Electronic Journal of Probability

Electron. J. Probab. **26** (2021), article no. 48, 1–22.
ISSN: 1083-6489 https://doi.org/10.1214/21-EJP606

Concentration inequalities for polynomials in α-sub-exponential random variables*

Friedrich Götze[†] Holger Sambale[‡] Arthur Sinulis[§]

Abstract

We derive multi-level concentration inequalities for polynomials in independent random variables with an 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 them are generalizations of some results proven by Rudelson and 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 and concentration inequalities for the distance between a random vector and a fixed subspace.

Keywords: concentration of measure phenomenon; Orlicz norms; Hanson-Wright inequality; Poisson chaos; sub-exponential random variables.
MSC2020 subject classifications: Primary 60E15, Secondary 46E30; 46N30.
Submitted to EJP on September 10, 2019, final version accepted on March 12, 2021.
Supersedes arXiv:1903.05964.

1 Introduction

Let $X_1, \ldots, X_n$ be independent random variables and $f : \mathbb{R}^n \to \mathbb{R}$ 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

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

*This research was supported by the German Research Foundation (DFG) via CRC 1283 "*Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications*".

[†]Faculty of Mathematics, Bielefeld University. E-mail: goetze@math.uni-bielefeld.de
[‡]Faculty of Mathematics, Bielefeld University. E-mail: hsambale@math.uni-bielefeld.de
[§]Faculty of Mathematics, Bielefeld University. E-mail: asinulis@math.uni-bielefeld.de

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 standard Gaussian random variables, and $f(X_1, \ldots, X_n) = n^{-1/2} \sum_{i=1}^n X_i$, then $f(X_1, \ldots, X_n)$ is a standard Gaussian and satisfies

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

Using the entropy method, it is possible to show that the estimate (1.2) remains true for any 1-Lipschitz function f (see e. g. [24, Chapter 5]).

On the other hand, if f is a quadratic form, then its tails are heavier. Indeed, the *Hanson–Wright inequality* (see [14], [35], [29]) states that for a quadratic form in independent, centered sub-Gaussian random variables $X_1, \ldots, X_n$ with $\mathbb{E}\, X_i^2 = 1$ we have for some absolute constant $C > 0$

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

Here, $\|A\|_{\mathrm{op}}$ is the operator norm and $\|A\|_{\mathrm{HS}}$ the Hilbert–Schmidt norm (Frobenius norm) of A respectively. Thus the tails of the quadratic form decay like $\exp(-ct/\|A\|_{\mathrm{op}})$ for large t. There are inequalities similar to (1.3) for multilinear chaos in Gaussian random variables proven in [22] (and in fact, a lower bound using the same quantities as well), and in [4] for polynomials in sub-Gaussian random variables. Moreover, extensions of the Hanson–Wright inequality to certain types of dependent random variables have been considered in [17] and [2], for instance. However, a key component is that the individual random variables X_i have a sub-Gaussian tail decay.

In recent works [6], [13], [12] we have studied similar concentration inequalities for bounded functions f of independent and 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 polynomials.

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 which we mean that there exists two constants c, C and a parameter $\alpha > 0$ such that for all $i = 1, \ldots, n$ and $t \geq 0$ it holds

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

There are many interesting choices of random variables X_i of this type, like 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 as present in 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 and $\alpha > 0$ define the (quasi-)norm

$$\|X\|_{\Psi_\alpha} := \inf \Big\{ t > 0 \colon \mathbb{E} \exp\Big(\frac{|X|^\alpha}{t^\alpha} \Big) \leq 2 \Big\}, \tag{1.5}$$

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 we refer to Appendix A.

Throughout the article, we denote by C an absolute constant and by $C_{l_1,\ldots,l_k}$ a constant that only depends on some parameters $l_1, \ldots, l_k$. These constants may vary from line to line. Moreover, we frequently write $X = (X_1, \ldots, X_n)$ for the vector consisting of the random variables under consideration.

Just to provide one example which might be of particular interest for applications, we start with a simplified version of a later result. More precisely, we present the following concentration inequality which may be considered as an analogue 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 $\mathbb{E}\, X_i = 0, \mathbb{E}\, X_i^2 = \sigma_i^2$, $\|X_i\|_{\Psi_\alpha} \leq M$ for some $\alpha \in (0,1]$, and $A = (a_{ij})$ be a symmetric $n \times n$ matrix. For any $t \geq 0$ we have*

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

As we will see in Corollary 1.4 (2), 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. Note that with slightly different techniques, it is possible to extend Proposition 1.1 to $\alpha \in (0,2]$, as was done in [30].

Our main results presented in the next section yield bounds for arbitrary polynomials in α-sub-exponential random variables for $\alpha \in (0,1]$. These inequalities typically involve larger families of norms, resulting in refined tail estimates.

1.1 Main results

Let us first introduce some notation. Define $[n] := \{1, \ldots, n\}$, and let $\mathbf{i} = (i_1, \ldots, i_d) \in [n]^d$ be a multiindex (which we typically write in boldface letters). 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 $A_{\mathbf{i}_{I^c}} = (a_{1j2k})_{jk}$. 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. If $I = \emptyset$, $\mathbf{i}_I$ does not indicate any specification, and $A_{\mathbf{i}_{I^c}} = A$.

Moreover, we write P_d for the set of all partitions of $[d]$ and $P(I^c)$ for the set of all partitions of I^c for any $I \subset [d]$. If $\mathcal{J} = \{J_1, \ldots, J_k\}$ is any partition, we denote by $|\mathcal{J}|$ the number of subsets it contains. For $I = [d]$, by convention there is a single element $\mathcal{J} \in P(I^c)$ which we may call the "empty" partition, thus justifying that we set $|\mathcal{J}| := 0$.

Next we introduce a family of tensor-product matrix norms $\|A\|_{\mathcal{J}}$ for a d-tensor A and a partition $\mathcal{J} = \{J_1, \ldots, J_k\} \in P_d$. For each $l = 1, \ldots, k$ we denote by $x^{(l)}$ a vector in $\mathbb{R}^{n^{J_l}}$. Then, we set

$$\|A\|_{\mathcal{J}} := \sup\Big\{\sum_{\mathbf{i} \in [n]^d} a_{\mathbf{i}} \prod_{l=1}^k x_{\mathbf{i}_{J_l}}^{(l)} : \sum_{\mathbf{i}_{J_l}} (x_{\mathbf{i}_{J_l}}^{(l)})^2 \leq 1 \text{ for all } l = 1, \ldots, k\Big\}.$$

In particular, this also defines norms $\|A_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}$ for any $I \subset [d]$ and any $\mathcal{J} \in P(I^c)$. If $I = [d]$, $\|A_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} = |a_{\mathbf{i}}|$ is just the Euclidean norm of $a_{\mathbf{i}}$.

The family $\|\cdot\|_{\mathcal{J}}$ was first introduced in [22], where it was used to prove two-sided estimates for L^p norms of Gaussian chaos, and the definitions given above agree with the ones from [22] and [4] (among others). We can regard the $\|A\|_{\mathcal{I}}$ as a family of operator-type norms. In particular, it is easy to see that $\|A\|_{\{1,\ldots,d\}} = \|A\|_{\mathrm{HS}} := (\sum_{i_1,\ldots,i_d} a^2_{i_1\ldots i_d})^{1/2}$ (Hilbert–Schmidt norm) and $\|A\|_{\{\{1\},\ldots,\{d\}\}} = \|A\|_{\mathrm{op}} := \sup\{\sum_{i_1,\ldots,i_d} a_{i_1\ldots i_d} x^{(1)}_{i_1} \cdots x^{(d)}_{i_d} : |x^{(l)}| \le 1 \text{ for all } l = 1,\ldots,d\}$ (operator norm).

In fact, these norms are monotone with respect to the underlying partition in the following sense. For two partitions $\mathcal{I} = \{I_1,\ldots,I_\mu\}$ and $\mathcal{J} = \{J_1,\ldots,J_\nu\}$ of $[d]$, we say that $\mathcal{I}$ is finer than $\mathcal{J}$ if for any $j = 1,\ldots,\mu$ there is a $k \in \{1,\ldots,\nu\}$ such that $I_j \subseteq J_k$. In this case, it is easy to see that $\|A\|_{\mathcal{I}} \le \|A\|_{\mathcal{J}}$. In particular, we always have

$$\|A\|_{\mathrm{op}} = \|A\|_{\{\{1\},\ldots,\{d\}\}} \le \|A\|_{\mathcal{J}} \le \|A\|_{\{1,\ldots,d\}} = \|A\|_{\mathrm{HS}}, \tag{1.6}$$

so that the two norms highlighted above can be regarded as "extremal cases" of the family $\|\cdot\|_{\mathcal{J}}$.

Before stating our main theorem, let us recall a result by Kolesko and Latała which provides bounds for a d-homogeneous chaos in independent α-sub-exponential random variables. By a d-th order homogeneous chaos, we mean the polynomial

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

Here, $A = (a_{i_1\ldots i_d})$ is a d-tensor 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 $\sigma \in \mathcal{S}_d$). Additionally, we often assume that A has vanishing diagonal in the sense that $a_{i_1\ldots i_d} = 0$ whenever $i_1,\ldots,i_d$ are not pairwise different. Using the characterization of the Ψ_α norms in terms of the growth of L^p norms (see Appendix A for details), [18, Corollary 2] yields the following result:

Proposition 1.2. *Let $X_1,\ldots,X_n$ be a set of independent, centered random variables with $\|X_i\|_{\Psi_\alpha} \le M$ 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 $t \ge 0$*

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

Our main result generalizes Proposition 1.2 to arbitrary polynomials in random variables with bounded Orlicz norms. For sub-Gaussian random variables, i. e. $\alpha = 2$, such a result has been obtained in [4, Theorem 1.4]. Our next theorem can be regarded as an analogous result for $\alpha \in (0,1]$. 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 1.3. *Let $X_1,\ldots,X_n$ be a set of independent random variables satisfying $\|X_i\|_{\Psi_\alpha} \le M$ for some $\alpha \in (0,1]$. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a polynomial of total degree $D \in \mathbb{N}$. Then, for any $t \ge 0$,*

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

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 Proposition 1.2. We discuss Theorem 1.3 and compare it to known results in Subsection 1.2. For illustration, let us consider some simple cases next. The following corollary evaluates the cases of $d = 1, 2$.

Corollary 1.4. *Let $X_1,\ldots,X_n$ be a set of independent, centered random variables with $\|X_i\|_{\Psi_\alpha} \le M$ for some $\alpha \in (0,1]$.*

1. *Let $a \in \mathbb{R}^n$. For any $t \ge 0$ it holds*

$$\mathbb{P}\Big(\big|\sum_{i=1}^n a_i X_i\big| \ge t\Big) \le 2\exp\Big(-\frac{1}{C_\alpha}\min\Big(\frac{t^2}{M^2|a|^2}, \frac{t^\alpha}{M^\alpha \max_i |a_i|^\alpha}\Big)\Big).$$

2. *Let A be a symmetric matrix. Writing $\mathbb{E}\, X_i^2 = \sigma_i^2$, we have for any $t \geq 0$*

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

where, setting $\|A\|_\infty := \max_{i,j} |a_{ij}|$, we have

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

Let us complement these results by some simple observations. First, up to constants, Corollary 1.4 (1) gives back a classical result for the tails of a linear form in random variables with sub-exponential tails for $\alpha = 1$. For more general functions and similar results under a Poincaré-type inequality, we refer to [7] (the first order case) and [11] (the higher order case).

Moreover, Corollary 1.4 (2) is a sharpened version of Proposition 1.1. In comparison, 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, we have $\|A\|_\infty = \max_{i,j} |\langle e_i, A e_j \rangle|$ for the standard basis $(e_i)_i$ of $\mathbb{R}^n$, and it can easily be seen that $\max_{i=1,\ldots,n} \|(a_{ij})_j\|_2 = \|A\|_{2\to\infty} := \max_{\|x\|_2 = 1} \|Ax\|_\infty$. Moreover, the norms might have a very different scaling in n. For example, for $e = (1, \ldots, 1)$ and $A = e^T e - \mathrm{Id}$ we have $\|A\|_{\mathrm{HS}} \sim \|A\|_{\mathrm{op}} \sim n$, $\max_i \|(a_{ij})_j\|_2 \sim n^{1/2}$ and $\|A\|_\infty = 1$.

Furthermore, let us provide a simplified version of Theorem 1.3 which only involves Hilbert–Schmidt norms. These norms are typically easiest to calculate, and another benefit is that in this situation, we may extend our results to any $\alpha \in (0, 2]$.

Theorem 1.5. *Let $X_1, \ldots, X_n$ be independent random variables satisfying $\|X_i\|_{\Psi_\alpha} \leq M$ for some $\alpha \in (0, 2]$ and let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a polynomial of total degree $D \in \mathbb{N}$. Then for all $t \geq 0$ it holds*

$$\mathbb{P}(|f(X) - \mathbb{E} f(X)| \geq t) \leq 2 \exp\Big(- \frac{1}{C_{D,\alpha}} \min_{1 \leq d \leq D} \Big(\frac{t}{M^d \|\mathbb{E} f^{(d)}(X)\|_{\mathrm{HS}}} \Big)^{\frac{\alpha}{d}} \Big). \tag{1.8}$$

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

$$\mathbb{E} \exp\Bigg(\frac{1}{C_{D,\alpha} M^\alpha} |f(X) - \mathbb{E} f(X)|^{\frac{\alpha}{D}} \Bigg) \leq 2,$$

or equivalently $\|f(X) - \mathbb{E} f(X)\|_{\Psi_{\frac{\alpha}{D}}} \leq C_{D,\alpha} M^D$.

Intuitively, Theorem 1.5 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 in (1.8) 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. For a d-th order homogeneous chaos, Theorem 1.5 reads as follows:

Corollary 1.6. *Let $X_1, \ldots, X_n$ be independent random variables with $\|X_i\|_{\Psi_\alpha} \leq M$ for some $\alpha \in (0, 2]$ and let A be a symmetric d-tensor with vanishing diagonal and $\|A\|_{\mathrm{HS}} \leq 1$. Then it holds*

$$\mathbb{E} \exp\Big(\frac{1}{C_{d,\alpha} M^\alpha} |f_{d,A}(X)|^{\frac{\alpha}{d}} \Big) \leq 2.$$

Remark 1.7. 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, [27, Lemma A.6] is an immediate consequence of Corollary 1.4 (combined with Lemma A.1 for $f(X, X') = X_i X_i'$). In a similar way, one can deduce [36, Lemma C.4] by applying Corollary 1.4 to the random variable $Z_i \coloneqq X_i Y_i$, 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 [9, Appendix B] can be sharpened for $\alpha \in (0, 1]$ by a more general version of Corollary 1.4 (2), using the same arguments as in [29, Section 3] to treat complex-valued matrices.

1.2 Discussion of related literature

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 [8, Chapter 1].

Starting with linear forms, generalizations to certain classes of random variables as well as multilinear forms of higher degree were shown. 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 or concave, respectively). Two-sided L^p norm estimates for the log-convex case were derived in [15] for linear forms and in [18] for chaoses of all orders. On the other hand, for measures with log-concave tails similar two-sided estimates were proven in [10, 20, 21, 23, 3] under different conditions. Moreover, two-sided estimates for non-negative random variables have been derived in [25] and for chaos of order two in symmetric random variables satisfying the inequality $\|X\|_{2p} \leq A\|X\|_p$ in [26].

Our approach is closer to the work of Adamczak and Wolff [4], where the case of polynomials in sub-Gaussian random variables has been treated. Lastly, let us mention the two results [9, Lemma B.2, Lemma B.3] and [34, Corollary 1.6], proving concentration inequalities for quadratic forms in independent random variables with α-sub-exponential tails.

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-concave 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 $\|X\|_{\Psi_1} < \infty$.

Secondly, 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 $\Psi_{\widetilde{\alpha}}$-norm for $\widetilde{\alpha} \coloneqq (\log_2 \alpha)^{-1}$. 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 [3, 18, 25, 26] are two-sided and require very different tools.

Finally, the two works of Schudy and Sviridenko [32, 33] contain concentration inequalities for polynomials in moment bounded random variables. Therein, a random variable Z is called moment bounded with parameter $L > 0$, if for all $i \geq 1$ it holds $\mathbb{E}|Z|^i \leq iL\,\mathbb{E}|Z|^{i-1}$. Actually, using Stirling's formula, it is easy to see that it 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 in the Bernoulli case $X \sim \mathrm{Ber}(p)$. Considering the case of quadratic forms in random variables X which are moment bounded and centered, one can easily see that (apart from the constants) the bound in Corollary 1.4 (2) is sharper than the corresponding inequality in [33, Theorem 1.1]. Since for log-convex distributions there are two-sided estimates, Corollary 1.4 (2) is sharp in this class. Apart from quadratic forms, due to the different conditions and quantities, it is difficult to compare [33] and Theorem 1.3 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 proofs of the results from Section 1. Lastly, Appendix A contains some elementary properties of the Orlicz norms in particular for $\alpha \in (0,1]$.

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 Corollary 1.4 (2).

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 a random vector X consisting of independent, normalized random variables with α-sub-exponential tails. We give two different forms thereof. The first form is inspired by the results in [29] for the sub-Gaussian case.

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

$$\mathbb{P}\Big(|\|BX\|_2 - \|B\|_{\mathrm{HS}}| \ge t\Big) \le 2\exp\Big(-\frac{1}{C_\alpha M^4}\min\Big(\frac{t^2}{\|B\|_{\mathrm{HS}}^{2-\alpha}\|B\|_{\mathrm{op}}^{\alpha}}, \frac{t^\alpha}{\|B\|_{\mathrm{op}}^{\alpha}}\Big)\Big). \tag{2.1}$$

Proof. First, we additionally assume that $\|B\|_{\mathrm{HS}} = 1$. In this situation, 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

$$\begin{aligned}\mathbb{P}\Big(|\|BX\|_2^2 - 1| \ge t\Big) &\le 2\exp\Big(-\frac{1}{C_\alpha M^4}\min\Big(\frac{t^2}{\|B\|_{\mathrm{op}}^2}, \Big(\frac{t}{\|B\|_{\mathrm{op}}^2}\Big)^{\frac{\alpha}{2}}\Big)\Big)\\ &\le 2\exp\Big(-\frac{1}{C_\alpha M^4\|B\|_{\mathrm{op}}^{\alpha}}\min(t^2, t^{\frac{\alpha}{2}})\Big).\end{aligned} \tag{2.2}$$

Here, in the first step we have used the estimates $\|A\|_{\mathrm{HS}}^2 \le \|B\|_{\mathrm{op}}^2\|B\|_{\mathrm{HS}}^2 = \|B\|_{\mathrm{op}}^2$ and $\|A\|_{\mathrm{op}} \le \|B\|_{\mathrm{op}}^2$ as well as the fact that by Lemma A.2, $\mathbb{E}X_i^2 = 1$ for any i implies $M \ge C_\alpha > 0$, while the second step follows from $\|B\|_{\mathrm{op}} < 1$.

Now, as in [29], we use the inequality $|z-1| \le \min(|z^2-1|, |z^2-1|^{1/2})$, giving for any $t \ge 0$

$$\mathbb{P}\Big(|\|BX\|_2 - 1| \ge t\Big) \le \mathbb{P}\Big(|\|BX\|_2^2 - 1| \ge \max(t, t^2)\Big). \tag{2.3}$$

Hence, a combination of (2.2), (2.3) and $\min(\max(t,t^2)^2, \max(t,t^2)^{\alpha/2}) = \min(t^2, t^\alpha)$ yields

$$\mathbb{P}\Big(|\|BX\|_2 - 1| \ge t\Big) \le 2\exp\Big(-\frac{1}{C_\alpha M^4\|B\|_{\mathrm{op}}^{\alpha}}\min(t^2, t^\alpha)\Big),$$

i.e. (2.1) for $\|B\|_{\mathrm{HS}} = 1$. The general case now follows by considering $\widetilde{B} := B\|B\|_{\mathrm{HS}}^{-1}$, noting that

$$\mathbb{P}\Big(|\|BX\|_2 - \|B\|_{\mathrm{HS}}| \ge t\Big) = \mathbb{P}\Big(|\|\widetilde{B}X\|_2 - 1| \ge t\|B\|_{\mathrm{HS}}^{-1}\Big). \qquad \square$$

The next corollary provides an alternative estimate for $\|BX\|_2^2$:

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

$$||\|BX\|_2^2 - \sum_{i=1}^n \sigma_i^2 \sum_{j=1}^n b_{ji}^2| \le M^2 \max\Big(\sqrt{x}\|A\|_{\mathrm{HS}}, x\|A\|_{\mathrm{op}}, x^{\frac{2+\alpha}{2\alpha}} \max_{i=1,\ldots,n} \|(a_{ij})_j\|_2, x^{\frac{2}{\alpha}} \|A\|_\infty\Big).$$

Let us briefly compare this result to the bound we obtain if we assume $\|X_i\|_{\Psi_2} \le M$ instead. In this case, proceeding in the same way as in the proof below but using the Hanson–Wright inequality (1.3) yields

$$||\|BX\|_2^2 - \sum_{i=1}^n \sigma_i^2 \sum_{j=1}^n b_{ji}^2| \le M^2(\sqrt{x}\|A\|_{\mathrm{HS}} + x\|A\|_{\mathrm{op}})$$

with probability $1 - 2\exp(-x/C)$. In particular, we have similar terms corresponding to $\sqrt{x}$ and x, whereas in the α-sub-exponential case we need two additional terms to account for the heavier tails of its components.

Proof. Define the quadratic form

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

Using Corollary 1.4 (2) with the matrix A gives with probability $1 - 2\exp(-x/C_\alpha)$

$$|Z - \mathbb{E}\, Z| \le M^2 \max\Big(\sqrt{x}\|A\|_{\mathrm{HS}}, x\|A\|_{\mathrm{op}}, x^{\frac{2+\alpha}{2\alpha}} \max_{i=1,\ldots,n} \|A_{i\cdot}\|_2, x^{\frac{2}{\alpha}} \|A\|_\infty\Big).$$

Noting that $\mathbb{E}\, Z = \mathbb{E}\langle X, AX\rangle = \sum_{i=1}^n \sigma_i^2 \sum_{j=1}^n b_{ji}^2$ finishes the proof. □

2.2 Projections of a random vector and distance to a fixed subspace

It is possible to apply Proposition 1.1 to any matrix A associated to an orthogonal projection. In this case, the norms involved can be explicitly calculated, and they do not depend on the structure of the subspace onto which one projects, but merely on its dimension. This leads to the following corollary.

Corollary 2.3. *Let $X_1, \ldots, X_n$ be independent random variables satisfying $\mathbb{E}\, X_i = 0, \mathbb{E}\, X_i^2 = 1$ and $\|X_i\|_{\Psi_\alpha} \le M$ for some $\alpha \in (0,1]$. Furthermore, let $m < n$ and let P denote a random orthogonal projection onto an m-dimensional subspace of $\mathbb{R}^n$, independent of X. For any $x > 0$, with probability at least $1 - 2\exp(-x/C_\alpha)$, we have*

$$\Big|\|PX\|_2^2 - m\Big| \le M^2 \max\Big(\sqrt{xm}, x^{\frac{2}{\alpha}}\Big). \tag{2.4}$$

Proof of Corollary 2.3. This is an application of Proposition 1.1. First, note that we have for any fixed projection P onto an m-dimensional subspace

$$\mathbb{E}_X \|PX\|_2^2 = \mathbb{E}_X \langle X, PX\rangle = \sum_{i,j} P_{ij}\, \mathbb{E}_X\, X_i X_j = \mathrm{tr}(P) = \sum_{i=1}^n \lambda_i(P) = m$$

as well as $\|P\|_{\mathrm{HS}}^2 = \sum_{i=1}^n \lambda_i(P)^2 = m$ and $\|P\|_{\mathrm{op}} = 1$. Consequently, by conditioning on P and using the independence of X and P we have

$$\mathbb{P}\left(|\|PX\|_2^2 - m| \ge t \mid P\right) \le 2\exp\Big(-\frac{1}{C_\alpha}\min\Big(\frac{t^2}{M^4 m}, \Big(\frac{t}{M^2}\Big)^{\frac{\alpha}{2}}\Big)\Big).$$

Finally, it remains to integrate with respect to P. □

A very similar result which follows from Proposition 2.1 is the following variant of [29, 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 $X_1, \ldots, X_n$ be independent random variables satisfying $\mathbb{E} X_i = 0, \mathbb{E} X_i^2 = 1$ and $\|X_i\|_{\Psi_\alpha} \le M$ for some $\alpha \in (0,1]$, and let E be a subspace of $\mathbb{R}^n$ of dimension d. For any $t > 0$ we have*

$$\mathbb{P}\Big(|d(X,E) - \sqrt{n-d}| \ge t\Big) \le 2\exp\Big(-\frac{1}{C_\alpha M^4}\min\Big(\frac{t^2}{(n-d)^{(2-\alpha)/2}}, t^\alpha\Big)\Big).$$

Proof. This follows exactly as in [29, Corollary 3.1] by using Proposition 2.1. □

2.3 Special cases

To pick out one example of random variables with finite Ψ_1 norms, 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

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

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 $\varphi_{|X|}$ in a suitably large neighborhood of 0, it can be used to explicitly calculate the Ψ_1-norm. Indeed, we have $\mathbb{E}\exp(|X|/t) = \varphi_{|X|}(t^{-1})$, and so $\|X\|_{\Psi_1} = 1/\varphi_{|X|}^{-1}(2)$. Moreover, it follows from Lemma A.3 that

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

Thus, as a special case of Corollary 1.4 (2), we obtain the following corollary.

Corollary 2.5. *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 $t \ge 0$*

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\sum_{i,j} a_{ij}(X_i - \lambda_i)(X_j - \lambda_j) - \sum_{i=1}^n a_{ii}\lambda_i\Big| \ge B^2 t\Big) \\
&\le 2\exp\left(-\frac{1}{C}\min\left(\frac{t^2}{\|A\|_{\mathrm{HS}}^2}, \frac{t}{\|A\|_{\mathrm{op}}}, \Big(\frac{t}{\max_i \|(a_{ij})_j\|_2}\Big)^{\frac{2}{3}}, \Big(\frac{t}{\|A\|_\infty}\Big)^{\frac{1}{2}}\right)\right) \\
&\le 2\exp\left(-\frac{1}{C}\min\left(\frac{t^2}{\|A\|_{\mathrm{HS}}^2}, \Big(\frac{t}{\|A\|_{\mathrm{op}}}\Big)^{\frac{1}{2}}\right)\right).
\end{aligned}$$

For Poisson chaos of arbitrary order $d \in \mathbb{N}$, one may derive similar results by evaluating Proposition 1.2 (for $\alpha = 1$). Note though that already for $d = 1$, we lose a logarithmic factor in the exponent.

Let us compare Corollary 2.5 to [16, Section 4], where results for U-statistics of order 2 for Poisson processes are shown (related estimates for higher orders can be found in [1, Section 4]). For simplicity, we assume $\lambda_1 = \ldots = \lambda_n =: \lambda$. Let $(N_t)_{t \ge 0}$ be a Poisson process with intensity λ and compensator $(\lambda t)_{t \ge 0}$, and write $M_t := N_t - \lambda t$. Taking

$$f(x,y) := \sum_{i<j} a_{ij} \mathbb{1}_{(i-1,i]}(x)\mathbb{1}_{(j-1,j]}(y)$$

and $t = n$, the functional Z_t from [16] reads $Z_n = \sum_{i<j} a_{ij}(X_i - \lambda_i)(X_j - \lambda_j)$, i. e. we obtain a quadratic form without diagonal. In a slightly rewritten form, [16, Theorem 4.2] now yields that for any $\varepsilon, u > 0$,

$$\mathbb{P}(Z_n > u) \le 2.77 \exp\left(-\min\left(\frac{u^2}{4(1+\varepsilon)^3 C^2}, \frac{u}{2\eta(\varepsilon)D}, \left(\frac{u}{\beta(\varepsilon)B}\right)^{2/3}, \left(\frac{u}{\gamma(\varepsilon)A}\right)^{1/2}\right)\right).$$

Here, $\eta(\varepsilon)$, $\beta(\varepsilon)$ and $\gamma(\varepsilon)$ are ε-dependent quantities which can be read off [16], while it not hard to relate the quantities A, B, C and D to the norms which appear in Corollary 2.5. Therefore, up to constants, we arrive at a result similar to Corollary 2.5.

Other interesting examples of sub-exponential random variables arise in stochastic geometry. Apart from standard examples like the uniform distribution or more generally log-concave measures on convex bodies, let us mention the cone measure. Indeed, if $K \subseteq \mathbb{R}^n$ is an isotropic, convex body and X is distributed according to the cone measure on K, then $\|\langle X, \theta\rangle\|_{\Psi_1} \le c$ for some constant c and any $\theta \in S^{n-1}$. For the details and the proof we refer to [28, Lemma 5.1].

3 Proofs

To begin with, let us introduce some notation. For any subset $C \subseteq [d]$ with cardinality $|C| > 1$, we may introduce the "generalized diagonal" of $[n]^d$ with respect to C by

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

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

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

To prove Theorem 1.3, we need a number of auxiliary results. The first group of them provides L^p norm estimates. In the proof of [4, Theorem 1.4], moments of sums of sub-Gaussian random variables are compared to moments of Gaussian random variables (cf. (3.2) below). A similar result is needed for α-sub-exponential random variables, but here we have to replace Gaussian by symmetric Weibull variables with shape parameter α (and scale parameter 1), i. e. symmetric random variables w with $\mathbb{P}(|w| \ge t) = \exp(-t^\alpha)$. Of course, in particular we have $\|w\|_{\Psi_\alpha} < \infty$. We now have the following lemma:

Lemma 3.1. *For any $k \in \mathbb{N}$, any $\alpha > 0$ and any $p \ge 1$ the following holds. For any set of independent, symmetric random variables $Y_1, \ldots, Y_n$ satisfying $\|Y_i\|_{\Psi_{\frac{\alpha}{k}}} \le M$ we have*

$$\Big\| \sum_{i=1}^n a_i Y_i \Big\|_p \le 2 c_{\alpha,k} M \Big\| \sum_{i=1}^n a_i w_{i,1} \cdots w_{i,k} \Big\|_p,$$

where $w_{i,j}$ are symmetric i. i. d. Weibull random variables with shape parameter α and $c_{\alpha,k} := (k/(1 - \log(2)))^{k/\alpha}$.

Recall that if $\alpha = 2$, by [4, Lemma 5.4] we have

$$\Big\| \sum_{i=1}^n a_i Y_i \Big\|_p \le C_k M \Big\| \sum_{i=1}^n a_i g_{i,1} \cdots g_{i,k} \Big\|_p, \tag{3.2}$$

where $g_{i,j}$ are independent standard Gaussians. In particular, this allows to prove analogues of Proposition 1.2 and Theorem 1.3 if $\alpha = 2/q$ for some $q \in \mathbb{N}$ solely based on the use of Gaussian random variables. We will not pursue this idea further in this note.

Proof of Lemma 3.1. Due to homogeneity we assume $M = 1$, and for brevity we set $c := c_{\alpha,k}$. By Markov's inequality we have $\mathbb{P}(|Y_i| \ge t) \le 2\exp(-t^{\alpha/k})$ for any $i \in [n]$ and all $t \ge 0$.

The inclusion $\{c^{1/k}|w_{i,1}| \ge t^{1/k}, \ldots, c^{1/k}|w_{i,k}| \ge t^{1/k}\} \subseteq \{c|w_{i,1}\cdots w_{i,k}| \ge t\}$ holds for any $i \in [n]$ and $t \ge 0$. This yields for all $t \ge 1$

$$\begin{aligned} \mathbb{P}(c|w_{i,1}\cdots w_{i,k}| \ge t) &\ge \prod_{j=1}^k \mathbb{P}(c^{1/k}|w_{i,j}| \ge t^{1/k}) = \exp\Big(-k\Big(\frac{t}{c}\Big)^{\alpha/k}\Big) \\ &= \exp\Big(-(1-\log(2))t^{\alpha/k}\Big) \ge 2\exp\Big(-t^{\alpha/k}\Big) \\ &\ge \mathbb{P}(|Y_i| \ge t), \end{aligned}$$

where the second inequality requires the condition $t \ge 1$. Now the rest follows exactly as in [4, Proof of Lemma 5.4].

Alternatively, one can extend the inequality to all $t \ge 0$ by multiplying the left hand side by a constant. Indeed, it is easy to see (by observing $\mathbb{P}(c|w_{i,1}\cdots w_{i,k}| \ge 1) \ge 2/e$) that for all $t \ge 0$ it holds

$$\mathbb{P}(|Y_i| \ge t) \le \frac{e}{2}\,\mathbb{P}(c|w_{i,1}\cdots w_{i,k}| \ge t).$$

Thus, the contraction principle [19, Theorem 1] tells us that for any $p \ge 1$ we have

$$\Big\| \sum_{i=1}^n a_i Y_i \Big\|_p \le \frac{e}{2} c_{\alpha,k} \Big\| \sum_{i=1}^n a_i w_{i,1}\cdots w_{i,k} \Big\|_p. \qquad \square$$

Moreover, we shall need estimates for the L^p norms of multilinear forms in Weibull random variables. Adapting [18, Example 3] yields the following lemma:

Lemma 3.2. *Let $A = (a_{\mathbf{i}})_{\mathbf{i}\in[n]^d}$ be a d-tensor and (w_i^j), $i \le n$, $j \le d$, an array of i. i. d. Weibull variables with shape parameter $\alpha \in (0,1]$. Then, for every $p \ge 2$,*

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

Note that a similar result for Weibull variables with shape parameter $\alpha > 1$ is not known yet if $d > 3$, which explains the restriction to $\alpha \in (0,1]$ in this article. The case of $\alpha \in [1,2]$ and polynomials of total degree $D \le 3$ has been discussed in [4, Proposition 6.2].

In the proof of Theorem 1.3, we actually show L^p estimates for $f(X)$. The following proposition provides the link to concentration inequalities. Results of this type are by now standard, and we cite them in the form given in [31] with some smaller modifications to address the situation considered in the present note.

Proposition 3.3. *Assume that a random variable Z satisfies for every $p \geq 2$*

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

for some $\nu \in \mathbb{N}$, some constants $C_1, \ldots, C_\nu \geq 0$ and some exponents $\gamma_k \in (0, \infty)$ for which we assume $\gamma_1 \leq \cdots \leq \gamma_\nu$. Let $L \coloneqq |\{k \colon C_k > 0\}|$ and $r \coloneqq \min\{k \in \{1, \ldots, \nu\} \colon C_k > 0\}$. Then, for any $t \geq 0$

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

In the second group of auxiliary results, we discuss some properties of the norms $\|A\|_{\mathcal{J}}$. To this end, recall the Hadamard product of two d-tensors A, B given by $A \circ B \coloneqq (a_{\mathbf{i}} b_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$. For any $C \subset [n]^d$, we may define "indicator tensors" 1_C 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

$$\|A \circ 1_C\|_{\mathcal{J}} \leq \|A\|_{\mathcal{J}} \tag{3.3}$$

in general. However, [4, Lemma 5.2 and Corollary 5.3] show a number of situations in which such an inequality does hold, e. g. "generalized rows" or "generalized diagonals" as well as certain sets $L(\mathcal{K})$. Here, for any partition $\mathcal{K} = \{K_1, \ldots, K_\nu\}$ of $[d]$ we define

$$L(\mathcal{K}) = \{\mathbf{i} \in [n]^d \colon i_k = i_l \Leftrightarrow \exists j \colon k, l \in K_j\}. \tag{3.4}$$

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

We need to extend these results to the "restricted" tensors $A_{\mathbf{i}_{I^c}}$. That is, we examine whether (a modification of) the inequality

$$\|(A \circ 1_C)_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \leq \|A_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \tag{3.5}$$

still holds in the situation where $\mathcal{J}$ is a partition of I^c. Additionally, we show an analogue of [4, Lemma 5.1].

Lemma 3.4. *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 $C = \{\mathbf{i} \colon i_{k_1} = j_1, \ldots, i_{k_l} = j_l\}$ for some $1 \leq k_1 < \ldots < k_l \leq d$ ("generalized row"), then (3.5) holds.*
2. *If $C = \{\mathbf{i} \colon i_k = i_l \ \forall k, l \in K\}$ for some $K \subset [d]$ ("generalized diagonal"), then (3.5) holds.*
3. *If $C_1, C_2 \subset [n]^d$ are such that (3.5) holds for any d-tensor A, then so is $C_1 \cap C_2$.*
4. *If $\mathcal{K} \in P_d$, then $\|(A \circ 1_{L(\mathcal{K})})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \leq 2^{|\mathcal{K}|(|\mathcal{K}|-1)/2} \|A_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}$.*
5. *For any vectors $v_1, \ldots, v_d \in \mathbb{R}^n$, $\|(A \circ \otimes_{i=1}^d v_i)_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \leq \|A_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \prod_{i=1}^d \|v_i\|_\infty$.*

Proof. To see (1), we may assume that $\{k_1, \ldots, k_l\} \cap I = \emptyset$ (note that if $\{k_1, \ldots, k_l\} \cap I \neq \emptyset$, either the conditions are not compatible, in which case $(A \circ 1_C)_{\mathbf{i}_{I^c}} = 0$, or we can remove some of the conditions and obtain a subset with $\{k_1, \ldots, k_{\tilde{l}}\} \cap I = \emptyset$). 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, proving (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 $K \cap I \neq \emptyset$, then either $(A \circ 1_C)_{\mathbf{i}_{I^c}} = 0$ or $(A \circ 1_C)_{\mathbf{i}_{I^c}} = A_{\mathbf{i}_{I^c}} \circ 1_{C'}$ for some generalized *row* C' in I^c, readily leading to (2) again.

(3) is clear. To see (4), observe that $1_{L(\mathcal{K})}$ is the indicator matrix of a set C which can be written as an intersection of $|\mathcal{K}|$ generalized diagonals and $|\mathcal{K}|(|\mathcal{K}|-1)/2$ sets of the form $\{\mathbf{i} : i_k \neq i_l\}$ for $k < l$. Recall that

$$\|(B \circ 1_{\{i_k \neq i_l\}})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} = \|(B - B \circ 1_{\{i_k = i_l\}})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \leq 2\|B_{\mathbf{i}_{I^c}}\|_{\mathcal{J}},$$

using (2) in the last step. As a consequence, the claim follows by applying (2) again and an interation of (3). Finally, noting that

$$\begin{aligned}\|(A \circ \otimes_{i=1}^d v_i)_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} &= \sup\Big\{\sum_{\mathbf{i}_{I^c}} a_{\mathbf{i}} (\otimes_{i=1}^d v_i)_{\mathbf{i}} \prod_{j=1}^k x^{(j)}_{\mathbf{i}_{J_j}} : \|x^{(j)}_{\mathbf{i}_{J_j}}\|_2 \leq 1\Big\} \\ &\leq \sup\Big\{\sum_{\mathbf{i}_{I^c}} a_{\mathbf{i}} \prod_{j=1}^k x^{(j)}_{\mathbf{i}_{J_i}} : \|x^{(j)}_{\mathbf{i}_{J_j}}\|_2 \leq 1\Big\} \prod_{i=1}^d \|v_i\|_\infty = \|A_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \prod_{i=1}^d \|v_i\|_\infty,\end{aligned}$$

we arrive at (5). $\square$

We are now ready to prove Theorem 1.3. Before we start, let us give some final definitions. For any multiindex $\mathbf{i}$ let $|\mathbf{i}| \coloneqq \sum_j i_j$. For the sake of brevity we define

$$\begin{aligned} I_{m,d} &\coloneqq \{(i_1, \ldots, i_m) \in \mathbb{N}^m : |\mathbf{i}| = d\}, \\ I_{m,\leq d} &\coloneqq \{(i_1, \ldots, i_m) \in \mathbb{N}^m : |\mathbf{i}| \leq d\}, \end{aligned}$$

where by convention $0 \notin \mathbb{N}$. Given two multiindices $\mathbf{k}, \mathbf{l}$ 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 $f \leq C_{D,\alpha} g$.

Proof of Theorem 1.3. The proof works by finding suitable estimates for the L^p norms $\|f(X) - \mathbb{E}f(X)\|_p$, from which we derive concentration bounds using Proposition 3.3. We may assume $M = 1$. For the general case, given random variables $X_1, \ldots, X_n$ with $\|X_i\|_{\Psi_\alpha} \leq M$, define $Y_i \coloneqq M^{-1} X_i$. The polynomial $f = f(X)$ can be written as a polynomial $\widetilde{f} = \widetilde{f}(Y)$ 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 $\partial_{i_1 \ldots i_j} \widetilde{f}(Y) = M^j \partial_{i_1 \ldots i_j} f(X)$.

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

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

where the constants satisfy $c^{(d)}_{(i_1,k_1),\ldots,(i_\nu,k_\nu)} = c^{(d)}_{(i_{\pi_1},k_{\pi_1}),\ldots,(i_{\pi_\nu},k_{\pi_\nu})}$ for any permutation $\pi \in \mathcal{S}_\nu$. As in [4], by rearranging and making use of the independence of $X_1, \ldots, X_n$, 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

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

Step 2. Let $X^{(1)}, \ldots, X^{(d)}$ be independent copies of the random vector X. Take a set of i. i. d. Rademacher variables $(\varepsilon_i^{(j)})$, $i \le n$, $j \le d$, which are independent of the $(X^{(j)})_j$. By standard decoupling and symmetrization inequalities (see [8, Theorem 3.1.1] and [8, Lemma 1.2.6]) we have

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

Noting that $\|X_i^k\|_{\Psi_{\alpha/k}} = \|X_i\|_{\Psi_\alpha}^k \le 1$ (cf. Appendix A), an iteration of Lemma 3.1 hence leads to

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

Here, $(w_{i,k}^{(j)})$ is an array of i. i. d. symmetric Weibull variables with shape parameter α.

Moreover, the family $(a_{\mathbf{i}}^{\mathbf{k}})_{\nu\in\{1,\ldots,d\},\mathbf{k}\in I_{\nu,d},\mathbf{i}\in[n]^{\underline{\nu}}}$ 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 $\nu \in \{1, \ldots, d\}$ of distinct elements $j_1, \ldots, j_\nu$ with each j_l appearing exactly k_l times in $\mathbf{i}$. Consequently, we set $a_{i_1\ldots i_d} := a_{j_1,\ldots,j_\nu}^{(k_1,\ldots,k_\nu)}$, 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 definition of $L(\mathcal{K})$ (3.4), rewriting and applying Lemma 3.2 together with Lemma 3.4 (4) yields

$$\begin{aligned}\|f(X) - \mathbb{E}f(X)\|_p &\lesssim \sum_{d=1}^{D}\sum_{\nu=1}^{d}\sum_{\mathbf{k}\in I_{\nu,d}} \|\langle A_d \circ 1_{L(\mathcal{K}(k_1,\ldots,k_\nu))}, \otimes_{j=1}^{\nu}\otimes_{k=1}^{k_j}(w_{i,k}^{(j)})_{i\le n}\rangle\|_p \\ &\lesssim \sum_{d=1}^{D}\sum_{\nu=1}^{d}\sum_{\mathbf{k}\in I_{\nu,d}}\sum_{I\subset[d]}\sum_{\mathcal{J}\in P(I^c)} p^{|I|/\alpha+|\mathcal{J}|/2}\max_{\mathbf{i}_I}\|(A_d \circ 1_{L(\mathcal{K}(k_1,\ldots,k_\nu))})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \\ &\lesssim \sum_{d=1}^{D}\sum_{I\subset[d]}\sum_{\mathcal{J}\in P(I^c)} p^{|I|/\alpha+|\mathcal{J}|/2}\max_{\mathbf{i}_I}\|(A_d)_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}.\end{aligned}$$

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

$$\begin{aligned}&\mathbb{E}\frac{\partial^d f}{\partial x_{i_1}\ldots\partial x_{i_d}}(X) = \sum_{\mathbf{k}:\mathbf{k}\ge\mathbf{l}}\sum_{m=\nu}^{D}\sum_{\substack{k_{\nu+1},\ldots,k_m>0\\ k_1+\ldots+k_m\le D}}\sum_{\substack{j_{\nu+1},\ldots,j_m\\ (j_1,\ldots,j_m)\in[n]^{\underline{m}}}} \\ &\left(\binom{m}{\nu}\nu!c_{(j_1,k_1),\ldots,(j_m,k_m)}^{(k_1+\ldots+k_m)}\prod_{\beta=1}^{\nu}\mathbb{E}X_{j_\beta}^{k_\beta-l_\beta}\prod_{\beta=\nu+1}^{m}\mathbb{E}X_{j_\beta}^{k_\beta}\prod_{\beta=1}^{\nu}\frac{k_\beta!}{(k_\beta-l_\beta)!}\right) \\ &= \nu!l_1!\cdots l_\nu!a_{\mathbf{i}} + R_{\mathbf{i}}^{(d)},\end{aligned}$$

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

$$\mathbb{E}\frac{\partial^D f}{\partial x_{i_1}\ldots\partial x_{i_D}}(X) = \nu!l_1!\cdots l_\nu!a_{\mathbf{i}} = \nu!|I_1|!\cdots|I_\nu|!a_{\mathbf{i}}, \tag{3.6}$$

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 $I \subset [D]$ and any partition $\mathcal{J} \in P(I^c)$,

$$\begin{aligned} \|(A_D)_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} &\le \sum_{\mathcal{K} \in P_D} \|(A_D \circ 1_{L(\mathcal{K})})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \le \sum_{\mathcal{K} \in P_D} \|(\mathbb{E} f^{(D)}(X) \circ 1_{L(\mathcal{K})})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \\ &\lesssim \|(\mathbb{E} f^{(D)}(X))_{\mathbf{i}_{I^c}}\|_{\mathcal{J}}, \end{aligned} \tag{3.7}$$

using the partition of unity $1 = \sum_{\mathcal{K} \in P_D} 1_{L(\mathcal{K})}$ and the triangle inequality in the first, equation (3.6) in the second and Lemma 3.4 (4) in the last step.

The proof is now completed by induction. More precisely, in the next step we show that for any $d = 1, \ldots, D-1$, any $I \subset [d]$ and any partitions $\mathcal{I} \in P_d$, $\mathcal{J} \in P([d] \backslash I)$,

$$\|(R^{(d)})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \lesssim \sum_{k=d+1}^{D} \sum_{\substack{\mathcal{K} \in P([k] \backslash I) \\ |\mathcal{K}| \ge |\mathcal{J}|}} \|(A_k)_{\mathbf{i}_{I^c}}\|_{\mathcal{K}}. \tag{3.8}$$

Actually, one can see below that it is possible to restrict the second sum to partitions $\mathcal{K}$ with $|\mathcal{K}| \in \{|\mathcal{J}|, |\mathcal{J}|+1\}$. Once having proven (3.8), it follows from reverse induction that

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

Here, performing the induction, we consider sums $\sum_{d=k}^{D}$ and proceed from $k = D$ to $k = 1$. The base case $k = D$ then immediately follows from (3.7). In the induction step we moreover use that for any $p \ge 2$ and any $|\mathcal{K}| \ge |\mathcal{J}|$ we have $p^{|\mathcal{J}|/2} \le p^{|\mathcal{K}|/2}$. In view of Step 2 and Proposition 3.3, this finishes the proof.

Step 4. Instead of (3.8), we actually show

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

from which (3.8) easily follows (cf. the arguments used in (3.7)). To this end, let us analyze the "remainder tensors" $R^{(d)}$ in more detail. Fix $d \in \{1, \ldots, D-1\}$, $I \subset [d]$, $\mathbf{i}_I$ and partitions $\mathcal{I} = \{I_1, \ldots, I_\nu\} \in P_d$, $\mathcal{J} = \{J_1, \ldots, J_\mu\} \in P([d] \backslash I)$. Here we assume $\mathcal{I}$ to be "admissible" in the sense that there exists some $\mathbf{j} \in L(\mathcal{I})$ such that $\mathbf{j}_I = \mathbf{i}_I$ (otherwise $(R^{(d)} \circ 1_{L(\mathcal{I})})_{\mathbf{i}_{I^c}} = 0$). For instance, if $d = 3$, $I = \{1, 2\}$ and $\mathbf{i}_I = (1, 1)$, $\mathcal{I} = \{\{1, 2, 3\}\}$ is admissible but $\mathcal{I} = \{\{1\}, \{2, 3\}\}$ is not. Moreover, let $\mathbf{l}$ be the vector with $l_\beta \coloneqq |I_\beta|$ (which implies $|\mathbf{l}| = d$).

For any $\mathbf{k} \in I_{\nu, \le D}$ 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:

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

Here, for $\beta \le \nu$ we denote by j_β the value of $\mathbf{i}$ on the level set I_β. Clearly,

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

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

$$\|(S_{\mathcal{I}}^{(d,\mathbf{k})})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \lesssim \|(A_{|\mathbf{k}|})_{\mathbf{i}_{I^c}}\|_{\mathcal{K}} \tag{3.9}$$

for any $\mathbf{i}_I$. To this end, we introduce an auxiliary tensor which is given by an appropriate embedding of $S_{\mathcal{I}}^{(d,\mathbf{k})}$. Choose any partition $\widetilde{\mathcal{I}} = \{\widetilde{I}_1, \ldots, \widetilde{I}_\nu\} \in P_{|\mathbf{k}|}$ with $|\widetilde{I}_\beta| = k_\beta$ and $I_\beta \subset \widetilde{I}_\beta$ for all β. Embedding the d-tensor $S_{\mathcal{I}}^{(d,\mathbf{k})}$ into the space of $|\mathbf{k}|$-tensors is done by defining a new tensor $\widetilde{S}^{|\mathbf{k}|} = (\widetilde{s}_{\mathbf{i}}^{|\mathbf{k}|})_{\mathbf{i}}$ given by

$$\widetilde{s}_{\mathbf{i}}^{|\mathbf{k}|} = s_{\mathbf{i}_{[d]}}^{(d)} 1_{\mathbf{i} \in L(\widetilde{\mathcal{I}})}. \tag{3.10}$$

We now define the partition $\mathcal{K} = \{K_1, \ldots, K_{\mu+1}\}$ 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 $\widetilde{\mathcal{I}}$ is a partition of $|\mathbf{k}|$, there is a unique $k \in \{1, \ldots, \nu\}$ such that $r \in \widetilde{I}_k$. Take the smallest element $t =: \pi(r)$ in $\widetilde{I}_k$ (since $I_k \subset \widetilde{I}_k$, 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 $\mathcal{K}$ for $d = 6$, $|\mathbf{k}| = 8$, $I = \{2, 3\}$, $\mathcal{J} = \{\{1\}, \{4\}, \{5, 6\}\}$ and $\widetilde{\mathcal{I}} = \{\{1, 2, 5\}, \{3, 6, 8\}, \{4, 7\}\}$. In both figures we used shapes to encode the partitions under consideration. In the first step (left figure), we assign all elements of $\mathcal{J}$ to $\mathcal{K}$. In the second step (right figure) we use the partition $\widetilde{\mathcal{I}}$ to assign the elements 7 and 8. Here, $\pi(7) = 4$ and $\pi(8) = 3$, so that 7 is added to K_2 (see the left figure) and 8 is added to K_4 as $3 \in I$.

First off, we claim

$$\|(S_{\mathcal{I}}^{(d,|\mathbf{k}|)})_{\mathbf{i}_{I^c}}\|_{\mathcal{J}} \le \|(\widetilde{S}^{|\mathbf{k}|})_{\mathbf{i}_{I^c}}\|_{\mathcal{K}}. \tag{3.11}$$

To see (3.11), let $x = (x^{(\beta)}) = ((x_{\mathbf{i}_{J_\beta}}^{(\beta)}))$ be such that $\max_{\beta=1,\ldots,\mu} \|x^{(\beta)}\|_2 \le 1$. Based on this, we define vectors $y = (y^{(\beta)})_{\beta=1,\ldots,\mu+1}$ via

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

As $y^{(\mu+1)}$ only has a single non-zero element, it is easy to see that $\max_{i=1,\ldots,\mu+1} \|y^{(\beta)}\|_2 \le 1$. Moreover, by the definition of the matrix $\widetilde{S}^{|k|}$ and the fact that if $\mathbf{i} \in L(\widetilde{\mathcal{I}})$, then for $r > d$, $i_r = i_{\pi(r)}$, which implies $y_{\mathbf{i}_{K_\beta}}^{(\beta)} = x_{\mathbf{i}_{K_\beta \cap [d]}}^{(\beta)} = x_{\mathbf{i}_{J_\beta}}^{(\beta)}$ for $\beta \le \mu$ as well as $y_{\mathbf{i}_{K_{\mu+1}}}^{(\mu+1)} = 1$ we have

$$\langle (S^{(d,\mathbf{k})})_{\mathbf{i}_{I^c}}, \bigotimes_{\beta=1}^{\mu} x^{(\beta)} \rangle = \langle (\widetilde{S}^{(|k|)})_{\mathbf{i}_{I^c}}, \bigotimes_{\beta=1}^{\mu+1} y^{(\beta)} \rangle. \tag{3.12}$$

Hence, the supremum on the left hand side of (3.11) is taken over a subset of the unit ball with respect to $\max_{i=1,\ldots,\mu+1} \|x^{(\beta)}\|_2$.

Finally, it remains to prove

$$\|(\widetilde{S}^{|\mathbf{k}|})_{\mathbf{i}_{I^c}}\|_{\mathcal{K}} \lesssim \|(A_{|\mathbf{k}|})_{\mathbf{i}_{I^c}}\|_{\mathcal{K}} \tag{3.13}$$

for any partition $\mathcal{K} \in P(I^c)$. To see this, note that $\mathbf{i} \in L(\widetilde{\mathcal{I}})$ implies $\widetilde{s}_{\mathbf{i}}^{|\mathbf{k}|} = a_{\mathbf{i}}^{|\mathbf{k}|} \prod_{\beta=1}^{\nu} \mathbb{E} X_{j_\beta}^{k_\beta - l_\beta}$. As a consequence,

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

where the vectors v_β are defined by $v_\beta = (\mathbb{E}X_i^{k_\beta - l_\beta})_{i\le n}$ if $\beta \in \{\min I_1, \ldots, \min I_\nu\}$ and $v_\beta = (1,\ldots,1)$, otherwise. In particular, recalling $M = 1$ we always have $\|v_\beta\|_\infty \lesssim 1$, and therefore, by Lemma 3.4 (5),

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

from where we easily arrive at (3.13) by applying Lemma 3.4 (4).

Combining (3.11) and (3.13) yields (3.9), which finishes the proof. □

It remains to prove the corollary-type results presented in Section 1.

Proof of Proposition 1.1 and Corollary 1.4. Corollary 1.4 follows directly from Theorem 1.3 by calculating the derivatives. Moreover, Proposition 1.1 follows directly from Corollary 1.4 (2). □

Proof of Theorem 1.5. First let $\alpha \in (0,1]$ and consider the bound given by Theorem 1.3. 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

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

(using (1.6)) as well as

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

If $t/(M^d\|\mathbb{E}f^{(d)}(X)\|_{\mathrm{HS}}) \ge 1$, this immediately yields the result. Otherwise, note that the tail bound given in Theorem 1.5 is trivial. (In fact, here one needs to ensure that $C_{D,\alpha}$ is sufficiently large, e. g. $C_{D,\alpha} \ge 1$. It is not hard to see that in general this condition will be satisfied anyway.)

To prove the result for $\alpha \in (1,2]$, we need to modify the proof of Theorem 1.3. Here we only provide a sketch, since many of the arguments can be easily adapted. We continue working with symmetric Weibull variables with shape parameter α. As Lemma 3.1 holds true for any $\alpha > 0$, the central task is to find a replacement for Lemma 3.2 (or more precisely, an upper bound on $\|\langle A, w^1 \otimes \ldots \otimes w^d\rangle\|_p$). Here we begin with the case of $d = 1$. Using the notation from the lemma, we may deduce from [20, Theorem 1] that

$$\|\langle A, w\rangle\|_p \le C(\|\langle A, w\rangle\|_1 + \|A\|_{\mathcal{N},p}),$$

where, writing $x_i = x_i I(|x_i| \le 1) + x_i I(|x_i| > 1) =: \widetilde{x}_i + \hat{x}_i$,

$$\begin{aligned}
\|A\|_{\mathcal{N},p} &= \sup\Big\{\sum_{i=1}^n a_i x_i \colon \sum_{i=1}^n \min(x_i^2, |x_i|^\alpha) \le p\Big\} \\
&\le \sup\Big\{\sum_{i=1}^n a_i \widetilde{x}_i \colon \|\widetilde{x}\|_2 \le p^{1/2}\Big\} + \sup\Big\{\sum_{i=1}^n a_i \hat{x}_i \colon \|\hat{x}\|_\alpha \le p^{1/\alpha}\Big\} \\
&\le 2p^{1/\alpha} \sup\Big\{\sum_{i=1}^n a_i x_i \colon \|x\|_2 \le 1\Big\} = 2p^{1/\alpha}\|A\|_2,
\end{aligned}$$

which together with $\|\langle A, w\rangle\|_1 \le \|\langle A, w\rangle\|_2 \le C_\alpha \|A\|_2$ leads to

$$\|\langle A, w\rangle\|_p \le C_\alpha p^{1/\alpha}\|A\|_2 \tag{3.14}$$

for any $p \ge 1$. This may be iterated to arrive at

$$\|\langle A, w^1 \otimes \ldots \otimes w^d\rangle\|_p \le C_\alpha^d p^{d/\alpha}\|A\|_{\mathrm{HS}} \tag{3.15}$$

for any d-tensor A and any $p \geq 2$ with the same constant C_α as in (3.14). Indeed, assuming we have proven (3.15) up to order $d-1$, we obtain that

$$\begin{aligned}\Big\| \sum_{i_1,\dots,i_d} a_{i_1\dots i_d} w_{i_1}^{(1)} \cdots w_{i_d}^{(d)} \Big\|_p^2 &\leq C_\alpha^{2(d-1)} p^{2(d-1)/\alpha} \Big\| \big(\sum_{i_1,\dots,i_{d-1}} \big(\sum_{i_d=1}^n a_{i_1\dots i_d} w_{i_d}^{(d)} \big)^2 \big)^{1/2} \Big\|_p^2 \\ &= C_\alpha^{2(d-1)} p^{2(d-1)/\alpha} \Big\| \sum_{i_1,\dots,i_{d-1}} \big(\sum_{i_d=1}^n a_{i_1\dots i_d} w_{i_d}^{(d)} \big)^2 \Big\|_{p/2} \\ &\leq C_\alpha^{2(d-1)} p^{2(d-1)/\alpha} \sum_{i_1,\dots,i_{d-1}} \Big\| \big(\sum_{i_d=1}^n a_{i_1\dots i_d} w_{i_d}^{(d)} \big)^2 \Big\|_{p/2} \\ &= C_\alpha^{2(d-1)} p^{2(d-1)/\alpha} \sum_{i_1,\dots,i_{d-1}} \Big\| \sum_{i_d=1}^n a_{i_1\dots i_d} w_{i_d}^{(d)} \Big\|_p^2 \\ &\leq C_\alpha^{2d} p^{2d/\alpha} \sum_{i_1,\dots,i_{d-1}} \| a_{i_1\cdots i_{d-1}\cdot} \|_2^2 = C_\alpha^{2d} p^{2d/\alpha} \|A\|_{\mathrm{HS}}^2.\end{aligned}$$

Here we have used the assumption in the first step and (3.14) in the last inequality. This establishes (3.15). Following the proof of Theorem 1.3 using (3.15), the conclusion of Step 2 (in particular recalling the d-tensor A_d introduced there) reads

$$\|f(X) - \mathbb{E}f(X)\|_p \lesssim \sum_{d=1}^D p^{d/\alpha} \|A_d\|_{\mathrm{HS}}.$$

The rest of the proof is easily adapted.

From here, the exponential moment bound follows by standard arguments, see for example [6, Proof of Theorem 1.1]. □

A Properties of Orlicz quasinorms

As mentioned in the introduction, the Orlicz norms (1.5) satisfy the triangle inequality only for $\alpha \geq 1$. However, for $\alpha \in (0,1)$ it is still 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.

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

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

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

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

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

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 $\alpha > 0$ define

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

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

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

For $\alpha \geq 1$, we obtain α-independent lower and upper bounds on d_α and D_α, i. e. $d_\alpha \geq 1/2$ and $D_\alpha \leq 2e$, agreeing with the bounds proven in [5, Section 8]. In the proof below, 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

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

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 $e^y y = y^{1/\alpha} \alpha^{-1} (\alpha e)^{1/\alpha}$, 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 $\alpha > 0$

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

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

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

For the second inequality in (A.2), again assume that $\sup_{p \geq 1} \frac{\|X\|_p}{p^{1/\alpha}} = 1$. First, we consider the case of $\alpha < 1$ and extend the supremum to $p \in [\alpha, \infty)$ as follows. For any $p \in [\alpha, 1)$ we have

$$\frac{\|X\|_p}{p^{1/\alpha}} \leq \frac{\|X\|_1}{p^{1/\alpha}} \leq \frac{1}{p^{1/\alpha}} \leq \frac{1}{\alpha^{1/\alpha}} \Longrightarrow \sup_{p \geq \alpha} \frac{\|X\|_p}{p^{1/\alpha}} \leq \frac{1}{\alpha^{1/\alpha}}.$$

Now, by Taylor's expansion and using the inequality $n^n \leq e^n n!$ 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, proving the assertion. These arguments can easily be adapted to the case of $\alpha \geq 1$. $\square$

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

$$\|X+Y\|_{\Psi_\alpha} \leq K_\alpha \left(\|X\|_{\Psi_\alpha} + \|Y\|_{\Psi_\alpha}\right), \tag{A.4}$$

$$\|\mathbb{E}\, X\|_{\Psi_\alpha} \leq \frac{1}{d_\alpha (\log 2)^{1/\alpha}} \|X\|_{\Psi_\alpha}, \tag{A.5}$$

$$\|X - \mathbb{E}\, X\|_{\Psi_\alpha} \leq K_\alpha \left(1 + (d_\alpha \log 2)^{-1/\alpha}\right) \|X\|_{\Psi_\alpha}, \tag{A.6}$$

where $K_\alpha \coloneqq 2^{1/\alpha}$ if $\alpha \in (0,1)$ and $K_\alpha = 1$ if $\alpha \geq 1$.

Proof. First assume $\alpha \in (0,1)$, let $K \coloneqq \|X\|_{\Psi_\alpha}$ and $L \coloneqq \|Y\|_{\Psi_\alpha}$ and define $t \coloneqq 2^{1/\alpha}(K + L)$. We have

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

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.

For $\alpha \geq 1$ it is easy to see that $\psi(x) = \exp(x^\alpha) - 1$ satisfies $\psi(0) = 0$, is convex and non-decreasing on $[0, \infty)$. As a consequence, for $K \coloneqq \|X\|_{\Psi_\alpha}$ and $L \coloneqq \|Y\|_{\Psi_\alpha}$ we have

$$\psi\Big(\frac{|X+Y|}{K+L}\Big) \leq \psi\Big(\frac{|X|+|Y|}{K+L}\Big) \leq \frac{K}{K+L}\psi\Big(\frac{|X|}{K}\Big) + \frac{L}{K+L}\psi\Big(\frac{|Y|}{L}\Big),$$

and it remains to integrate this with respect to $\mathbb{P}$.

To see (A.5), assuming $\|X\|_{\Psi_\alpha} < \infty$, an application of Lemma A.2 gives

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

From here, (A.6) follows readily. □

References

[1] Radosław Adamczak, *Moment inequalities for U-statistics*, Ann. Probab. **34** (2006), no. 6, 2288–2314. MR-2294982

[2] Radosław Adamczak, *A note on the Hanson–Wright inequality for random vectors with dependencies*, Electron. Commun. Probab. **20** (2015), no. 72, 13. MR-3407216

[3] Radosław Adamczak and Rafał Latała, *Tail and moment estimates for chaoses generated by symmetric random variables with logarithmically concave tails*, Ann. Inst. Henri Poincaré Probab. Stat. **48** (2012), no. 4, 1103–1136. MR-3052405

[4] Radosław Adamczak and Paweł Wolff, *Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order*, Probab. Theory Related Fields **162** (2015), no. 3-4, 531–586. MR-3383337

[5] Sergey G. Bobkov, *The growth of L^p-norms in presence of logarithmic Sobolev inequalities*, Vestnik Syktyvkar Univ. **11** (2010), no. 2, 92–111.

[6] Sergey G. Bobkov, Friedrich Götze, and Holger Sambale, *Higher order concentration of measure*, Commun. Contemp. Math. **21** (2019), no. 03. MR-3947067

[7] Sergey G. Bobkov and Michel Ledoux, *Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution*, Probab. Theory Related Fields **107** (1997), no. 3, 383–400. MR-1440138

[8] Víctor H. de la Peña and Evarist Giné, *Decoupling*, Probability and its Applications, Springer-Verlag, New York, 1999. MR-1666908

[9] László Erdős, Horng-Tzer Yau, and Jun Yin, *Bulk universality for generalized Wigner matrices*, Probab. Theory Related Fields **154** (2012), no. 1-2, 341–407. MR-2981427

[10] Efim D. Gluskin and Stanisław Kwapień, *Tail and moment estimates for sums of independent random variables with logarithmically concave tails*, Studia Math. **114** (1995), no. 3, 303–309. MR-1338834

[11] Friedrich Götze and Holger Sambale, *Higher order concentration in presence of Poincaré-type inequalities*, High Dimensional Probability VIII, Birkhäuser, Springer, Progress in Probability **74** (2019), pp. 55–69.

[12] Friedrich Götze, Holger Sambale, and Arthur Sinulis, *Concentration inequalities for bounded functionals via generalized log-Sobolev inequalities*, J. Theor. Probab. (2020)

[13] Friedrich Götze, Holger Sambale, and Arthur Sinulis, *Higher order concentration for functions of weakly dependent random variables*, Electron. J. Probab. **24** (2019), no. 85, 19 pp. MR-4003138

[14] David L. Hanson and Farroll T. Wright, *A bound on tail probabilities for quadratic forms in independent random variables*, Ann. Math. Statist. **42** (1971), no. 3, 1079–1083. MR-0279864

[15] Paweł Hitczenko, Stephen J. Montgomery-Smith, and Krzysztof Oleszkiewicz, *Moment inequalities for sums of certain independent symmetric random variables*, Studia Math. **123** (1997), no. 1, 15–42. MR-1438303

[16] Christian Houdré and Patricia Reynaud-Bouret, *Exponential inequalities, with constants, for U-statistics of order two*, Stochastic inequalities and applications, Birkhäuser, Springer, Progress in Probability **56** (2003), pp. 55–69.

[17] Daniel Hsu, Sham M. Kakade, and Tong Zhang, *A tail inequality for quadratic forms of subgaussian random vectors*, Electron. Commun. Probab. **17** (2012), no. 52, 6. MR-2994877

[18] Konrad Kolesko and Rafał Latała, *Moment estimates for chaoses generated by symmetric random variables with logarithmically convex tails*, Statist. Probab. Lett. **107** (2015), 210–214. MR-3412778

[19] Stanisław Kwapień, *Decoupling inequalities for polynomial chaos*, Ann. Probab. **15** (1987), no. 3, 1062–1071. MR-0893914

[20] Rafał Latała, *Tail and moment estimates for sums of independent random vectors with logarithmically concave tails*, Studia Math. **118** (1996), no. 3, 301–304. MR-1388035

[21] Rafał Latała, *Tail and moment estimates for some types of chaos*, Studia Math. **135** (1999), no. 1, 39–53. MR-1686370

[22] Rafał Latała, *Estimates of moments and tails of Gaussian chaoses*, Ann. Probab. **34** (2006), no. 6, 2315–2331. MR-2294983

[23] Rafał Latała and Rafał Łochowski, *Moment and tail estimates for multidimensional chaos generated by positive random variables with logarithmically concave tails*, Stochastic inequalities and applications, Progr. Probab., vol. 56, Birkhäuser, Basel, 2003, pp. 77–92. MR-2073428

[24] Michel Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR-1849347

[25] Rafał Meller, *Two-sided moment estimates for a class of nonnegative chaoses*, Statist. Probab. Lett. **119** (2016), 213–219. MR-3555289

[26] Rafał Meller, *Tail and moment estimates for a class of random chaoses of order two*, Studia Math. **249** (2019), no. 1, 1–32. MR-3984282

[27] Alexey Naumov, Vladimir Spokoiny, and Vladimir Ulyanov, *Bootstrap confidence sets for spectral projectors of sample covariance*, Probab. Theory Related Fields **174** (2019), no. 3-4, 1091–1132. MR-3980312

[28] Joscha Prochno, Christoph Thäle, and Nicola Turchi, *The isotropic constant of random polytopes with vertices on convex surfaces*, J. Complexity **54** (2019), 101394, 17 pp. MR-3983208

[29] Mark Rudelson and Roman Vershynin, *Hanson–Wright inequality and sub-Gaussian concentration*, Electron. Commun. Probab. **18** (2013), no. 82, 9. MR-3125258

[30] Holger Sambale, *Some notes on concentration for α-subexponential random variables*, arXiv preprint (2020).

[31] Holger Sambale and Arthur Sinulis, *Logarithmic Sobolev inequalities for finite spin systems and applications*, Bernoulli **26** (2020), no. 3, 1863–1890. MR-4091094

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

[33] Warren Schudy and Maxim Sviridenko, *Concentration and moment inequalities for polynomials of independent random variables*, Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, ACM, New York, 2012, pp. 437–446. MR-3205227

[34] Van H. Vu and Ke Wang, *Random weighted projections, random quadratic forms and random eigenvectors*, Random Structures Algorithms **47** (2015), no. 4, 792–821. MR-3418916

[35] Farroll T. Wright, *A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric*, Ann. Probability **1** (1973), no. 6, 1068–1070. MR-0353419

[36] Zhuoran Yang, Lin F. Yang, Ethan X. Fang, Tuo Zhao, Zhaoran Wang, and Matey Neykov, *Misspecified nonconvex statistical optimization for sparse phase retrieval*, Math. Program. **176** (2019), no. 1-2, Ser. B, 545–571. MR-3960819

Acknowledgments. We are very grateful to an anonymous referee for numerous valuable remarks which helped improving this paper significantly.