\mathbf{i}_{[|\mathbf{k}|]}| \leq n} \tilde{s}^{|\mathbf{k}|}_{\mathbf{i}_{|\mathbf{k}|}} \prod_{\beta=1}^{\mu} x^{(\beta)}_{(e_q(\mathbf{i}))_{J_\beta}} = \sum_{|\mathbf{i}_{[|\mathbf{k}|]}| \leq n} \tilde{s}^{(|\mathbf{k}|)}_{\mathbf{i}_{[|\mathbf{k}|]}} \prod_{\beta=1}^{\mu} y^{(\beta)}_{(e_q(\mathbf{i}))_{K_\beta}}.$$

These equations follow from the definition of the matrix $\tilde{S}^{|\mathbf{k}|}$ and the fact that if $\mathbf{i} \in e_q(L(\tilde{\mathcal{I}}))$, then for $r > qd$, $i_r = i_{\pi(r)}$, which implies $y^{(\beta)}_{\mathbf{i}_{K_\beta}} = x^{(\beta)}_{\mathbf{i}_{K_\beta \cap [qd]}} = x^{(\beta)}_{\mathbf{i}_{J_\beta}}$. As this holds true for any collection $x^{(\beta)}$, we obtain (5.8).

Finally, we prove

(5.9) $$\|\tilde{S}^{|\mathbf{k}|}\|_{\mathcal{K}} \lesssim \|A_{|\mathbf{k}|}\|_{\mathcal{K}}$$

for any partition $\mathcal{K} \in P_{q|\mathbf{k}|}$. To see this, note that if $\mathbf{i} \in L(\tilde{\mathcal{I}})$, we have $\tilde{s}^{|\mathbf{k}|}_{\mathbf{i}} = a^{|\mathbf{k}|}_{\mathbf{i}} \prod_{\alpha=1}^{\nu} \mathbb{E} X^{k_\alpha - l_\alpha}_{i_\alpha}$. As a consequence,

$$\tilde{S}^{|\mathbf{k}|} = (A_{|\mathbf{k}|} \circ 1_{L(\tilde{\mathcal{I}})}) \circ \otimes_{\alpha=1}^{|\mathbf{k}|} v_\alpha,$$

where the vectors v_α are defined by $v_\alpha = (\mathbb{E} X^{k_\alpha - l_\alpha}_i)_{i \leq n}$ if $\alpha \in \{\min I_1, \ldots, \min I_\nu\}$ and $v_\alpha = (1, \ldots, 1)$, otherwise. In particular, we always have $\|v_\alpha\|_\infty \lesssim 1$, and therefore, by Lemma 5.3,

$$\|\tilde{S}^{|\mathbf{k}|}\|_{\mathcal{K}} \lesssim \|A_{|\mathbf{k}|} \circ 1_{L(\tilde{\mathcal{I}})}\|_{\mathcal{K}},$$

from where we easily arrive at (5.9) by applying Lemma 5.2.

Now, (5.6) follows by combining (5.8) and (5.9), which finishes the proof. $\square$

6. The general sub-exponential case: $\alpha \in (0, 1]$

Using slightly different techniques than in the proofs of Theorem 1.4 and Theorem 1.6, we may obtain concentration results for polynomials in independent random variables with bounded ψ_α-norms for any $\alpha \in (0, 1]$. Here, the key difference is that we will not compare their moments to products of Gaussians but to Weibull variables.

To this end, we need some more notation. Let $A = (a_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$ be a d-tensor and $I \subset [d]$ a set of indices. Then, for any $\mathbf{i}_I := (i_j)_{j \in I}$, we denote by $A_{\mathbf{i}_{I^c}} = (a_{\mathbf{i}})_{\mathbf{i}_{I^c}}$ the $(d - |I|)$-tensor defined by fixing i_j, $j \in I$. For instance, if $d = 4$, $I = \{1, 3\}$ and $i_1 = 1, i_3 = 2$, then $A_{\mathbf{i}_{I^c}} = (a_{1j2k})_{jk}$. We will also need the notation $P(I^c)$ for the set of all partitions of I^c.

For $I = [d]$, i.e. we fix all indices of $\mathbf{i}$, we interpret $A_{\mathbf{i}_{I^c}} = a_{\mathbf{i}}$ as the $\mathbf{i}$-th entry of A. Moreover, in this case, we assume that there is a single element $\mathcal{J} \in P(I^c)$ (which we may call the "empty" partition), and $\|A_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} = |a_{\mathbf{i}}|$ is just the Euclidean norm of $a_{\mathbf{i}}$. Finally, note that if $I = \emptyset$, $\mathbf{i}_I$ does not indicate any specification, and $A_{\mathbf{i}_{I^c}} = A$.

Using the characterization of the Ψ_α norms in terms of the growth of L^p norms (see Appendix A for details), [KL15, Corollary 2] now yields a result similar to Theorem 1.4 for all $\alpha \in (0, 1]$:

Corollary 6.1. *Let $X_1, \ldots, X_n$ be a set of independent, centered random variables with $\|X\|_{\Psi_\alpha} \leq M$ for some $\alpha \in (0, 1]$, A be a symmetric d-tensor with vanishing diagonal and consider $f_{d,A}$ as in (1.7). We have for any $t > 0$*

$$\mathbb{P}\Big(|f_{d,A}(X)| \geq t\Big) \leq 2 \exp\left(-\frac{1}{C_{d,\alpha}} \min_{I \subset [d]} \min_{\mathcal{J} \in P(I^c)} \Big(\frac{t}{M^d \max_{\mathbf{i}_I} \|A_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}}\Big)^{\frac{2\alpha}{2|I| + \alpha|\mathcal{J}|}}\right).$$

The main goal of this section is to generalize Corollary 6.1 to arbitrary polynomials similarly to Theorem 1.6. This yields the following result:

Theorem 6.2. *Let $X_1, \ldots, X_n$ be a set of independent random variables satisfying $\|X_i\|_{\psi_\alpha} \leq M$ for some $\alpha \in (0, 1]$ and $M > 0$. Let $f \colon \mathbb{R}^n \to \mathbb{R}$ a polynomial of total degree $D \in \mathbb{N}$. Then, for any $t > 0$,*

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq t) \leq 2 \exp\left(-\frac{1}{C_{D,\alpha}} \min_{1 \leq d \leq D} \min_{I \subset [d]} \min_{\mathcal{J} \in P(I^c)} \Big(\frac{t}{M^d \max_{\mathbf{i}_I} \|(\mathbb{E}f^{(d)}(X))_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}}\Big)^{\frac{2\alpha}{2|I| + \alpha|\mathcal{J}|}}\right).$$

To prove Theorem 6.2, note that one particular example of centered random variables with $\|X\|_{\Psi_\alpha} \leq M$ is given by symmetric Weibull variables with shape parameter α (and scale parameter 1), i.e. symmetric random variables w with $\mathbb{P}(|w| \geq t) = \exp(-t^\alpha)$. In fact, [KL15, Example 3] especially implies the following analogue of of Lemma 3.3:

Lemma 6.3. *Let $A = (a_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$ be a d-tensor and (w_i^j), $i \leq n$, $j \leq d$, an array of i.i.d. Weibull variables with shape parameter $\alpha \in (0, 1]$. Then, for every $p \geq 2$,*

$$C_{\alpha,d}^{-1} \sum_{I \subset [d]} \sum_{\mathcal{J} \in P(I^c)} p^{|I|/\alpha + |\mathcal{J}|/2} \max_{\mathbf{i}_I} \|A_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}$$

$$\leq \|\langle A, w^1 \otimes \ldots \otimes w^d \rangle\|_p \leq C_{\alpha,d} \sum_{I \subset [d]} \sum_{\mathcal{J} \in P(I^c)} p^{|I|/\alpha + |\mathcal{J}|/2} \max_{\mathbf{i}_I} \|A_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}.$$

Moreover, we need a replacement of Lemma 3.2. Here, instead of Gaussian random variables we use Weibull random variables to compare the p-th moments:

Lemma 6.4. *For any $k \in \mathbb{N}$, any $\alpha \in (0, 1]$ and any $p \geq 2$, if $Y_1, \ldots, Y_n$ are independent symmetric random variables with $\|Y_i\|_{\psi_{\alpha/k}} \leq M$, then*

$$\Big\|\sum_{i=1}^n a_i Y_i\Big\|_p \leq C_{\alpha,k} M \Big\|\sum_{i=1}^n a_i w_{i_1} \cdots w_{i_k}\Big\|_p,$$

where w_{ij} are i.i.d. Weibull variables with shape parameter α.

Proof. We extend the arguments given in the proof of [KL15, Corollary 2]. As always, we assume $M = 1$. Moreover, note that it suffices to prove Lemma 6.4 for $p \in 2\mathbb{N}$. It follows from Lemma 6.3 that $\|w_{ij}\|_p \geq C_\alpha p^{1/\alpha}$ for any i, j, from where we easily arrive at $\|w_{i1} \cdots w_{ik}\|_p \geq C_{\alpha,k} p^{k/\alpha}$. Consequently, for a set of independent Rademacher variables $\varepsilon_1, \ldots, \varepsilon_n$ which are independent of the $(Y_i)_i$, $\|Y_i\|_p = \|\varepsilon_i Y_i\|_p \leq C_\alpha p^{k/\alpha} \leq C_{\alpha,k} \|w_{i1} \ldots w_{ik}\|_p$. Therefore, for any $m \in \mathbb{N}$ and using standard symmetrization inequalities,

$$\Big\|\sum_{i=1}^n a_i Y_i\Big\|_{2m} \leq 2 \Big\|\sum_{i=1}^n a_i \varepsilon_i Y_i\Big\|_{2m} \leq C_{\alpha,k} \Big\|\sum_{i=1}^n a_i w_{i1} \cdots w_{ik}\Big\|_{2m}.$$

$\square$

Our next goal is to adapt Lemmas 5.1, 5.2 and 5.3 to the "restricted" tensors $A_{\mathbf{i}_{I^c}}$. That is, we examine whether (a modification of) the inequality

(6.1) $$\|(A \circ 1_C)_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \le \|A_{\mathbf{i}_{I_c}}\|_{\mathcal{J}}$$

still holds in this situation, where $\mathcal{J}$ is a partition of I^c.

Lemma 6.5. *Let $A = (a_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$ be a d-tensor, $I \subset [d]$ and $\mathbf{i}_I \in [n]^I$ fixed.*

(1) *If $C = \{\mathbf{i}: i_{k_1} = j_1, \dots, i_{k_l} = j_l\}$ for some $1 \le k_1 < \dots < k_l \le d$ ("generalized row"), then (6.1) holds.*

(2) *If $C = \{\mathbf{i}: i_k = i_l \ \forall k,l \in K\}$ for some $K \subset [d]$ ("generalized diagonal"), then (6.1) holds.*

(3) *If $C_1, C_2 \subset [n]^d$ are such that (6.1) holds, then so is $C_1 \cap C_2$.*

(4) *If $\mathcal{K} \in P_d$, then $\|(A \circ 1_{L(\mathcal{K})})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \le 2^{|\mathcal{K}|(|\mathcal{K}|-1)/2} \|A_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}$.*

(5) *For any vectors $v_1, \dots, v_d \in \mathbb{R}^n$, $\|(A \circ \otimes_{i=1}^d v_i)_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \le \|A_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \prod_{i=1}^d \|v_i\|_\infty$.*

Proof. To see (1), we may assume that $\{k_1, \dots, k_l\} \cap I = \emptyset$ (note that if $\{k_1, \dots, k_l\} \cap I \ne \emptyset$, either the conditions are not compatible, in which case $(A \circ 1_C)_{i_{I^c}} = 0$, or we can remove some of the conditions and obtain a subset with $\{k_1, \dots, k_{\tilde{l}}\} \cap I = \emptyset$). In this case, if C is a generalized row, then $(A \circ 1_C)_{\mathbf{i}_{I^c}} = A_{\mathbf{i}_{I^c}} \circ 1_{C'}$ for some generalized row C' in I^c. This proves (1).

If C is a generalized diagonal, we have to consider two situations. Assuming $K \cap I = \emptyset$, i.e. K is subset of I^c, we immediately obtain (2). On the other hand, if $K \cap I \ne \emptyset$, then $(A \circ 1_C)_{\mathbf{i}_{I^c}} = A_{\mathbf{i}_{I^c}} \circ 1_{C'}$ for some generalized *row* C' in I^c, readily leading to (2) again.

(3) is clear. To see (4), one may argue as in the proof of Lemma 5.2 (for $q = 1$), replacing Lemma 5.1 (2) and (3) by their analogues we just proved. Finally, an easy modification of the proof of Lemma 5.3 yields (5). $\square$

We are now ready to prove Theorem 6.2. Here, we recall the notation used in the proof of Theorem 1.6, with the only difference that now, by $f \lesssim g$ we mean an inequality of the form $f \le C(D, \alpha) g$, where $C(D, \alpha)$ may depend on D, α.

Proof of Theorem 6.2. We will follow the proof of Theorem 1.6. In particular, let us assume $M = 1$.

Step 1. Recall the inequality

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

from the proof of Theorem 1.6.

Step 2. Applying Lemma 6.5, we arrive at

$$\|f(X) - \mathbb{E} f(X)\|_p \lesssim \sum_{d=1}^D \sum_{\nu=1}^d \sum_{\mathbf{k} \in I_{\nu,d}} \left\| \sum_{\mathbf{i} \in [n]^{\underline{\nu}}} a_{\mathbf{i}}^{\mathbf{k}} (w_{i_1,1}^{(1)} \cdots w_{i_1,k_1}^{(1)}) \cdots (w_{i_\nu,1}^{(\nu)} \cdots w_{i_\nu,k_\nu}^{(\nu)}) \right\|_p .$$

Here, $(w_{i,k}^{(j)})$ is an array of i.i.d. symmetric Weibull variables with shape parameter α. Now we may define d-tensors A_d as in the proof of Theorem 1.6. Similarly as in (5.3), rewriting and applying Lemma 6.3 together with Lemma 6.5 (4) then yields

$$\|f(X) - \mathbb{E}f(X)\|_p \lesssim \sum_{d=1}^{D} \sum_{\nu=1}^{d} \sum_{\mathbf{k} \in I_{\nu,d}} \|\langle A_d \circ 1_{L(\mathcal{K}(k_1,\ldots,k_\nu))}, \otimes_{j=1}^{\nu} \otimes_{k=1}^{k_j} (w_{i,k}^{(j)})_{i \leq n}\rangle\|_p$$

$$\lesssim \sum_{d=1}^{D} \sum_{\nu=1}^{d} \sum_{\mathbf{k} \in I_{\nu,d}} \sum_{I \subset [d]} \sum_{\mathcal{J} \in P(I^c)} p^{|I|/r + |\mathcal{J}|/2} \max_{\mathbf{i}_I} \|(A_d \circ 1_{L(\mathcal{K}(k_1,\ldots,k_\nu))})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}$$

$$\lesssim \sum_{d=1}^{D} \sum_{I \subset [d]} \sum_{\mathcal{J} \in P(I^c)} p^{|I|/r + |\mathcal{J}|/2} \max_{\mathbf{i}_I} \|(A_d)_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}.$$

Step 3. In the proof of Theorem 1.6 we have decomposed

$$\mathbb{E} \frac{\partial^d f}{\partial x_{i_1} \ldots \partial x_{i_d}}(X) = \nu! l_1! \cdots l_\nu! a_{i_1,\ldots,i_d} + R_{\mathbf{i}}^{(d)}$$

with a remainder tensor $R_{\mathbf{i}}^{(d)}$ corresponding to the set of indices $\mathbf{k}$ with $\mathbf{k} > \mathbf{l}$ and $R_{\mathbf{i}}^{(d)} = 0$ for $d = D$. Again, for any $I \subset [D]$ and any partition $\mathcal{J} \in P(I^c)$,

$$\|(A_D)_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \leq \sum_{\mathcal{K} \in P_D} \|(A_D \circ 1_{L(\mathcal{K})})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \leq \sum_{\mathcal{K} \in P_D} \|(\mathbb{E}f^{(D)}(X) \circ 1_{L(\mathcal{K})})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}$$

$$\lesssim \|(\mathbb{E}f^{(D)}(X))_{\mathbf{i}_{I^c}}\|_{\mathcal{J}},$$

using Lemma 6.5 (4) in the last step. To complete the proof, we need to show that for any $d = 1, \ldots, D - 1$, any $I \subset [d]$ and any partitions $\mathcal{I} \in P([d])$, $\mathcal{J} \in P([d]\backslash I)$,

$$(6.2) \qquad \|(R^{(d)} \circ 1_{L(\mathcal{I})})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \lesssim \sum_{k=d+1}^{D} \sum_{\substack{\mathcal{K} \in P([k]\backslash I) \\ |\mathcal{K}| \geq |\mathcal{J}|}} \|(A_k)_{\mathbf{i}_{I^c}}\|_{\mathcal{K}}.$$

Actually, analyzing the proof one can see it is possible to restrict the second sum on the right-hand side to partitions $\mathcal{K}$ with $|\mathcal{K}| \in \{|\mathcal{J}|, |\mathcal{J}| + 1\}$. Once having proven (6.2), it follows from reverse induction that

$$\sum_{d=1}^{D} \sum_{I \subset [d]} \sum_{\mathcal{J} \in P(I^c)} p^{|I|/r + |\mathcal{J}|/2} \max_{\mathbf{i}_I} \|(A_d)_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \lesssim \sum_{d=1}^{D} \sum_{I \subset [d]} \sum_{\mathcal{J} \in P(I^c)} p^{|I|/r + |\mathcal{J}|/2} \max_{\mathbf{i}_I} \|(\mathbb{E}f^{(d)}(X))_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}.$$

Here, we use that for any $p \geq 2$ and any $|\mathcal{K}| \geq |\mathcal{J}|$ we have $p^{|\mathcal{J}|/2} \leq p^{|\mathcal{K}|/2}$. In view of Step 2 and Proposition 3.4, this finishes the proof.

Step 4. One last time we need to recall some definitions from the proof of Theorem 1.6. We fix some $I \subset [d]$ and $\mathbf{i}_I$, an admissible partition $\mathcal{I} \in P([d])$ and some associated extension $\tilde{\mathcal{I}} \in P([k])$, the d-tensor $S_{\mathcal{I}}^{(d,\mathbf{k})} = (s_{\mathbf{i}}^{(d,k_1,\ldots,k_\nu)})_{\mathbf{i} \in [n]^d} = (s_{\mathbf{i}}^{(d)})_{\mathbf{i} \in [n]^d}$ and for any $\mathbf{k} \in I_{\nu, \leq D}$ with $\mathbf{k} > \mathbf{l}$ a $|\mathbf{k}|$-tensor $\tilde{S}^{|\mathbf{k}|}$. The notion of admissibility was not relevant in Theorem 1.6, as we have not fixed any indices I and values $\mathbf{i}_I \in [n]^I$. Here, it simply means that the level sets have to compatible with the fact that we have fixed some of the partial derivatives by I and $\mathbf{i}_I$. Also, note that $\mathcal{I}$ is a partition of $[d]$ and not of $[d]\backslash I$, since it arises from level sets of partial derivatives and includes the partial derivatives taken in I.

Our aim is to prove that there is a partition $\mathcal{K} \in P([k]\backslash I)$ with $|\mathcal{K}| \in \{|\mathcal{J}|, |\mathcal{J}|+1\}$ such that

$$(6.3) \qquad \|(S_{\mathcal{I}}^{(d,\mathbf{k})})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \lesssim \|(A_{|\mathbf{k}|})_{\mathbf{i}_{I^c}}\|_{\mathcal{K}}.$$

$\mathcal{K} = \{K_1, \ldots, K_{\mu+1}\}$ will be defined as follows: for $j = 1, \ldots, \mu$, we add all elements of J_j to K_j, so that it remains to assign the elements $r \in \{d+1, \ldots, |\mathbf{k}|\}$ to the sets K_j. Since $\tilde{\mathcal{I}}$ is a partition of $|\mathbf{k}|$, there is a unique $k \in \{1, \ldots, \nu\}$ such that $r \in \tilde{I}_k$. Take the smallest element $t =: \pi(r)$ in $\tilde{I}_k$ (since $I_k \subset \tilde{I}_k$, we have $t \in [d]$). If $t \in I^c$, it follows that $t \in K_j$ for some set K_j and we add r to K_j. If $t \in I$, we assign r to an "extra set" $K_{\mu+1}$. In particular, it may happen that $K_{\mu+1} = \emptyset$. In this case, we ignore $\beta = \mu + 1$ in the rest of the proof.
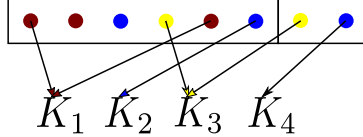

FIGURE 1. An illustration of the procedure of producing the partition $\mathcal{K}$; here, $I = \{1, 2\}$ and we used colors to indicate the partition, i.e. $\tilde{\mathcal{I}} = \{\{1, 2, 5\}, \{3, 6, 8\}, \{4, 7\}\}$. $\{8\}$ belongs to $K_4 = K_{\mu+1}$ since $\{3\} \in I$. Changing its color to yellow would produce a partition $\mathcal{K}$ with 3 subsets.

First off, we claim

$$(6.4) \qquad \|(S_{\mathcal{I}}^{(d,|\mathbf{k}|)})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \leq \|(\tilde{S}^{|\mathbf{k}|})_{\mathbf{i}_{I^c}}\|_{\mathcal{K}}.$$

To see (6.4), let $x = (x^{(\beta)})_{\beta=1,\ldots,\mu} = ((x_{\mathbf{i}_{J_\beta}}^{(\beta)}))_{\beta=1,\ldots,\mu}$, be such that

$$|x|_{\mathcal{J}} = \max_{\beta=1,\ldots,\mu} \|x^{(\beta)}\|_2 \leq 1.$$

We embed this in the unit ball with respect to $|x|_{\mathcal{K}} = \max_{i=1,\ldots,\mu+1} \|x^\beta\|_2$ by defining $y = (y^{(\beta)})_{\beta=1,\ldots,\mu+1}$ via

$$y_{\mathbf{i}_{K_\beta}}^{(\beta)} = \begin{cases} x_{\mathbf{i}_{K_\beta \cap [d]}}^{(\beta)} \prod_{r \in K_\beta \setminus [d]} 1_{i_r = i_{\pi(r)}} & \beta = 1, \ldots, \mu \\ \prod_{r \in K_{\mu+1}} 1_{i_r = i_{\pi(r)}} & \beta = \mu + 1. \end{cases}$$

Note that $y^{(\mu+1)}$ only has a single non-zero element, and thus it is easy to see that $|y|_{\mathcal{K}} \leq 1$. Moreover, by the definition of the matrix $\tilde{S}^{|k|}$ and the fact that if $\mathbf{i} \in L(\tilde{\mathcal{I}})$, then for $r > d$, $i_r = i_{\pi(r)}$, which implies $y_{\mathbf{i}_{K_\beta}}^{(\beta)} = x_{\mathbf{i}_{K_\beta \cap [d]}}^{(\beta)} = x_{\mathbf{i}_{J_\beta}}^{(\beta)}$ as well as $y_{\mathbf{i}_{K_{\mu+1}}}^{(\mu+1)} = 1$ we have

$$(6.5) \qquad \langle (S^{(d,\mathbf{k})})_{\mathbf{i}_{I^c}}, \bigotimes_{\beta=1}^{\mu} x^{(\beta)} \rangle = \langle (\tilde{S}^{(|k|)})_{\mathbf{i}_{I^c}}, \bigotimes_{\beta=1}^{\mu+1} y^{(\beta)} \rangle.$$

Hence, the supremum on the left hand side of (6.4) is taken over a subset of the unit ball with respect to $|x|_{\mathcal{K}}$.

Finally, it remains to prove

$$(6.6) \qquad \|(\tilde{S}^{|\mathbf{k}|})_{\mathbf{i}_{I^c}}\|_{\mathcal{K}} \lesssim \|(A_{|\mathbf{k}|})_{\mathbf{i}_{I^c}}\|_{\mathcal{K}}$$

for any partition $\mathcal{K} \in P(I^c)$. This may be achieved as in the proof of Theorem 1.6, replacing Lemma 5.3 by Lemma 6.5 (5).

Combining (6.4) and (6.6) yields (6.3), which finishes the proof. $\qquad \square$

It remains to prove Proposition 1.1 and Theorem 1.2 (from which Corollary 1.3 follows immediately).

Proof of Proposition 1.1. The case $\alpha \in (0, 1]$ follows immediately from the $d = 2$ case of Corollary 6.1. $\alpha = 2$ corresponds to the well-known Hanson–Wright inequality, see e. g. [RV13]. $\qquad \square$

Proof of Theorem 1.2. Let $\alpha \in (0, 1]$ and consider the bound given by Theorem 6.2. Fix any $d = 1, \ldots, D$. Then, for any $I \subset [d]$, any $\mathbf{i}_I$ and any $\mathcal{J} \in P(I^c)$, we have

$$\|(\mathbb{E}f^{(d)}(X))_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \leq \|(\mathbb{E}f^{(d)}(X))_{\mathbf{i}_{I^c}}\|_{\mathrm{HS}} \leq \|\mathbb{E}f^{(d)}(X)\|_{\mathrm{HS}}$$

(using (3.7)) as well as

$$\frac{\alpha}{d} \leq \frac{2\alpha}{2|I| + \alpha|\mathcal{J}|} \leq 2.$$

If $t/(M^d\|\mathbb{E}f^{(d)}(X)\|_{\mathrm{HS}}) \geq 1$, this immediately yields the result. Otherwise, note that the tail bound given in Theorem 1.2 is trivial. (In fact, here one needs to ensure that $C_{D,\alpha}$ is sufficiently large, e. g. $C_{D,\alpha} \geq 1$. It is not hard to see that in general this condition will be satisfied anyway.)

In a similar way, it is possible to derive the same results for $\alpha = 2/q$ and any $q \in \mathbb{N}$ from Theorem 1.6.

From these results, the exponential moment bound follows by standard arguments, see for example [BGS18, Proof of Theorem 1.1]. $\qquad \square$

APPENDIX A. PROPERTIES OF ORLICZ QUASINORMS

As mentioned in the introduction, Orlicz norms (1.5) satisfy the triangle inequality only for $\alpha \geq 1$. However, for any $\alpha \in (0, 1)$ (1.5) still is a quasinorm, which for many purposes is sufficient. We shall collect some elementary results on Orlicz quasinorms in this appendix. The first result is a Hölder-type inequality for the Ψ_α norms.

Lemma A.1. *Let $X_1, \ldots, X_k$ be random variables such that $\|X_i\|_{\Psi_{\alpha_i}} < \infty$ for some $\alpha_i \in (0, 1]$ and let $t := (\sum_{i=1}^k \alpha_i^{-1})^{-1}$. Then $\|\prod_{i=1}^k X_i\|_{\Psi_t} < \infty$ and*

$$\Big\|\prod_{i=1}^k X_i\Big\|_{\Psi_t} \leq \prod_{j=1}^k \|X_i\|_{\Psi_{\alpha_i}}.$$

Proof. By homogeneity we can assume $\|X\|_{\Psi_{\alpha_i}} = 1$ for all $i = 1, \ldots, k$. We will need the general form of Young's inequality, i. e. for all $p_1, \ldots, p_k > 1$ satisfiyng $\sum_{i=1}^k p_i^{-1} = 1$ and any $x_1, \ldots, x_k \geq 0$ we have

$$\prod_{i=1}^k x_i \leq \sum_{i=1}^k p_i^{-1} x_i^{p_i},$$

which follows easily from the concavity of the logarithm. If we apply this to $p_i := \alpha_i t^{-1}$ and use the convexity of the exponential function, we obtain

$$\mathbb{E}\exp\Big(\prod_{i=1}^k |X_i|^t\Big) \leq \mathbb{E}\exp\Big(\sum_{j=1}^k p_i^{-1}|X_i|^{\alpha_i}\Big) \leq \sum_{j=1}^k p_i^{-1}\mathbb{E}\exp\Big(|X_i|^{\alpha_i}\Big) \leq 2.$$

Consequently, we have $\|\prod_{i=1}^k X_i\|_{\Psi_t} \leq 1$. $\qquad \square$

The random variables $X_1, \ldots, X_k$ need not be independent, i.e. we can consider a random vector $X = (X_1, \ldots, X_k)$ with marginals having α-sub-exponential tails. The special case $\alpha_i = \alpha$ for all $i = 1, \ldots, k$ gives

$$\Big\| \prod_{i=1}^{k} X_i \Big\|_{\Psi_{\alpha/k}} \leq \prod_{j=1}^{k} \|X_i\|_{\Psi_\alpha}.$$

To state the other lemmas, for any $0 < \alpha < 1$ define

(A.1) $$d_\alpha := (\alpha e)^{1/\alpha}/2 \qquad \text{and} \qquad D_\alpha := (2e)^{1/\alpha}.$$

Lemma A.2. *For any $0 < \alpha < 1$ we have*

(A.2) $$d_\alpha \sup_{p \geq 1} \frac{\|X\|_p}{p^{1/\alpha}} \leq \|X\|_{\Psi_\alpha} \leq D_\alpha \sup_{p \geq 1} \frac{\|X\|_p}{p^{1/\alpha}}.$$

The statement of the lemma remains true for $\alpha \geq 1$, with (α-independent constants) $d_\alpha = 1/2$ and $D_\alpha = 2e$, see [Bob10, Section 8]. In the proof, we will closely follow the proof therein, but keep track of the α-dependent constants.

Proof. We begin with the left inequality. By homogeneity, we assume $\|X\|_{\Psi_\alpha} = 1$. First let us show that we have

(A.3) $$g(x) := (\alpha e)^{-1/\alpha} e^{x^\alpha} - x \geq 0 \qquad \text{for} \quad x \geq 0.$$

Note that g is continuous on $[0, \infty)$ and differentiable on $(0, \infty)$ with $g(0) > 0$ and $g(x) \to \infty$ as $x \to \infty$. Therefore, it suffices to find the critical points. We can rewrite the condition $g'(x) = 0$ as $e^y y = y^{1/\alpha} (\alpha e)^{1/\alpha}$, setting $y := x^\alpha$. From this representation it can be seen that there can be at most two points x_0 and x_1 satisfying this condition. One of these points is $x_\alpha := \alpha^{-1/\alpha}$, and we have $g(x_\alpha) = 0$. A short calculation shows that $g''(x_\alpha) = \alpha^{1/\alpha+1} > 0$, so that x_α is a global minimum, from which $g \geq 0$ follows.

Next, from this we can infer for all $p \geq 1$ and $\alpha > 0$

(A.4) $$x^p \leq \left(\frac{p}{\alpha e} \right)^{p/\alpha} e^{x^\alpha}.$$

Indeed, by a transformation $y = x^p$ and the change $\tilde{\alpha} = \frac{\alpha}{p}$ this is just an application of (A.3). Consequently, for any $p \geq 1$ we have

$$\mathbb{E}|X|^p \leq \left(\frac{p}{\alpha e} \right)^{p/\alpha} \mathbb{E} \exp \left(|X|^\alpha \right) \leq 2 \left(\frac{p}{\alpha e} \right)^{p/\alpha} \leq 2^p \left(\frac{p}{\alpha e} \right)^{p/\alpha},$$

i.e.

$$\|X\|_p \leq 2(\alpha e)^{-1/\alpha} p^{1/\alpha}.$$

For the second inequality, again assume that $\sup_{p \geq 1} \frac{\|X\|_p}{p^{1/\alpha}} = 1$. First, we need to extend the the supremum to $p \in [\alpha, \infty)$, which can be done as follows. For any $p \in [\alpha, 1)$ we have

$$\frac{\|X\|_p}{p^{1/\alpha}} \leq \frac{\|X\|_1}{p^{1/\alpha}} \leq \frac{1}{p^{1/\alpha}} \leq \frac{1}{\alpha^{1/\alpha}}$$

and therefore

$$\sup_{p \geq \alpha} \frac{\|X\|_p}{p^{1/\alpha}} \leq \frac{1}{\alpha^{1/\alpha}}.$$

Now, by Taylor's expansion and using the inequality $n^n \leq e^n n!$ this gives

$$\mathbb{E} \exp\left(\frac{|X|^\alpha}{t^\alpha}\right) = 1 + \sum_{n=1}^{\infty} \frac{\mathbb{E}|X|^{\alpha n}}{t^{\alpha n} n!} \leq 1 + \sum_{n=1}^{\infty} \frac{n^n}{n! t^{\alpha n}} \leq 1 + \sum_{n=1}^{\infty} \left(\frac{e}{t^\alpha}\right)^n = \frac{1}{1 - e/t^\alpha}.$$

For $t = (2e)^{1/\alpha}$ this is less or equal to 2, so that

$$\|X\|_{\Psi_\alpha} \leq (2e)^{1/\alpha} \sup_{p \geq 1} \frac{\|X\|^p}{p^{1/\alpha}}.$$

$\square$

Lemma A.3. *For any $0 < \alpha < 1$ and any random variables X, Y we have*

(A.5) $$\|X + Y\|_{\Psi_\alpha} \leq 2^{1/\alpha} \left(\|X\|_{\Psi_\alpha} + \|Y\|_{\Psi_\alpha}\right).$$

Proof. Let $K := \|X\|_{\Psi_\alpha}$ and $L := \|Y\|_{\Psi_\alpha}$ and define $t := 2^{1/\alpha}(K + L)$. We have

$$\mathbb{E} \exp\left(\frac{|X + Y|^\alpha}{t^\alpha}\right) \leq \mathbb{E} \exp\left(\frac{(|X| + |Y|)^\alpha}{t^\alpha}\right) \leq \mathbb{E} \exp\left(\frac{|X|^\alpha + |Y|^\alpha}{2(K + L)^\alpha}\right)$$

$$\leq \mathbb{E} \exp\left(\frac{|X|^\alpha}{2K^\alpha}\right) \exp\left(\frac{|Y|^\alpha}{2L^\alpha}\right)$$

$$\leq \frac{1}{2} \mathbb{E} \exp\left(\frac{|X|^\alpha}{K^\alpha}\right) + \frac{1}{2} \mathbb{E} \exp\left(\frac{|Y|^\alpha}{L^\alpha}\right) \leq 2.$$

Here, the second step follows from the inequality $(x + y)^\alpha \leq x^\alpha + y^\alpha$ valid for all $x, y \geq 0$ and $\alpha \in [0, 1]$, and the fourth one is an application of Young's inequality $ab \leq a^2/2 + b^2/2$ for all positive a, b. $\square$

Lemma A.4. *Let $0 < \alpha < 1$. For all random variables X we have*

$$\|\mathbb{E} X\|_{\Psi_\alpha} \leq \frac{1}{d_\alpha (\log 2)^{1/\alpha}} \|X\|_{\Psi_\alpha}.$$

Proof. Assuming $\|X\|_{\Psi_\alpha} < \infty$, an application of Lemma A.2 gives

$$\|\mathbb{E} X\|_{\Psi_\alpha} = \frac{|\mathbb{E} X|}{(\log 2)^{1/\alpha}} \leq \frac{\|X\|_1}{(\log 2)^{1/\alpha}} \leq \frac{1}{d_\alpha (\log 2)^{1/\alpha}} \|X\|_{\Psi_\alpha}.$$

$\square$

As a consequence of the last two results, we can readily infer the following corollary.

Corollary A.5. *For any $\alpha > 0$ and any random variable X we have*

$$\|X - \mathbb{E} X\|_{\Psi_\alpha} \leq 2^{1/\alpha} \left(1 + (d_\alpha \log 2)^{-1/\alpha}\right) \|X\|_{\Psi_\alpha}.$$

References

[Ada06] Radosław Adamczak. "Moment inequalities for U-statistics". In: *Ann. Probab.* 34.6 (2006), pp. 2288–2314. DOI: 10.1214/009117906000000476.

[Ada15] Radosław Adamczak. "A note on the Hanson-Wright inequality for random vectors with dependencies". In: *Electron. Commun. Probab.* 20 (2015), no. 72, 13. DOI: 10.1214/ECP.v20-3829.

[AL12] Radosław Adamczak and Rafał Latała. "Tail and moment estimates for chaoses generated by symmetric random variables with logarithmically concave tails". In: *Ann. Inst. Henri Poincaré Probab. Stat.* 48.4 (2012), pp. 1103–1136. DOI: 10.1214/11-AIHP441.

[AW15] Radosław Adamczak and Paweł Wolff. "Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order". In: *Probab. Theory Related Fields* 162.3-4 (2015), pp. 531–586. DOI: 10.1007/s00440-014-0579-3.

[BGS18]   Sergey G. Bobkov, Friedrich Götze, and Holger Sambale. "Higher order concentration of measure". In: *Commun. Contemp. Math.* online first (2018), p. 0. DOI: `10.1142/S0219199718500438`.

[BL97]   Sergey G. Bobkov and Michel Ledoux. "Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution". In: *Probab. Theory Related Fields* 107.3 (1997), pp. 383–400. DOI: `10.1007/s004400050090`.

[Bob10]   Sergey G. Bobkov. "The growth of L^p-norms in presence of logarithmic Sobolev inequalities". In: *Vestnik Syktyvkar Univ.* 11.2 (2010), pp. 92–111. URL: `http://www-users.math.umn.edu/~bobko001/papers/2010_Vestnik_Growth.under.log-Sobolev.pdf`.

[Cha08]   Sourav Chatterjee. "A new method of normal approximation". In: *Ann. Probab.* 36.4 (2008), pp. 1584–1610. DOI: `10.1214/07-AOP370`.

[CTP17]   Joshua Cape, Minh Tang, and Carey E. Priebe. "The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics". In: *arXiv preprint* (2017). arXiv: `1705.10735`.

[EYY12]   László Erdős, Horng-Tzer Yau, and Jun Yin. "Bulk universality for generalized Wigner matrices". In: *Probab. Theory Related Fields* 154.1-2 (2012), pp. 341–407. DOI: `10.1007/s00440-011-0390-3`.

[GK95]   Efim D. Gluskin and Stanisław Kwapień. "Tail and moment estimates for sums of independent random variables with logarithmically concave tails". In: *Studia Math.* 114.3 (1995), pp. 303–309. DOI: `10.4064/sm-114-3-303-309`.

[GS18]   Friedrich Götze and Holger Sambale. "Higher order concentration in presence of Poincaré-type inequalities". In: *arXiv preprint* (2018). arXiv: `1803.05190`.

[GSS18a] Friedrich Götze, Holger Sambale, and Arthur Sinulis. "Concentration inequalities for bounded functionals via generalized log-Sobolev inequalities". In: *arXiv preprint* (2018). arXiv: `1812.01092`.

[GSS18b] Friedrich Götze, Holger Sambale, and Arthur Sinulis. "Higher order concentration for functions of weakly dependent random variables". In: *arXiv preprint* (2018). arXiv: `1801.06348`.

[GT99]   Friedrich Götze and Alexander N. Tikhomirov. "Asymptotic distribution of quadratic forms". In: *Ann. Probab.* 27.2 (1999), pp. 1072–1098. DOI: `10.1214/aop/1022677395`.

[HKZ12]   Daniel Hsu, Sham M. Kakade, and Tong Zhang. "A tail inequality for quadratic forms of subgaussian random vectors". In: *Electron. Commun. Probab.* 17 (2012), no. 52, 6. DOI: `10.1214/ECP.v17-2079`.

[HMO97]   Paweł Hitczenko, Stephen J. Montgomery-Smith, and Krzysztof Oleszkiewicz. "Moment inequalities for sums of certain independent symmetric random variables". In: *Studia Math.* 123.1 (1997), pp. 15–42. URL: `http://eudml.org/doc/216377`.

[Jon87]   Peter de Jong. "A central limit theorem for generalized quadratic forms". In: *Probab. Theory Related Fields* 75.2 (1987), pp. 261–277. DOI: `10.1007/BF00354037`.

[KL15]   Konrad Kolesko and Rafał Latała. "Moment estimates for chaoses generated by symmetric random variables with logarithmically convex tails". In: *Statist. Probab. Lett.* 107 (2015), pp. 210–214. DOI: `10.1016/j.spl.2015.08.019`.

[Lat06]   Rafał Latała. "Estimates of moments and tails of Gaussian chaoses". In: *Ann. Probab.* 34.6 (2006), pp. 2315–2331. DOI: `10.1214/009117906000000421`.

[Lat96]   Rafał Latała. "Tail and moment estimates for sums of independent random vectors with logarithmically concave tails". In: *Studia Math.* 118.3 (1996), pp. 301–304. DOI: `10.4064/sm-118-3-301-304`.

[Lat99]   Rafał Latała. "Tail and moment estimates for some types of chaos". In: *Studia Math.* 135.1 (1999), pp. 39–53. DOI: `10.4064/sm-135-1-39-53`.

[Led01]   Michel Ledoux. *The concentration of measure phenomenon*. Vol. 89. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2001, pp. x+181. ISBN: 0-8218-2864-9.

[LŁ03]   Rafał Latała and Rafał Łochowski. "Moment and tail estimates for multidimensional chaos generated by positive random variables with logarithmically concave tails". In: *Stochastic inequalities and applications*. Vol. 56. Progr. Probab. Birkhäuser, Basel, 2003, pp. 77–92. DOI: `10.1007/978-3-0348-8069-5_7`.

[Mel16]   Rafał Meller. "Two-sided moment estimates for a class of nonnegative chaoses". In: *Statist. Probab. Lett.* 119 (2016), pp. 213–219. DOI: `10.1016/j.spl.2016.08.005`.

[Mel17]   Rafał Meller. "Tail and moment estimates for a class of random chaoses of order two". In: *arXiv preprint* (2017). arXiv: `1708.05856`.

[NSU17]   Alexey Naumov, Vladimir Spokoiny, and Vladimir Ulyanov. "Bootstrap confidence sets for spectral projectors of sample covariance". In: *arXiv preprint* (2017). arXiv: `1703.00871`.

[PG99]   Víctor H. de la Peña and Evarist Giné. *Decoupling*. Probability and its Applications (New York). From dependence to independence, Randomly stopped processes. *U*-statistics and processes. Martingales and beyond. Springer-Verlag, New York, 1999, pp. xvi+392. ISBN: 0-387-98616-2.

[PTT18]   Joscha Prochno, Christoph Thäle, and Nicola Turchi. "The isotropic constant of random polytopes with vertices on convex surfaces". In: *arXiv preprint* (2018). arXiv: `1807.02396`.

[RV13]   Mark Rudelson and Roman Vershynin. "Hanson-Wright inequality and sub-Gaussian concentration". In: *Electron. Commun. Probab.* 18 (2013), no. 82, 9. DOI: `10.1214/ECP.v18-2865`.

[SS11]   Warren Schudy and Maxim Sviridenko. "Bernstein-like concentration and moment inequalities for polynomials of independent random variables: multilinear case". In: *arXiv preprint* (2011). arXiv: `1109.5193`.

[SS12]   Warren Schudy and Maxim Sviridenko. "Concentration and moment inequalities for polynomials of independent random variables". In: *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*. ACM, New York, 2012, pp. 437–446. DOI: `10.1137/1.9781611973099.37`.

[SS18]   Holger Sambale and Arthur Sinulis. "Logarithmic Sobolev inequalities for finite spin systems and applications". In: *arXiv preprint* (2018). arXiv: `1807.07765`.

[Ver12]   Roman Vershynin. "Introduction to the non-asymptotic analysis of random matrices". In: *Compressed sensing*. Cambridge Univ. Press, Cambridge, 2012, pp. 210–268.

[Ver18]   Roman Vershynin. *High-dimensional probability*. Vol. 47. Cambridge Series in Statistical and Probabilistic Mathematics. An introduction with applications in data science, With a foreword by Sara van de Geer. Cambridge University Press, Cambridge, 2018, pp. xiv+284. ISBN: 978-1-108-41519-4.

[VW15]   Van H. Vu and Ke Wang. "Random weighted projections, random quadratic forms and random eigenvectors". In: *Random Structures Algorithms* 47.4 (2015), pp. 792–821. DOI: 10.1002/rsa.20561.

[Yan+17] Zhuoran Yang et al. "Misspecified Nonconvex Statistical Optimization for Phase Retrieval". In: *arXiv preprint* (2017). arXiv: 1712.06245.