Some Notes on Concentration for α-Subexponential Random Variables

Holger Sambale

2020 Mathematics Subject Classification: 60E15; 60F10; 46E30; 46N30

1 Introduction

The aim of this note is to compile a number of smaller results that extend some classical as well as more recent concentration inequalities for bounded or sub-Gaussian random variables to random variables with heavier (but still exponential type) tails. In detail, we shall consider random variables X that satisfy

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq 2\exp(-t^\alpha / C_{1,\alpha}^\alpha) \tag{1.1}$$

for any $t \geq 0$, some $\alpha \in (0, 2]$, and a suitable constant $C_{1,\alpha} > 0$. Such random variables are sometimes called α-subexponential (for $\alpha = 2$, they are sub-Gaussian) or sub-Weibull(α) (cf. [23, Definition 2.2]).

There are several equivalent reformulations of (1.1), e. g., in terms of L^p norms:

$$\|X\|_{L^p} \leq C_{2,\alpha} p^{1/\alpha} \tag{1.2}$$

for any $p \geq 1$. Another characterization is that these random variables have finite Orlicz norms of order α, i. e.,

$$C_{3,\alpha} \coloneqq \|X\|_{\Psi_\alpha} \coloneqq \inf\{t > 0 \colon \mathbb{E}\exp((|X|/t)^\alpha) \leq 2\} < \infty. \tag{1.3}$$

If $\alpha < 1$, $\|\cdot\|_{\Psi_\alpha}$ is actually a quasi-norm; however, many norm-like properties (such as a triangle-type inequality) can nevertheless be recovered up to α-dependent

H. Sambale (✉)
Faculty of Mathematics, Bielefeld University, Bielefeld, Germany
e-mail: hsambale@math.uni-bielefeld.de

R. Adamczak et al. (eds.), *High Dimensional Probability IX*, Progress in Probability 80, https://doi.org/10.1007/978-3-031-26979-0_7

constants (see, e. g., [12, Appendix A]). In fact, $C_{1,\alpha}$, $C_{2,\alpha}$, and $C_{3,\alpha}$ can be chosen such that they only differ by a constant α-dependent factor.

Note that α-subexponential random variables have log-convex (if $\alpha \le 1$) or log-concave (if $\alpha \ge 1$) tails, i. e., $t \mapsto -\log \mathbb{P}(|X| \ge t)$ is convex or concave, respectively. For log-convex or log-concave measures, two-sided L^p norm estimates for polynomial chaos (and as a consequence, concentration bounds) have been established over the last 25 years. In the log-convex case, results of this type have been derived for linear forms in [17] and for forms of any order in [12, 21]. For log-concave measures, starting with linear forms again in [10], important contributions have been made in [3, 24, 25, 27].

In this note, we mainly present four different results for functions of α-subexponential random variables: a Hanson–Wright-type inequality in Sect. 2, a version of the convex concentration inequality in Sect. 3, a uniform Hanson–Wright inequality in Sect. 4, and finally a convex concentration inequality for simple random tensors in Sect. 5. These results are partly based on and generalize recent research, e. g., [20] and [42]. In fact, they partially build upon each other: for instance, in the proofs of Sect. 5, we apply results both from Sects. 2 and 3. A more detailed discussion is provided in each of the sections.

Finally, let us introduce some conventions that we will use in this chapter.

Notations. If $X_1, \ldots, X_n$ is a sequence of random variables, we denote by $X = (X_1, \ldots, X_n)$ the corresponding random vector. Moreover, we shall need the following types of norms throughout the paper:

- The norms $\|x\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $x \in \mathbb{R}^n$
- The L^p norms $\|X\|_{L^p} := (\mathbb{E}|X|^p)^{1/p}$ for random variables X (cf. (1.2))
- The Orlicz (quasi-)norms $\|X\|_{\Psi_\alpha}$ as introduced in (1.3)
- The Hilbert–Schmidt and operator norms $\|A\|_{\mathrm{HS}} := (\sum_{i,j} a_{ij}^2)^{1/2}$, $\|A\|_{\mathrm{op}} := \sup\{\|Ax\|_2 \colon \|x\|_2 = 1\}$ for matrices $A = (a_{ij})$

The constants appearing in this chapter (typically denoted C or c) may vary from line to line. Without subscript, they are assumed to be absolute, and if they depend on α (only), we shall write C_α or c_α.

2 A Generalized Hanson–Wright Inequality

Arguably, the most famous concentration result for quadratic form is the Hanson–Wright inequality, which first appeared in [16]. We may state it as follows: assuming $X_1, \ldots, X_n$ are centered, independent random variables satisfying $\|X_i\|_{\Psi_2} \le K$ for any i and $A = (a_{ij})$ is a symmetric matrix, we have for any $t \ge 0$

$$\mathbb{P}\big(|X^T AX - \mathbb{E}X^T AX| \ge t\big) \le 2\exp\Big(-\frac{1}{C}\min\Big(\frac{t^2}{K^4\|A\|_{\mathrm{HS}}^2}, \frac{t}{K^2\|A\|_{\mathrm{op}}}\Big)\Big).$$

For a modern proof, see [33], and for various developments, cf. [2, 4, 18, 43].

In this note, we provide an extension of the Hanson–Wright inequality to random variables with bounded Orlicz norms of any order $\alpha \in (0, 2]$. This complements the results in [12], where the case of $\alpha \in (0, 1]$ was considered, while for $\alpha = 2$, we get back the actual Hanson–Wright inequality.

Theorem 2.1 *For any $\alpha \in (0, 2]$, let $X_1, \ldots, X_n$ be independent, centered random variables such that $\|X_i\|_{\Psi_\alpha} \le K$ for any i and $A = (a_{ij})$ be a symmetric matrix. Then, for any $t \ge 0$,*

$$\mathbb{P}\big(|X^T AX - \mathbb{E}X^T AX| \ge t\big) \le 2\exp\Big(-\frac{1}{C_\alpha}\min\Big(\frac{t^2}{K^4\|A\|_{\mathrm{HS}}^2}, \Big(\frac{t}{K^2\|A\|_{\mathrm{op}}}\Big)^{\frac{\alpha}{2}}\Big)\Big).$$

Theorem 2.1 generalizes and implies a number of inequalities for quadratic forms in α-subexponential random variables (in particular for $\alpha = 1$) that are spread throughout the literature. For a detailed discussion, see [12, Remark 1.7]. Note that it is possible to sharpen the tail estimate given by Theorem 2.1, cf., e. g., [12, Corollary 1.4] for $\alpha \in (0, 1]$ or [3, Theorem 3.2] for $\alpha \in [1, 2]$ (in fact, the proof of Theorem 2.1 works by evaluating the family of norms used therein). The main benefit of Theorem 2.1 is that it uses norms that are easily calculable and in many situations already sufficient for applications.

Before we give the proof of Theorem 2.1, let us briefly mention that for the standard Hanson–Wright inequality, a number of selected applications can be found in [33]. Some of them were generalized to α-subexponential random variables with $\alpha \le 1$ in [12], and it is no problem to extend these proofs to any order $\alpha \in (0, 2]$ using Theorem 2.1. Here, we just focus on a single example that yields a concentration result for the Euclidean norm of a linear transformation of a vector X having independent components with bounded Orlicz norms around the Hilbert–Schmidt norm of the transformation matrix. This is a variant and extension of [12, Proposition 2.1] and will be applied in Sect. 5.

Proposition 2.2 *Let $X_1, \ldots, X_n$ be independent, centered random variables such that $\mathbb{E}X_i^2 = 1$ and $\|X_i\|_{\Psi_\alpha} \le K$ for some $\alpha \in (0, 2]$ and let $B \neq 0$ be an $m \times n$ matrix. For any $t \ge 0$, we have*

$$\mathbb{P}(|\|BX\|_2 - \|B\|_{\mathrm{HS}}| \ge tK^2\|B\|_{\mathrm{op}}) \le 2\exp(-t^\alpha/C_\alpha). \tag{2.1}$$

In particular, for any $t \ge 0$, it holds

$$\mathbb{P}(|\|X\|_2 - \sqrt{n}| \ge tK^2) \le 2\exp(-t^\alpha/C_\alpha). \tag{2.2}$$

For the proofs, let us recall some elementary relations that we will use throughout the paper to adjust the constants in the tail bounds we derive.

Adjusting constants. For any two constants $C_1 > C_2 > 1$, we have for all $r \ge 0$ and $C > 0$

$$C_1 \exp(-r/C) \le C_2 \exp\Big(-\frac{\log(C_2)}{C\log(C_1)}r\Big) \tag{2.3}$$

whenever the left-hand side is smaller or equal to 1 (cf., e. g., [35, Eq. (3.1)]). Moreover, for any $\alpha \in (0, 2)$, any $\gamma > 0$, and all $t \ge 0$, we may always estimate

$$\exp(-(t/C)^2) \le 2\exp(-(t/C')^\alpha), \tag{2.4}$$

using $\exp(-s^2) \le \exp(1 - s^\alpha)$ for any $s > 0$ and (2.3). More precisely, we may choose $C' := C/\log^{1/\alpha}(2)$. Note that strictly speaking, the range of $t/C \le 1$ is not covered by (2.3); however, in this case (in particular, choosing C' as suggested), both sides of (2.4) are at least 1 anyway so that the right-hand side still provides a valid upper bound for any probability.

Let us now turn to the proof of Theorem 2.1. In what follows, we actually show that for any $p \ge 2$,

$$\|X^T AX - \mathbb{E}X^T AX\|_{L^p} \le C_\alpha K^2\big(p^{1/2}\|A\|_{\mathrm{HS}} + p^{2/\alpha}\|A\|_{\mathrm{op}}\big). \tag{2.5}$$

From here, Theorem 2.1 follows by standard means (cf. [34, Proof of Theorem 3.6]). Moreover, we may restrict ourselves to $\alpha \in (1, 2]$, since the case of $\alpha \in (0, 1]$ has been proven in [12].

Proof of Theorem 2.1 First we shall treat the off-diagonal part of the quadratic form. Let $w_i^{(1)}, w_i^{(2)}$ be independent (of each other as well as of the X_i) symmetrized Weibull random variables with scale 1 and shape α, i. e., $w_i^{(j)}$ are symmetric random variables with $\mathbb{P}(|w_i^{(j)}| \ge t) = \exp(-t^\alpha)$. In particular, the $w_i^{(j)}$ have logarithmically concave tails.

Using standard decoupling and symmetrization arguments (cf. [8, Theorem 3.1.1 & Lemma 1.2.6]) as well as [3, Theorem 3.2] in the second inequality, for any $p \ge 2$, it holds

$$\begin{aligned}\|\sum_{i\neq j} a_{ij}X_iX_j\|_{L^p} &\le C_\alpha K^2\|\sum_{i\neq j} a_{ij}w_i^{(1)}w_j^{(2)}\|_{L^p}\\ &\le C_\alpha K^2(\|A\|^{\mathcal{N}}_{\{1,2\},p} + \|A\|^{\mathcal{N}}_{\{\{1\},\{2\}\},p}),\end{aligned} \tag{2.6}$$

where the norms $\|A\|^{\mathcal{N}}_{\mathcal{J},p}$ are defined as in [3]. Instead of repeating the general definitions, we will only focus on the case we need in our situation. Indeed, for the symmetric Weibull distribution with parameter α, we have (again, in the notation of [3]) $N(t) = t^\alpha$, and so for $\alpha \in (1, 2]$, it follows that $\hat{N}(t) = \min(t^2, |t|^\alpha)$. Hence, the norms can be written as follows:

$$\|A\|^{\mathcal{N}}_{\{1,2\},p} = 2\sup\Big\{\sum_{i,j} a_{ij}x_{ij} : \sum_{i=1}^n \min\Big(\sum_j x_{ij}^2, \Big(\sum_j x_{ij}^2\Big)^{\alpha/2}\Big) \le p\Big\},$$

$$\|A\|^{\mathcal{N}}_{\{\{1\},\{2\}\},p} = \sup\Big\{ \sum_{i,j} a_{ij}x_i y_j : \sum_{i=1}^n \min(x_i^2, |x_i|^\alpha) \le p, \sum_{j=1}^n \min(y_j^2, |y_j|^\alpha) \le p\Big\}.$$

Before continuing with the proof, we next introduce a lemma that will help to rewrite the norms in a more tractable form. □

Lemma 2.3 *For any $p \ge 2$, define*

$$I_1(p) := \Big\{x = (x_{ij}) \in \mathbb{R}^{n\times n} : \sum_{i=1}^n \min\Big(\Big(\sum_{j=1}^n x_{ij}^2\Big)^{\alpha/2}, \sum_{j=1}^n x_{ij}^2\Big) \le p\Big\},$$

$$I_2(p) := \Big\{x_{ij} = z_i y_{ij} \in \mathbb{R}^{n\times n} : \sum_{i=1}^n \min(|z_i|^\alpha, z_i^2) \le p, \max_{i=1,\dots,n} \sum_{j=1}^n y_{ij}^2 \le 1\Big\}.$$

Then $I_1(p) = I_2(p)$.

Proof The inclusion $I_1(p) \supseteq I_2(p)$ is an easy calculation, and the inclusion $I_1(p) \subseteq I_2(p)$ follows by defining $z_i = \|(x_{ij})_j\|$ and $y_{ij} = x_{ij}/\|(x_{ij})_j\|$ (or 0, if the norm is zero). □

Proof of Theorem 2.1, continued For brevity, for any matrix $A = (a_{ij})$, let us write $\|A\|_m := \max_{i=1,\dots,n}(\sum_{j=1}^n a_{ij}^2)^{1/2}$. Note that clearly, $\|A\|_m \le \|A\|_{\mathrm{op}}$.

Now, fix some vector $z \in \mathbb{R}^n$ such that $\sum_{i=1}^n \min(|z_i|^\alpha, z_i^2) \le p$. The condition also implies

$$p \ge \sum_{i=1}^n |z_i|^\alpha \mathbb{1}_{\{|z_i|>1\}} + \sum_{i=1}^n z_i^2 \mathbb{1}_{\{|z_i|\le 1\}} \ge \max\Big(\sum_{i=1}^n z_i^2 \mathbb{1}_{\{|z_i|\le 1\}}, \sum_{i=1}^n |z_i| \mathbb{1}_{\{|z_i|>1\}}\Big),$$

where in the second step we used $\alpha \in [1, 2]$ to estimate $|z_i|^\alpha \mathbb{1}_{\{|z_i|>1\}} \ge |z_i| \mathbb{1}_{\{|z_i|>1\}}$. So, given any z and y satisfying the conditions of $I_2(p)$, we can write

$$\begin{aligned}
\Big|\sum_{i,j} a_{ij} z_i y_{ij}\Big| &\le \sum_{i=1}^n |z_i| \Big(\sum_{j=1}^n a_{ij}^2\Big)^{1/2} \Big(\sum_{j=1}^n y_{ij}^2\Big)^{1/2} \le \sum_{i=1}^n |z_i| \Big(\sum_{j=1}^n a_{ij}^2\Big)^{1/2} \\
&\le \sum_{i=1}^n |z_i| \mathbb{1}_{\{|z_i|\le 1\}} \Big(\sum_{j=1}^n a_{ij}^2\Big)^{1/2} + \sum_{i=1}^n |z_i| \mathbb{1}_{\{|z_i|>1\}} \Big(\sum_{j=1}^n a_{ij}^2\Big)^{1/2} \\
&\le \|A\|_{\mathrm{HS}} \Big(\sum_{i=1}^n z_i^2 \mathbb{1}_{\{|z_i|\le 1\}}\Big)^{1/2} + \|A\|_m \sum_{i=1}^n |z_i| \mathbb{1}_{\{|z_i|>1\}}.
\end{aligned}$$

So, this yields

$$\|A\|^{\mathcal{N}}_{\{1,2\},p} \le 2p^{1/2}\|A\|_{\mathrm{HS}} + 2p\|A\|_m \le 2p^{1/2}\|A\|_{\mathrm{HS}} + 2p\|A\|_{\mathrm{op}}. \tag{2.7}$$

As for $\|A\|^{\mathcal{N}}_{\{\{1\},\{2\}\},p}$, we can use the decomposition $z = z_1 + z_2$, where $(z_1)_i = z_i \mathbb{1}_{\{|z_i|>1\}}$ and $z_2 = z - z_1$, and obtain

$$\begin{aligned}\|A\|^{\mathcal{N}}_{\{\{1\},\{2\}\},p} \le \sup\Big\{\sum_{ij} a_{ij}(x_1)_i(y_1)_j : \|x_1\|_\alpha \le p^{1/\alpha}, \|y_1\|_\alpha \le p^{1/\alpha}\Big\}\\ + 2\sup\Big\{\sum_{ij} a_{ij}(x_1)_i(y_2)_j : \|x_1\|_\alpha \le p^{1/\alpha}, \|y_2\|_2 \le p^{1/2}\Big\}\\ + \sup\Big\{\sum_{ij} a_{ij}(x_2)_i(y_2)_j : \|x_2\|_2 \le p^{1/2}, \|y_2\|_2 \le p^{1/2}\Big\}\\ = p^{2/\alpha}\sup\{\ldots\} + 2p^{1/\alpha+1/2}\sup\{\ldots\} + p\|A\|_{\mathrm{op}}\end{aligned}$$

(in the braces, the conditions $\|\cdot\|_\beta \le p^{1/\beta}$ have been replaced by $\|\cdot\|_\beta \le 1$). Clearly, since $\|x_1\|_\alpha \le 1$ implies $\|x_1\|_2 \le 1$ (and the same for y_1), all of the norms can be upper bounded by $\|A\|_{\mathrm{op}}$, i. e., we have

$$\|A\|^{\mathcal{N}}_{\{\{1\},\{2\}\},p} \le (p^{2/\alpha} + 2p^{1/\alpha+1/2} + p)\|A\|_{\mathrm{op}} \le 4p^{2/\alpha}\|A\|_{\mathrm{op}}, \tag{2.8}$$

where the last inequality follows from $p \ge 2$ and $1/2 \le 1/\alpha \le 1 \le (\alpha+2)/(2\alpha) \le 2/\alpha$.

Combining the estimates (2.6), (2.7), and (2.8) yields

$$\|\sum_{i,j} a_{ij}X_iX_j\|_{L^p} \le C_\alpha K^2\big(2p^{1/2}\|A\|_{\mathrm{HS}} + 6p^{2/\alpha}\|A\|_{\mathrm{op}}\big).$$

To treat the diagonal terms, we use Corollary 6.1 in [12], as X_i^2 are independent and satisfy $\|X_i^2\|_{\Psi_{\alpha/2}} \le K^2$, so that it yields

$$\mathbb{P}\big(|\sum_{i=1}^n a_{ii}(X_i^2 - \mathbb{E}X_i^2)| \ge t\big) \le 2\exp\Big(-\frac{1}{C_\alpha K^2}\min\Big(\frac{t^2}{\sum_{i=1}^n a_{ii}^2}, \Big(\frac{t}{\max_{i=1,\ldots,n}|a_{ii}|}\Big)^{\alpha/2}\Big)\Big).$$

Now it is clear that $\max_{i=1,\ldots,n}|a_{ii}| \le \|A\|_{\mathrm{op}}$ and $\sum_{i=1}^n a_{ii}^2 \le \|A\|_{\mathrm{HS}}^2$. In particular,

$$\|\sum_{i=1}^n a_{ii}(X_i^2 - \mathbb{E}X_i^2)\|_{L^p} \le C_\alpha K^2(p^{1/2}\|A\|_{\mathrm{HS}} + p^{2/\alpha}\|A\|_{\mathrm{op}}).$$

The claim (2.5) now follows from Minkowski's inequality. □

Finally, we prove Proposition 2.2.

Proof of Proposition 2.2 It suffices to prove (2.1) for matrices satisfying $\|B\|_{\mathrm{HS}} = 1$, as otherwise we set $\widetilde{B} = B\|B\|_{\mathrm{HS}}^{-1}$ and use the equality

$$\{|\|BX\|_2 - \|B\|_{\mathrm{HS}}| \geq \|B\|_{\mathrm{op}} t\} = \{|\|\widetilde{B}X\|_2 - 1| \geq \|\widetilde{B}\|_{\mathrm{op}} t\}.$$

Now let us apply Theorem 2.1 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 for any $t \geq 0$

$$\begin{aligned}\mathbb{P}\big(|\|BX\|_2 - 1| \geq t\big) &\leq \mathbb{P}\big(|\|BX\|_2^2 - 1| \geq \max(t, t^2)\big) \\ &\leq 2\exp\Big(-\frac{1}{C_\alpha}\min\Big(\frac{\max(t,t^2)^2}{K^4\|B\|_{\mathrm{op}}^2}, \Big(\frac{\max(t,t^2)}{K^4\|B\|_{\mathrm{op}}^2}\Big)^{\alpha/2}\Big)\Big) \\ &\leq 2\exp\Big(-\frac{1}{C_\alpha}\min\Big(\frac{t^2}{K^4\|B\|_{\mathrm{op}}^2}, \Big(\frac{t^2}{K^4\|B\|_{\mathrm{op}}^2}\Big)^{\alpha/2}\Big)\Big) \\ &\leq 2\exp\Big(-\frac{1}{C_\alpha}\Big(\frac{t}{K^2\|B\|_{\mathrm{op}}}\Big)^{\alpha}\Big).\end{aligned}$$

Here, the first step follows from $|z-1| \leq \min(|z^2-1|, |z^2-1|^{1/2})$ for $z \geq 0$, in the second step, we have used the estimates $\|A\|_{\mathrm{HS}}^2 \leq \|B\|_{\mathrm{op}}^2\|B\|_{\mathrm{HS}}^2 = \|B\|_{\mathrm{op}}^2$ and $\|A\|_{\mathrm{op}} \leq \|B\|_{\mathrm{op}}^2$, and moreover, the fact that since $\mathbb{E}X_i^2 = 1$, $K \geq C_\alpha > 0$ (cf., e. g., [12, Lemma A.2]), while the last step follows from (2.4) and (2.3). Setting $t = K^2 s\|B\|_{\mathrm{op}}$ for $s \geq 0$ finishes the proof of (2.1). Finally, (2.2) follows by taking $m = n$ and $B = I$. □

3 Convex Concentration for Random Variables with Bounded Orlicz Norms

Assume $X_1, \ldots, X_n$ are independent random variables each taking values in some bounded interval $[a, b]$. Then, by convex concentration as established in [19, 29, 38], for every convex 1-Lipschitz function $f: [a, b]^n \to \mathbb{R}$,

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \leq 2\exp\Big(-\frac{t^2}{2(b-a)^2}\Big) \tag{3.1}$$

for any $t \geq 0$ (see, e. g., [36, Corollary 3]).

While convex concentration for bounded random variables is by now standard, there is less literature for unbounded random variables. In [31], a Martingale-type approach is used, leading to a result for functionals with stochastically bounded increments. The special case of suprema of unbounded empirical processes was treated in [1, 28, 40]. Another branch of research, begun in [29] and continued, e. g., in [5, 13–15, 36, 37], is based on functional inequalities (such as Poincaré or log-Sobolev inequalities) restricted to convex functions and weak transport-entropy inequalities. In [20, Lemma 1.8], a generalization of (3.1) for sub-Gaussian random variables ($\alpha = 2$) was proven, which we may extend to any order $\alpha \in (0, 2]$.

Proposition 3.1 *Let $X_1, \ldots, X_n$ be independent random variables, $\alpha \in (0, 2]$ and $f : \mathbb{R}^n \to \mathbb{R}$ convex and 1-Lipschitz. Then, for any $t \geq 0$,*

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \leq 2\exp\Big(-\frac{t^\alpha}{C_\alpha \|\max_i |X_i|\|_{\Psi_\alpha}^\alpha}\Big).$$

In particular,

$$\|f(X) - \mathbb{E}f(X)\|_{\Psi_\alpha} \leq C_\alpha \|\max_i |X_i|\|_{\Psi_\alpha}. \tag{3.2}$$

Note that the main results of the following two sections can be regarded as applications of Proposition 3.1. If f is separately convex only (i. e., convex is every coordinate with the other coordinates being fixed), it is still possible to prove a corresponding result for the upper tails. Indeed, it is no problem to modify the proof below accordingly, replacing (3.1) by [7, Theorem 6.10]. Moreover, note that $\|\max_i |X_i|\|_{\Psi_\alpha}$ cannot be replaced by $\max_i \||X_i|\|_{\Psi_\alpha}$ (a counterexample for $\alpha = 2$ is provided in [20]). In general, the Orlicz norm of $\max_i |X_i|$ will be of order $(\log n)^{1/\alpha}$ (cf. Lemma 5.6).

Proof of Proposition 3.1 Following the lines of the proof of [20, Lemma 3.5], the key step is a suitable truncation that goes back to [1]. Indeed, write

$$X_i = X_i 1_{\{|X_i| \leq M\}} + X_i 1_{\{|X_i| > M\}} =: Y_i + Z_i \tag{3.3}$$

with $M := 8\mathbb{E}\max_i |X_i|$ (in particular, $M \leq C_\alpha \|\max_i |X_i|\|_{\Psi_\alpha}$, cf. [12, Lemma A.2]), and let $Y = (Y_1, \ldots, Y_n)$, $Z = (Z_1, \ldots, Z_n)$. By the Lipschitz property of f,

$$\begin{aligned}
&\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t)\\
&\leq \mathbb{P}(|f(Y) - \mathbb{E}f(Y)| + |f(X) - f(Y)| + |\mathbb{E}f(Y) - \mathbb{E}f(X)| > t) \qquad (3.4)\\
&\leq \mathbb{P}(|f(Y) - \mathbb{E}f(Y)| + \|Z\|_2 + \mathbb{E}\|Z\|_2 > t),
\end{aligned}$$

and hence, it suffices to bound the terms in the last line.

Applying (3.1) to Y and using (2.4) and (2.3), we obtain

$$\mathbb{P}(|f(Y) - \mathbb{E}f(Y)| > t) \leq 2\exp\Big(-\frac{t^\alpha}{C_\alpha^\alpha \|\max_i |X_i|\|_{\Psi_\alpha}^\alpha}\Big). \tag{3.5}$$

Furthermore, below we will show that

$$\|\|Z\|_2\|_{\Psi_\alpha} \leq C_\alpha \|\max_i |X_i|\|_{\Psi_\alpha}. \tag{3.6}$$

Hence, for any $t \geq 0$,

$$\mathbb{P}(\|Z\|_2 \geq t) \leq 2\exp\Big(-\frac{t^\alpha}{C_\alpha^\alpha \|\max_i |X_i|\|_{\Psi_\alpha}^\alpha}\Big), \tag{3.7}$$

and by [12, Lemma A.2],

$$\mathbb{E}\|Z\|_2 \leq C_\alpha \|\max_i |X_i|\|_{\Psi_\alpha}. \tag{3.8}$$

Temporarily writing $K := C_\alpha \|\max_i |X_i|\|_{\Psi_\alpha}$, where C_α is large enough so that (3.5), (3.7), and (3.8) hold, (3.4) and (3.8) yield

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \leq \mathbb{P}(|f(Y) - \mathbb{E}f(Y)| + \|Z\|_2 > t - K)$$

if $t \geq K$. Using subadditivity and invoking (3.5) and (3.7), we obtain

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \leq 4\exp\Big(-\frac{(t-K)^\alpha}{(2K)^\alpha}\Big) \leq 4\exp\Big(-\frac{t^\alpha}{c_\alpha (2K)^\alpha}\Big),$$

where the last step holds for $t \geq K + \delta$ for some $\delta > 0$. This bound extends trivially to any $t \geq 0$ (if necessary, by a suitable change of constants). Finally, the constant in front of the exponential may be adjusted to 2 by (2.3), which finishes the proof.

It remains to show (3.6). To this end, recall the Hoffmann-Jørgensen inequality (cf. [30, Theorem 6.8]) in the following form: if $W_1, \ldots, W_n$ are independent random variables, $S_k := W_1 + \ldots + W_k$, and $t \geq 0$ is such that $\mathbb{P}(\max_k |S_k| > t) \leq 1/8$, then

$$\mathbb{E}\max_k |S_k| \leq 3\mathbb{E}\max_i |W_i| + 8t.$$

In our case, we set $W_i := Z_i^2$, $t = 0$, and note that by Chebyshev's inequality,

$$\mathbb{P}(\max_i Z_i^2 > 0) = \mathbb{P}(\max_i |X_i| > M) \leq \mathbb{E}\max_i |X_i|/M = 1/8,$$

and consequently, recalling that $S_k = Z_1^2 + \ldots + Z_k^2$,

$$\mathbb{P}(\max_k |S_k| > 0) \leq \mathbb{P}(\max_i Z_i^2 > 0) \leq 1/8.$$

Thus, together with [12, Lemma A.2], we obtain

$$\mathbb{E}\|Z\|_2^2 \le 3\mathbb{E}\max_i Z_i^2 \le C_\alpha \|\max_i Z_i^2\|_{\Psi_{\alpha/2}}.$$

Now it is easy to see that $\|\max_i Z_i^2\|_{\Psi_{\alpha/2}} \le \|\max_i |X_i|\|_{\Psi_\alpha}^2$, so that altogether we arrive at

$$\mathbb{E}\|Z\|_2^2 \le C_\alpha \|\max_i |X_i|\|_{\Psi_\alpha}^2. \tag{3.9}$$

Furthermore, by [30, Theorem 6.21], if $W_1, \dots, W_n$ are independent random variables with zero mean and $\alpha \in (0, 1]$,

$$\|\sum_{i=1}^n W_i\|_{\Psi_\alpha} \le C_\alpha (\|\sum_{i=1}^n W_i\|_{L^1} + \|\max_i |W_i|\|_{\Psi_\alpha}).$$

In our case, we consider $W_i = Z_i^2 - \mathbb{E}Z_i^2$ and $\alpha/2$ (instead of α). Together with the previous arguments (in particular, (3.9)) and [12, Lemma A.3], this yields

$$\begin{aligned}\|\sum_{i=1}^n (Z_i^2 - \mathbb{E}Z_i^2)\|_{\Psi_{\alpha/2}} &\le C_\alpha (\mathbb{E}|\|Z\|_2^2 - \mathbb{E}\|Z\|_2^2| + \|\max_i |Z_i^2 - \mathbb{E}Z_i^2|\|_{\Psi_{\alpha/2}}) \\ &\le C_\alpha (\mathbb{E}\|Z\|_2^2 + \|\max_i Z_i^2\|_{\Psi_{\alpha/2}}) \le C_\alpha \|\max_i |X_i|\|_{\Psi_\alpha}^2.\end{aligned}$$

Combining this with [12, Lemma A.3] and (3.9), we arrive at (3.6). □

4 Uniform Tail Bounds for First- and Second-Order Chaos

In this section, we discuss bounds for the tails of the supremum of certain chaos-type classes of functions. Even if we are particularly interested in quadratic forms, i. e., uniform Hanson–Wright inequalities, let us first consider linear forms.

Let $X_1, \dots, X_n$ be independent random variables, let $\alpha \in (0, 2]$, and let $\{a_{i,t} : i = 1, \dots, n, t \in \mathcal{T}\}$ be a compact set of real numbers, where $\mathcal{T}$ is some index set. Consider $g(X) := \sup_{t\in\mathcal{T}} \sum_{i=1}^n a_{i,t} X_i$. Clearly, g is convex and has Lipschitz constant $D := \sup_{t\in\mathcal{T}} (\sum_{i=1}^n a_{i,t}^2)^{1/2}$. Therefore, applying Proposition 3.1, we immediately obtain that for any $t \ge 0$,

$$\mathbb{P}(|g(X) - \mathbb{E}g(X)| \ge t) \le 2\exp\Big(-\frac{t^\alpha}{C_\alpha D^\alpha \|\max_i |X_i|\|_{\Psi_\alpha}^\alpha}\Big). \tag{4.1}$$

For bounded random variables, corresponding tail bounds can be found, e. g., in [32, Eq. (14)], and choosing $\alpha = 2$, we get back this result up to constants.

Our main aim is to derive a second-order analogue of (4.1), i. e., a uniform Hanson–Wright inequality. A pioneering result in this direction (for Rademacher variables) can be found in [39]. Later results include [2] (which requires the so-called concentration property), [22], [9], and [11] (certain classes of weakly dependent random variables). In [20], a uniform Hanson–Wright inequality for sub-Gaussian random variables was proven. We may show a similar result for random variables with bounded Orlicz norms of any order $\alpha \in (0, 2]$.

Theorem 4.1 *Let $X_1, \dots, X_n$ be independent, centered random variables and $K := \|\max_i |X_i|\|_{\Psi_\alpha}$, where $\alpha \in (0, 2]$. Let $\mathcal{A}$ be a compact set of real symmetric $n \times n$ matrices, and let $f(X) := \sup_{A \in \mathcal{A}}(X^T AX - \mathbb{E}X^T AX)$. Then, for any $t \geq 0$,*

$$\mathbb{P}(f(X) - \mathbb{E}f(X) \geq t) \leq 2\exp\Big(- \frac{1}{C_\alpha K^\alpha} \min \Big(\frac{t^\alpha}{(\mathbb{E}\sup_{A \in \mathcal{A}} \|AX\|_2)^\alpha}, \frac{t^{\alpha/2}}{\sup_{A \in \mathcal{A}} \|A\|_{\mathrm{op}}^{\alpha/2}} \Big)\Big).$$

For $\alpha = 2$, this gives back [20, Theorem 1.1] (up to constants and a different range of t). Comparing Theorems 4.1 to 2.1, we note that instead of a sub-Gaussian term, we obtain an α-subexponential term (which can be trivially transformed into a sub-Gaussian term for $t \leq \mathbb{E}\sup_{A \in \mathcal{A}} \|AX\|_2$, but this does not cover the complete α-subexponential regime). Moreover, Theorem 4.1 only gives a bound for the upper tails. Therefore, if $\mathcal{A}$ just consists of a single matrix, Theorem 2.1 is stronger. These differences have technical reasons.

To prove Theorem 4.1, we shall follow the basic steps of [20] and modify those where the truncation comes in. Let us first repeat some tools and results. In the sequel, for a random vector $W = (W_1, \dots, W_n)$, we shall denote

$$f(W) := \sup_{A \in \mathcal{A}} (W^T AW - g(A)), \tag{4.2}$$

where $g \colon \mathbb{R}^{n \times n} \to \mathbb{R}$ is some function. Moreover, if A is any matrix, we denote by $\mathrm{Diag}(A)$ its diagonal part (regarded as a matrix with zero entries on its off-diagonal). The following lemma combines [20, Lemmas 3.2 & 3.5].

Lemma 4.2

(1) *Assume the vector W has independent components that satisfy $W_i \leq K$ a.s. Then, for any $t \geq 1$, we have*

$$f(W) - \mathbb{E}f(W) \leq C\Big(K(\mathbb{E}\sup_{A \in \mathcal{A}} \|AW\|_2 + \mathbb{E}\sup_{A \in \mathcal{A}} \|\mathrm{Diag}(A)W\|_2)\sqrt{t} + K^2 \sup_{A \in \mathcal{A}} \|A\|_{\mathrm{op}} t\Big)$$

with probability at least $1 - e^{-t}$.

(2) *Assuming the vector W has independent (but not necessarily bounded) components with mean zero, we have*

$$\mathbb{E} \sup_{A \in \mathcal{A}} \|\mathrm{Diag}(A)W\|_2 \le C\mathbb{E} \sup_{A \in \mathcal{A}} \|AW\|_2.$$

From now on, let X be the random vector from Theorem 4.1, and recall the truncated random vector Y that we introduced in (3.3) (and the corresponding "remainder" Z). Then, Lemma 4.2 (1) for $f(Y)$ with $g(A) = \mathbb{E}X^T AX$ yields

$$f(Y) - \mathbb{E}f(Y) \le C\big(M(\mathbb{E} \sup_{A \in \mathcal{A}} \|AY\|_2 + \mathbb{E} \sup_{A \in \mathcal{A}} \|\mathrm{Diag}(A)\|_2)t^{1/\alpha} + M^2 t^{2/\alpha} \sup_{A \in \mathcal{A}} \|A\|_{\mathrm{op}}\big) \tag{4.3}$$

with probability at least $1 - e^{-t}$ (actually, (4.3) even holds with $\alpha = 2$, but in the sequel we will have to use the weaker version given above anyway). Here we recall that $M \le C_\alpha \|\max_i |X_i|\|_{\Psi_\alpha}$.

To prove Theorem 4.1, it remains to replace the terms involving the truncated random vector Y by the original vector X. First, by Proposition 3.1 and since $\sup_{A \in \mathcal{A}} \|AX\|_2$ is $\sup_{A \in \mathcal{A}} \|A\|_{\mathrm{op}}$-Lipschitz, we obtain

$$\mathbb{P}(\sup_{A \in \mathcal{A}} \|AX\|_2 > \mathbb{E} \sup_{A \in \mathcal{A}} \|AX\|_2 + C_\alpha \|\max_i |X_i|\|_{\Psi_\alpha} \sup_{A \in \mathcal{A}} \|A\|_{\mathrm{op}} t^{1/\alpha}) \le 2e^{-t}. \tag{4.4}$$

Moreover, by (3.8),

$$|\mathbb{E} \sup_{A \in \mathcal{A}} \|AY\|_2 - \mathbb{E} \sup_{A \in \mathcal{A}} \|AX\|_2| \le C_\alpha \|\max_i |X_i|\|_{\Psi_\alpha} \sup_{A \in \mathcal{A}} \|A\|_{\mathrm{op}}. \tag{4.5}$$

Next we estimate the difference between the expectations of $f(X)$ and $f(Y)$.

Lemma 4.3 *We have*

$$|\mathbb{E}f(Y) - \mathbb{E}f(X)| \le C_\alpha\big(\|\max_i |X_i|\|_{\Psi_\alpha} \mathbb{E} \sup_{A \in \mathcal{A}} \|AX\|_2 + \|\max_i |X_i|\|_{\Psi_\alpha}^2 \sup_{A \in \mathcal{A}} \|A\|_{\mathrm{op}}\big).$$

Proof First note that

$$\begin{aligned} f(X) &= \sup_{A \in \mathcal{A}} (Y^T AY - \mathbb{E}X^T AX + Z^T AX + Z^T AY) \\ &\le \sup_{A \in \mathcal{A}} (Y^T AY - \mathbb{E}X^T AX) + \sup_{A \in \mathcal{A}} |Z^T AX| + \sup_{A \in \mathcal{A}} |Z^T AY| \\ &\le f(Y) + \|Z\|_2 \sup_{A \in \mathcal{A}} \|AX\|_2 + \|Z\|_2 \sup_{A \in \mathcal{A}} \|AY\|_2. \end{aligned}$$

The same holds if we reverse the roles of X and Y. As a consequence,

$$|f(X) - f(Y)| \le \|Z\|_2 \sup_{A\in\mathcal{A}} \|AX\|_2 + \|Z\|_2 \sup_{A\in\mathcal{A}} \|AY\|_2, \tag{4.6}$$

and thus, taking expectations and applying Hölder's inequality,

$$|\mathbb{E}f(X) - \mathbb{E}f(Y)| \le (\mathbb{E}\|Z\|_2^2)^{1/2}((\mathbb{E}\sup_{A\in\mathcal{A}} \|AX\|_2^2)^{1/2} + (\mathbb{E}\sup_{A\in\mathcal{A}} \|AY\|_2^2)^{1/2}). \tag{4.7}$$

We may estimate $(\mathbb{E}\|Z\|_2^2)^{1/2}$ using (3.9). Moreover, by related arguments as in (3.8), from (4.4), we get that

$$\mathbb{E}\sup_{A\in\mathcal{A}} \|AX\|_2^2 \le C_\alpha((\mathbb{E}\sup_{A\in\mathcal{A}} \|AX\|_2)^2 + \|\max_i |X_i|\|_{\Psi_\alpha}^2 \sup_{A\in\mathcal{A}} \|A\|_{\mathrm{op}}^2).$$

Arguing similarly and using (4.5), the same bound also holds for $(\mathbb{E}\sup_{A\in\mathcal{A}} \|AY\|_2^2)^{1/2}$. Taking roots and plugging everything into (4.7) complete the proof. □

Finally, we prove the central result of this section.

Proof of Theorem 4.1 First, it immediately follows from Lemma 4.3 that

$$\mathbb{E}f(Y) \le \mathbb{E}f(X) + C_\alpha\big(\|\max_i |X_i|\|_{\Psi_\alpha} \mathbb{E}\sup_{A\in\mathcal{A}} \|AX\|_2 + \|\max_i |X_i|\|_{\Psi_\alpha}^2 \sup_{A\in\mathcal{A}} \|A\|_{\mathrm{op}}\big). \tag{4.8}$$

Moreover, by (4.5) and Lemma 4.2 (2),

$$\begin{aligned}\mathbb{E}\sup_{A\in\mathcal{A}} \|AY\|_2 + \mathbb{E}\sup_{A\in\mathcal{A}} \|\mathrm{Diag}(A)Y\|_2 \le C_\alpha(&\mathbb{E}\sup_{A\in\mathcal{A}} \|AX\|_2 \\ &+ \|\max_i |X_i|\|_{\Psi_\alpha} \sup_{A\in\mathcal{A}} \|A\|_{\mathrm{op}}).\end{aligned} \tag{4.9}$$

Finally, it follows from (4.6), (4.4), and (4.5) that

$$\begin{aligned}|f(X) - f(Y)| &\le \|Z\|_2 \sup_{A\in\mathcal{A}} \|AX\|_2 + \|Z\|_2 \sup_{A\in\mathcal{A}} \|AY\|_2 \\ &\le C_\alpha(\|Z\|_2 \mathbb{E}\sup_{A\in\mathcal{A}} \|AX\|_2 + \|Z\|_2 \|\max_i |X_i|\|_{\Psi_\alpha} \sup_{A\in\mathcal{A}} \|A\|_{\mathrm{op}} t^{1/\alpha})\end{aligned}$$

with probability at least $1 - 4e^{-t}$ for all $t \ge 1$. Using (3.7), it follows that

$$\begin{aligned}|f(X) - f(Y)| \le C_\alpha(&\|\max_i |X_i|\|_{\Psi_\alpha} \mathbb{E}\sup_{A\in\mathcal{A}} \|AX\|_2 t^{1/\alpha} \\ &+ \|\max_i |X_i|\|_{\Psi_\alpha}^2 \sup_{A\in\mathcal{A}} \|A\|_{\mathrm{op}} t^{2/\alpha})\end{aligned} \tag{4.10}$$

with probability at least $1 - 6e^{-t}$ for all $t \ge 1$. Combining (4.8), (4.9), and (4.10) and plugging into (4.3) thus yield that with probability at least $1 - 6e^{-t}$ for all $t \ge 1$,

$$\begin{aligned} f(X) - \mathbb{E}f(X) &\le C_\alpha(\|\max_i |X_i|\|_{\Psi_\alpha} \mathbb{E} \sup_{A \in \mathcal{A}} \|AX\|_2 t^{1/\alpha} \\ &\quad + \|\max_i |X_i|\|_{\Psi_\alpha}^2 \sup_{A \in \mathcal{A}} \|A\|_{\mathrm{op}} t^{2/\alpha}) \\ &=: C_\alpha(at^{1/\alpha} + bt^{2/\alpha}). \end{aligned}$$

If $u \ge \max(a, b)$, it follows that

$$\mathbb{P}(f(X) - \mathbb{E}f(X) \ge u) \le 6 \exp\Big(- \frac{1}{C_\alpha} \min\Big(\Big(\frac{u}{a}\Big)^\alpha, \Big(\frac{u}{b}\Big)^{\alpha/2}\Big)\Big).$$

By standard means (a suitable change of constants, using (2.3)), this bound may be extended to any $u \ge 0$, and the constant may be adjusted to 2. □

5 Random Tensors

By a *simple random tensor*, we mean a random tensor of the form

$$X := X_1 \otimes \cdots \otimes X_d = (X_{1,i_1} \cdots X_{d,i_d})_{i_1,\dots,i_d} \in \mathbb{R}^{n^d}, \tag{5.1}$$

where all X_k are independent random vectors in $\mathbb{R}^n$ whose coordinates are independent, centered random variables with variance one. Concentration results for random tensors (typically for polynomial-type functions) have been shown in [6, 12, 26], for instance.

Recently, in [42], new and interesting concentration bounds for simple random tensors were shown. In comparison to previous work, these inequalities focus on *small* values of t, e. g., a regime where sub-Gaussian tail decay holds. Moreover, in contrast to previous papers, [42] provides constants with optimal dependence on d. One of these results is the following convex concentration inequality: assuming that n and d are positive integers, $f : \mathbb{R}^{n^d} \to \mathbb{R}$ is convex and 1-Lipschitz, and the X_{ij} are bounded a.s., then for any $t \in [0, 2n^{d/2}]$,

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \le 2 \exp\Big(- \frac{t^2}{Cdn^{d-1}}\Big), \tag{5.2}$$

where $C > 0$ only depends on the bound of the coordinates. Using Theorem 2.1 and Proposition 3.1, we may extend this result to unbounded random variables as follows:

Theorem 5.1 *Let $n, d \in \mathbb{N}$ and $f : \mathbb{R}^{n^d} \to \mathbb{R}$ be convex and 1-Lipschitz. Consider a simple random tensor $X := X_1 \otimes \cdots \otimes X_d$ as in (5.1). Fix $\alpha \in [1, 2]$, and assume that $\|X_{i,j}\|_{\Psi_\alpha} \le K$. Then, for any $t \in [0, c_\alpha n^{d/2} (\log n)^{1/\alpha}/K]$,*

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \le 2\exp\Big(-\frac{1}{C_\alpha}\Big(\frac{t}{d^{1/2}n^{(d-1)/2}(\log n)^{1/\alpha}K}\Big)^\alpha\Big).$$

On the other hand, if $\alpha \in (0,1)$*, then, for any* $t \in [0, c_\alpha n^{d/2}(\log n)^{1/\alpha} d^{1/\alpha - 1/2}/K]$,

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \le 2\exp\Big(-\frac{1}{C_\alpha}\Big(\frac{t}{d^{1/\alpha}n^{(d-1)/2}(\log n)^{1/\alpha}K}\Big)^\alpha\Big).$$

The logarithmic factor stems from the Orlicz norm of $\max_i |X_i|$ in Proposition 3.1. For a slightly sharper version that includes the explicit dependence on these norms (and also gives back (5.2) for bounded random variables and $\alpha = 2$), see (5.12) in the proof of Theorem 5.1. We believe that Theorem 5.1 is non-optimal for $\alpha < 1$ as we would expect a bound of the same type as for $\alpha \in [1,2]$. However, a key difference in the proofs is that in the case of $\alpha \ge 1$ we can make use of moment-generating functions. This is clearly not possible if $\alpha < 1$, so that less subtle estimates must be invoked instead.

For the proof of Theorem 5.1, we first adapt some preliminary steps and compile a number of auxiliary lemmas whose proofs are deferred to the appendix. As a start, we need some additional characterizations of α-subexponential random variables via the behavior of the moment-generating functions:

Proposition 5.2 *Let* X *be a random variable and* $\alpha \in (0,2]$*. Then, the properties* (1.1)*,* (1.2)*, and* (1.3) *are equivalent to*

$$\mathbb{E}\exp(\lambda^\alpha |X|^\alpha) \le \exp(C_{4,\alpha}^\alpha \lambda^\alpha) \tag{5.3}$$

for all $0 \le \lambda \le 1/C_{4,\alpha}$*. If* $\alpha \in [1,2]$ *and* $\mathbb{E}X = 0$*, then the above properties are moreover equivalent to*

$$\mathbb{E}\exp(\lambda X) \le \begin{cases} \exp(C_{5,\alpha}^2\lambda^2) & \text{if } |\lambda| \le 1/C_{5,\alpha} \\ \exp(C_{5,\alpha}^{\alpha/(\alpha-1)}|\lambda|^{\alpha/(\alpha-1)}) & \text{if } |\lambda| \ge 1/C_{5,\alpha} \text{ and } \alpha > 1. \end{cases} \tag{5.4}$$

The parameters $C_{i,\alpha}$*,* $i = 1,\dots,5$*, can be chosen such that they only differ by constant* α*-dependent factors. In particular, we can take* $C_{i,\alpha} = c_{i,\alpha}\|X\|_{\Psi_\alpha}$.

To continue, note that $\|X\|_2 = \prod_{i=1}^d \|X_i\|_2$. A key step in the proofs of [42] is a maximal inequality that simultaneously controls the tails of $\prod_{i=1}^k \|X_i\|_2$, $k = 1,\dots,d$, where the X_i have independent sub-Gaussian components, i. e., $\alpha = 2$. Generalizing these results to any order $\alpha \in (0,2]$ is not hard. The following preparatory lemma extends [42, Lemma 3.1]. Note that in the proof (given in the appendix again), we apply Proposition 2.2.

Lemma 5.3 *Let* $X_1,\dots,X_d \in \mathbb{R}^n$ *be independent random vectors with independent, centered coordinates such that* $\mathbb{E}X_{i,j}^2 = 1$ *and* $\|X_{i,j}\|_{\Psi_\alpha} \le K$ *for some* $\alpha \in (0,2]$*. Then, for any* $t \in [0, 2n^{d/2}]$,

$$\mathbb{P}\Big(\prod_{i=1}^{d}\|X_i\|_2 > n^{d/2}+t\Big) \le 2\exp\Big(-\frac{1}{C_\alpha}\Big(\frac{t}{K^2 d^{1/2} n^{(d-1)/2}}\Big)^{\alpha}\Big).$$

To control all $k = 1, \dots, d$ simultaneously, we need a generalized version of the maximal inequality [42, Lemma 3.2] that we state next.

Lemma 5.4 *Let $X_1, \dots, X_d \in \mathbb{R}^n$ be independent random vectors with independent, centered coordinates such that $\mathbb{E}X_{i,j}^2 = 1$ and $\|X_{i,j}\|_{\Psi_\alpha} \le K$ for some $\alpha \in (0, 2]$. Then, for any $u \in [0, 2]$,*

$$\mathbb{P}\Big(\max_{1\le k\le d} n^{-k/2}\prod_{i=1}^{k}\|X_i\|_2 > 1+u\Big) \le 2\exp\Big(-\frac{1}{C_\alpha}\Big(\frac{n^{1/2}u}{K^2 d^{1/2}}\Big)^{\alpha}\Big).$$

The following Martingale-type bound is directly taken from [42]:

Lemma 5.5 ([42], Lemma 4.1) *Let $X_1, \dots X_d$ be independent random vectors. For each $k = 1, \dots, d$, let $f_k = f_k(X_k, \dots, X_d)$ be an integrable real-valued function and $\mathcal{E}_k$ be an event that is uniquely determined by the vectors $X_k, \dots, X_d$. Let $\mathcal{E}_{d+1}$ be the entire probability space. Suppose that for every $k = 1, \dots, d$, we have*

$$\mathbb{E}_{X_k}\exp(f_k) \le \pi_k$$

for every realization of $X_{k+1}, \dots, X_d$ in $\mathcal{E}_{k+1}$. Then, for $\mathcal{E} := \mathcal{E}_2 \cap \dots \cap \mathcal{E}_d$, we have

$$\mathbb{E}\exp(f_1 + \ldots + f_d)1_{\mathcal{E}} \le \pi_1 \cdots \pi_d.$$

Finally, we need a bound for the Orlicz norm of $\max_i |X_i|$.

Lemma 5.6 *Let $X_1, \dots, X_n$ be independent, centered random variables such that $\|X_i\|_{\Psi_\alpha} \le K$ for any i and some $\alpha > 0$. Then,*

$$\|\max_i |X_i|\|_{\Psi_\alpha} \le C_\alpha K \max\Big\{\Big(\frac{\sqrt{2}+1}{\sqrt{2}-1}\Big)^{1/\alpha}, (\log n)^{1/\alpha}\Big(\frac{2}{\log 2}\Big)^{1/\alpha}\Big\}.$$

Here, we may choose $C_\alpha = \max\{2^{1/\alpha-1}, 2^{1-1/\alpha}\}$.

Note that for $\alpha \ge 1$, [8, Proposition 4.3.1] provides a similar result. However, we are also interested in the case of $\alpha < 1$ in the present note. The condition $\mathbb{E}X_i = 0$ in Lemma 5.6 can easily be removed only at the expense of a different absolute constant.

We are now ready to prove Theorem 5.1.

Proof of Theorem 5.1 We shall adapt the arguments from [42]. First let

$$\mathcal{E}_k := \Big\{ \prod_{i=k}^{d} \|X_i\|_2 \le 2n^{(d-k+1)/2} \Big\}, \qquad k = 1, \ldots, d,$$

and let $\mathcal{E}_{d+1}$ be the full space. It then follows from Lemma 5.4 for $u = 1$ that

$$\mathbb{P}(\mathcal{E}) \ge 1 - 2\exp\Big(-\frac{1}{C_\alpha}\Big(\frac{n^{1/2}}{K^2 d^{1/2}}\Big)^{\alpha}\Big), \tag{5.5}$$

where $\mathcal{E} := \mathcal{E}_2 \cap \cdots \cap \mathcal{E}_d$.

Now fix any realization $x_2, \ldots, x_d$ of the random vectors $X_2, \ldots, X_d$ in $\mathcal{E}_2$, and apply Proposition 3.1 to the function $f_1(x_1)$ given by $x_1 \mapsto f(x_1, \ldots x_d)$. Clearly, f_1 is convex, and since

$$|f(x\otimes x_2\otimes\cdots\otimes x_d) - f(y\otimes x_2\otimes\cdots\otimes x_d)| \le \|x-y\|_2 \prod_{i=2}^{d} \|x_i\|_2 \le \|x-y\|_2 2n^{(d-1)/2},$$

we see that it is $2n^{(d-1)/2}$-Lipschitz. Hence, it follows from (3.2) that

$$\|f - \mathbb{E}_{X_1} f\|_{\Psi_\alpha(X_1)} \le c_\alpha n^{(d-1)/2} \|\max_j |X_{1,j}|\|_{\Psi_\alpha} \tag{5.6}$$

for any $x_2, \ldots, x_d$ in $\mathcal{E}_2$, where $\mathbb{E}_{X_1}$ denotes taking the expectation with respect to X_1 (which, by independence, is the same as conditionally on $X_2, \ldots, X_d$).

To continue, fix any realization $x_3, \ldots, x_d$ of the random vectors $X_3, \ldots, X_d$ that satisfy $\mathcal{E}_3$ and apply Proposition 3.1 to the function $f_2(x_2)$ given by $x_2 \mapsto \mathbb{E}_{X_1} f(X_1, x_2, \ldots, x_d)$. Again, f_2 is a convex function, and since

$$\begin{aligned}
&|\mathbb{E}_{X_1} f(X_1 \otimes x \otimes x_3 \otimes \ldots \otimes x_d) - \mathbb{E}_{X_1} f(X_1 \otimes y \otimes x_3 \otimes \ldots \otimes x_d)| \\
&\le \mathbb{E}_{X_1} \|X_1 \otimes (x-y) \otimes x_3 \otimes \ldots \otimes x_d\|_2 \le (\mathbb{E}\|X_1\|_2^2)^{1/2} \|x-y\|_2 \prod_{i=3}^{d} \|x_i\|_2 \\
&\le \sqrt{n}\|x-y\|_2 \cdot 2n^{(d-2)/2} = \|x-y\|_2 \cdot 2n^{(d-1)/2},
\end{aligned}$$

f_2 is $2n^{(d-1)/2}$-Lipschitz. Applying (3.2), we thus obtain

$$\|\mathbb{E}_{X_1} f - \mathbb{E}_{X_1,X_2} f\|_{\Psi_\alpha(X_2)} \le c_\alpha n^{(d-1)/2} \|\max_j |X_{2,j}|\|_{\Psi_\alpha} \tag{5.7}$$

for any $x_3, \ldots, x_d$ in $\mathcal{E}_3$. Iterating this procedure, we arrive at

$$\|\mathbb{E}_{X_1,\ldots,X_{k-1}} f - \mathbb{E}_{X_1,\ldots,X_k} f\|_{\Psi_\alpha(X_k)} \le c_\alpha n^{(d-1)/2} \|\max_j |X_{k,j}|\|_{\Psi_\alpha} \tag{5.8}$$

for any realization $x_{k+1}, \dots, x_d$ of $X_{k+1}, \dots, X_d$ in $\mathcal{E}_{k+1}$.

We now combine (5.8) for $k = 1, \dots, d$. To this end, we write

$$\Delta_k := \Delta_k(X_k, \dots, X_d) := \mathbb{E}_{X_1,\dots,X_{k-1}} f - \mathbb{E}_{X_1,\dots,X_k} f$$

and apply Proposition 5.2. Here we have to distinguish between the cases where $\alpha \in [1, 2]$ and $\alpha \in (0, 1)$. If $\alpha \geq 1$, we use (5.4) to arrive at a bound for the moment-generating function. Writing $M_k := \|\max_j |X_{k,j}|\|_{\Psi_\alpha}$, we obtain

$$\mathbb{E}\exp(\lambda\Delta_k) \leq \begin{cases} \exp((c_\alpha n^{(d-1)/2} M_k)^2 \lambda^2) \\ \exp((c_\alpha n^{(d-1)/2} M_k)^{\alpha/(\alpha-1)} |\lambda|^{\alpha/(\alpha-1)}) \end{cases}$$

for all $x_{k+1}, \dots, x_d$ in $\mathcal{E}_{k+1}$, where the first line holds if $|\lambda| \leq 1/(c_\alpha n^{(d-1)/2} M_k)$ and the second one if $|\lambda| \geq 1/(c_\alpha n^{(d-1)/2} M_k)$ and $\alpha > 1$. For the simplicity of presentation, temporarily assume that $c_\alpha n^{(d-1)/2} = 1$ (alternatively, replace M_k by $c_\alpha n^{(d-1)/2} M_k$ in the following arguments) and that $M_1 \leq \dots \leq M_d$. Using Lemma 5.5, we obtain

$$\begin{aligned} \mathbb{E}\exp(\lambda(f - \mathbb{E}f))1_{\mathcal{E}} &= \mathbb{E}\exp(\lambda(\Delta_1 + \cdots + \Delta_d))1_{\mathcal{E}} \\ &\leq \exp((M_1^2 + \dots + M_k^2)\lambda^2 + (M_{k+1}^{\alpha/(\alpha-1)} + \dots + M_d^{\alpha/(\alpha-1)})|\lambda|^{\alpha/(\alpha-1)}) \end{aligned}$$

for $|\lambda| \in [1/M_{k+1}, 1/M_k]$, where we formally set $M_0 := 0$ and $M_{d+1} := \infty$. In particular, setting $M := (M_1^2 + \dots + M_d^2)^{1/2}$, we have

$$\mathbb{E}\exp(\lambda(f - \mathbb{E}f))1_{\mathcal{E}} \leq \exp(M^2\lambda^2)$$

for all $|\lambda| \leq 1/M_d = 1/(\max_k M_k)$. Furthermore, for $\alpha > 1$, it is not hard to see that

$$(M_1^2 + \dots + M_k^2)\lambda^2 + (M_{k+1}^{\alpha/(\alpha-1)} + \dots + M_d^{\alpha/(\alpha-1)})|\lambda|^{\alpha/(\alpha-1)} \leq M^{\alpha/(\alpha-1)}|\lambda|^{\alpha/(\alpha-1)}$$

If $|\lambda| \in [1/M_{k+1}, 1/M_k]$ for some $k = 0, 1, \dots, d-1$ or $|\lambda| \in [1/M, 1/M_d]$ for $k = d$. Indeed, by monotonicity (divide by λ^2 and compare the coefficients), it suffices to check this for $\lambda = 1/M_{k+1}$ or $\lambda = 1/M$ if $k = d$. The cases of $k = 0$ and $k = d$ follow by simple calculations. In the general case, set $x^2 = (M_1^2 + \dots + M_{k+1}^2)/M_{k+1}^2$ and $y^{\alpha/(\alpha-1)} = (M_{k+2}^{\alpha/(\alpha-1)} + \dots + M_d^{\alpha/(\alpha-1)})/M_{k+1}^{\alpha/(\alpha-1)}$. Clearly, $(x^2 + y^{\alpha/(\alpha-1)})^{(\alpha-1)/\alpha} \leq (x^2 + y^2)^{1/2}$ since $x \geq 1$ and $\alpha/(\alpha-1) \geq 2$. Moreover, $y^2 \leq (M_{k+2}^2 + \dots + M_d^2)/M_{k+1}^2$, which proves the inequality. Altogether, inserting the factor $c_\alpha n^{(d-1)/2}$ again, we therefore obtain

$$\mathbb{E}\exp(\lambda(f - \mathbb{E}f))1_{\mathcal{E}} = \mathbb{E}\exp(\lambda(\Delta_1 + \cdots + \Delta_d))1_{\mathcal{E}}$$

$$\leq \begin{cases} \exp((c_\alpha n^{(d-1)/2})^2 M^2 \lambda^2) \\ \exp((c_\alpha n^{(d-1)/2})^{\alpha/(\alpha-1)} M^{\alpha/(\alpha-1)} |\lambda|^{\alpha/(\alpha-1)}), \end{cases} \tag{5.9}$$

where the first line holds if $|\lambda| \leq 1/(c_\alpha n^{(d-1)/2} M)$ and the second one if $|\lambda| \geq 1/(c_\alpha n^{(d-1)/2} M)$ and $\alpha > 1$.

On the other hand, if $\alpha < 1$, we use (5.3). Together with Lemma 5.5 and the subadditivity of $|\cdot|^\alpha$ for $\alpha \in (0, 1)$, this yields

$$\begin{aligned} &\mathbb{E} \exp(\lambda^\alpha |f - \mathbb{E} f|^\alpha) 1_{\mathcal{E}} \leq \mathbb{E} \exp(\lambda^\alpha (|\Delta_1|^\alpha + \cdots + |\Delta_d|^\alpha)) 1_{\mathcal{E}} \\ &\leq \exp((c_\alpha n^{(d-1)/2})^\alpha (M_1^\alpha + \cdots + M_d^\alpha) \lambda^\alpha) \end{aligned} \tag{5.10}$$

for $\lambda \in [0, 1/(c_\alpha n^{(d-1)/2} \max_k M_k)]$.

To finish the proof, first consider $\alpha \in [1, 2]$. Then, for any $\lambda > 0$, we have

$$\begin{aligned} \mathbb{P}(f - \mathbb{E} f > t) &\leq \mathbb{P}(\{f - \mathbb{E} f > t\} \cap \mathcal{E}) + \mathbb{P}(\mathcal{E}^c) \\ &\leq \mathbb{P}(\exp(\lambda(f - \mathbb{E} f)) 1_{\mathcal{E}} > \exp(\lambda t)) + \mathbb{P}(\mathcal{E}^c) \\ &\leq \exp\Big(-\Big(\frac{t}{c_\alpha n^{(d-1)/2} M}\Big)^\alpha\Big) + 2 \exp\Big(-\frac{1}{C_\alpha}\Big(\frac{n^{1/2}}{K^2 d^{1/2}}\Big)^\alpha\Big), \end{aligned} \tag{5.11}$$

where the last step follows by standard arguments (similarly as in the proof of Proposition 5.2 given in the appendix), using (5.9) and (5.5). Now, assume that $t \leq c_\alpha n^{d/2} M/(K^2 d^{1/2})$. Then, the right-hand side of (5.11) is dominated by the first term (possibly after adjusting constants), so that we arrive at

$$\mathbb{P}(f - \mathbb{E} f > t) \leq 3 \exp\Big(-\frac{1}{C_\alpha}\Big(\frac{t}{n^{(d-1)/2} M}\Big)^\alpha\Big).$$

The same arguments hold if f is replaced by $-f$. Adjusting constants by (2.3), we obtain that for any $t \in [0, c_\alpha n^{d/2} M/(K^2 d^{1/2})]$,

$$\mathbb{P}(|f(X) - \mathbb{E} f(X)| > t) \leq 2 \exp\Big(-\frac{1}{C_\alpha}\Big(\frac{t}{n^{(d-1)/2} M}\Big)^\alpha\Big). \tag{5.12}$$

Now it remains to note that by Lemma 5.6, we have

$$\|\max_j |X_{i,j}|\|_{\Psi_\alpha} \leq C_\alpha (\log n)^{1/\alpha} \max_j \|X_{i,j}\|_{\Psi_\alpha} \leq C_\alpha (\log n)^{1/\alpha} K.$$

If $\alpha \in (0, 1)$, similarly to (5.11), using (5.10), (5.5) and Proposition 5.2,

$$\mathbb{P}(|f - \mathbb{E} f| > t) \leq \mathbb{P}(\{|f - \mathbb{E} f| > t\} \cap \mathcal{E}) + \mathbb{P}(\mathcal{E}^c)$$

$$\leq 2\exp\Big(-\Big(\frac{t}{c_\alpha n^{(d-1)/2}M_\alpha}\Big)^\alpha\Big) + 2\exp\Big(-\frac{1}{C_\alpha}\Big(\frac{n^{1/2}}{K^2 d^{1/2}}\Big)^\alpha\Big),$$

where $M_\alpha := (M_1^\alpha + \ldots + M_d^\alpha)^{1/\alpha}$. The rest follows as above. □

Appendix A

Proof of Proposition 5.2 The equivalence of (1.1), (1.2), (1.3), and (5.3) is easily seen by directly adapting the arguments from the proof of [41, Proposition 2.5.2]. To see that these properties imply (5.4), first note that since in particular $\|X\|_{\Psi_1} < \infty$, the bound for $|\lambda| \leq 1/C'_{5,\alpha}$ directly follows from [41], Proposition 2.7.1 (e). To see the bound for large values of $|\lambda|$, we infer that by the weighted arithmetic–geometric mean inequality (with weights $\alpha - 1$ and 1),

$$y^{(\alpha-1)/\alpha} z^{1/\alpha} \leq \frac{\alpha-1}{\alpha} y + \frac{1}{\alpha} z$$

for any $y, z \geq 0$. Setting $y := |\lambda|^{\alpha/(\alpha-1)}$ and $z := |x|^\alpha$, we may conclude that

$$\lambda x \leq \frac{\alpha-1}{\alpha}|\lambda|^{\alpha/(\alpha-1)} + \frac{1}{\alpha}|x|^\alpha$$

for any $\lambda, x \in \mathbb{R}$. Consequently, using (5.3), assuming $C_{4,\alpha} = 1$, for any $|\lambda| \geq 1$,

$$\begin{aligned}\mathbb{E}\exp(\lambda X) &\leq \exp\big(\frac{\alpha-1}{\alpha}|\lambda|^{\alpha/(\alpha-1)}\big)\mathbb{E}\exp(|X|^\alpha/\alpha)\\ &\leq \exp\big(\frac{\alpha-1}{\alpha}|\lambda|^{\alpha/(\alpha-1)}\big)\exp(1/\alpha) \leq \exp(|\lambda|^{\alpha/(\alpha-1)}).\end{aligned}$$

This yields (5.4) for $|\lambda| \geq 1/C''_{5,\alpha}$. The claim now follows by taking $C_{5,\alpha} := \max(C'_{5,\alpha}, C''_{5,\alpha})$.

Finally, starting with (5.4), assuming $C_{5,\alpha} = 1$, let us check (1.1). To this end, note that for any $\lambda > 0$,

$$\mathbb{P}(X \geq t) \leq \exp(-\lambda t)\mathbb{E}\exp(\lambda X) \leq \exp(-\lambda t + \lambda^2 \mathbb{1}_{\{\lambda \leq 1\}} + \lambda^{\alpha/(\alpha-1)}\mathbb{1}_{\{\lambda > 1\}}).$$

Now choose $\lambda := t/2$ if $t \leq 2$, $\lambda := ((\alpha-1)t/\alpha)^{\alpha-1}$ if $t \geq \alpha/(\alpha-1)$, and $\lambda := 1$ if $t \in (2, \alpha/(\alpha-1))$. This yields

$$\mathbb{P}(X \geq t) \leq \begin{cases}\exp(-t^2/4) & \text{if } t \leq 2,\\ \exp(-(t-1)) & \text{if } t \in (2, \alpha/(\alpha-1)),\\ \exp(-\frac{(\alpha-1)^{\alpha-1}}{\alpha^\alpha}t^\alpha) & \text{if } t \geq \alpha/(\alpha-1).\end{cases}$$

Now use (2.3), (2.4), and the fact that $\exp(-(t-1)) \le \exp(-t^\alpha/C_\alpha^\alpha)$ for any $t \in (2, \alpha/(\alpha-1))$. It follows that

$$\mathbb{P}(X \ge t) \le 2\exp(-t^\alpha/C_{1,\alpha}'^\alpha)$$

for any $t \ge 0$. The same argument for $-X$ completes the proof. □

Proof of Lemma 5.3 By the arithmetic and geometric means inequality and since $\mathbb{E}\|X_i\|_2 \le \sqrt{n}$, for any $s \ge 0$,

$$\begin{aligned}\mathbb{P}\Big(\prod_{i=1}^d \|X_i\|_2 > (\sqrt{n}+s)^d\Big) &\le \mathbb{P}\Big(\frac{1}{d}\sum_{i=1}^d (\|X_i\|_2 - \sqrt{n}) > s\Big)\\ &\le \mathbb{P}\Big(\frac{1}{d}\sum_{i=1}^d (\|X_i\|_2 - \mathbb{E}\|X_i\|_2) > s\Big).\end{aligned} \tag{A.1}$$

Moreover, by (2.2) and [12, Corollary A.5],

$$\big\| \|X_i\|_2 - \mathbb{E}\|X_i\|_2 \big\|_{\Psi_\alpha} = \big\| \|X_i\|_2 - \sqrt{n} - (\mathbb{E}\|X_i\|_2 - \sqrt{n}) \big\|_{\Psi_\alpha} \le C_\alpha K^2$$

for any $i = 1, \ldots, d$. On the other hand, if $Y_1, \ldots, Y_d$ are independent centered random variables with $\|Y_i\|_{\Psi_\alpha} \le M$, we have

$$\begin{aligned}\mathbb{P}\Big(\frac{1}{d}\Big|\sum_{i=1}^d Y_i\Big| \ge s\Big) &\le 2\exp\Big(-\frac{1}{C_\alpha}\min\Big(\Big(\frac{s\sqrt{d}}{M}\Big)^2, \Big(\frac{s\sqrt{d}}{M}\Big)^\alpha\Big)\Big)\\ &\le 2\exp\Big(-\frac{1}{C_\alpha}\Big(\frac{s\sqrt{d}}{M}\Big)^\alpha\Big).\end{aligned}$$

Here, the first estimate follows from [10] ($\alpha > 1$) and [17] ($\alpha \le 1$), while the last step follows from (2.4). As a consequence, (A.1) can be bounded by $2\exp(-s^\alpha d^{\alpha/2}/(K^{2\alpha}C_\alpha))$.

For $u \in [0, 2]$ and $s = u\sqrt{n}/2d$, we have $(\sqrt{n}+s)^d \le n^{d/2}(1+u)$. Plugging in, we arrive at

$$\mathbb{P}\Big(\prod_{i=1}^d \|X_i\|_2 > n^{d/2}(1+u)\Big) \le 2\exp\Big(-\frac{1}{C_\alpha}\Big(\frac{n^{1/2}u}{K^2 d^{1/2}}\Big)^\alpha\Big).$$

Now set $u := t/n^{d/2}$. □

Proof of Lemma 5.4 Let us first recall the partition into "binary sets" that appears in the proof of [42, Lemma 3.2]. Here we assume that $d = 2^L$ for some $L \in \mathbb{N}$ (if not, increase d). Then, for any $\ell \in \{0, 1, \ldots, L\}$, we consider the partition $\mathcal{I}_\ell$ of $\{1, \ldots, d\}$ into 2^ℓ successive (integer) intervals of length $d_\ell := d/2^\ell$ that we call "binary intervals." It is not hard to see that for any $k = 1, \ldots, d$, we can partition $[1, k]$ into binary intervals of different lengths such that this partition contains at most one interval of each family $\mathcal{I}_\ell$.

Now it suffices to prove that

$$\mathbb{P}\Big(\exists 0 \le \ell \le L, \exists I \in \mathcal{I}_\ell \colon \prod_{i \in I} \|X_i\|_2 > (1 + 2^{-\ell/4}u)n^{d_\ell/2}\Big) \le 2\exp\Big(-\frac{1}{C_\alpha}\Big(\frac{n^{1/2}u}{K^2 d^{1/2}}\Big)^\alpha\Big)$$

(cf. Step 3 of the proof of [42, Lemma 3.2], where the reduction to this case is explained in detail). To this end, for any $\ell \in \{0, 1, \ldots, L\}$, any $I \in \mathcal{I}_\ell$, and $d_\ell := |I| = d/2^\ell$, we apply Lemma 5.3 for d_ℓ and $t := 2^{-\ell/4} n^{d_\ell/2} u$. This yields

$$\mathbb{P}\Big(\prod_{i \in I} \|X_i\|_2 > (1 + 2^{-\ell/4}u)n^{d_\ell/2}\Big) \le 2\exp\Big(-\frac{1}{C_\alpha}\Big(\frac{n^{1/2}u}{2^{\ell/4} K^2 d_\ell^{1/2}}\Big)^\alpha\Big) = 2\exp\Big(-\frac{1}{C_\alpha}\Big(2^{\ell/4}\frac{n^{1/2}u}{K^2 d^{1/2}}\Big)^\alpha\Big).$$

Altogether, we arrive at

$$\mathbb{P}\Big(\exists \ell \in \{0, 1, \ldots, L\}, \exists I \in \mathcal{I}_\ell \colon \prod_{i \in I} \|X_i\|_2 > (1 + 2^{-\ell/4}u)n^{d_\ell/2}\Big) \le \sum_{\ell=0}^{L} 2^\ell \cdot 2\exp\Big(-\frac{1}{C_\alpha}\Big(2^{\ell/4}\frac{n^{1/2}u}{K^2 d^{1/2}}\Big)^\alpha\Big). \tag{A.2}$$

We may now assume that $(n^{1/2}u/(K^2 d^{1/2}))^\alpha / C_\alpha \ge 1$ (otherwise the bound in Lemma 5.4 gets trivial by adjusting C_α). Using the elementary inequality $ab \ge (a + b)/2$ for all $a, b \ge 1$, we arrive at

$$2^{\ell\alpha/4}\frac{1}{C_\alpha}\Big(\frac{n^{1/2}u}{K^2 d^{1/2}}\Big)^\alpha \ge \frac{1}{2}\Big(2^{\ell\alpha/4} + \frac{1}{C_\alpha}\Big(\frac{n^{1/2}u}{K^2 d^{1/2}}\Big)^\alpha\Big).$$

Using this in (A.2), we obtain the upper bound

$$2\exp\Big(-\frac{1}{2C_\alpha}\Big(\frac{n^{1/2}u}{K^2 d^{1/2}}\Big)^\alpha\Big)\sum_{\ell=0}^{L} 2^\ell \exp(-2^{\ell\alpha/4-1}) \le c_\alpha \exp\Big(-\frac{1}{2C_\alpha}\Big(\frac{n^{1/2}u}{K^2 d^{1/2}}\Big)^\alpha\Big).$$

By (2.3), we can assume $c_\alpha = 2$. □

To prove Lemma 5.6, we first present a number of lemmas and auxiliary statements. In particular, recall that if $\alpha \in (0, \infty)$, then for any $x, y \in (0, \infty)$,

$$c_\alpha(x^\alpha + y^\alpha) \le (x + y)^\alpha \le \widetilde{c}_\alpha(x^\alpha + y^\alpha), \tag{A.3}$$

where $c_\alpha := 2^{\alpha-1} \wedge 1$ and $\widetilde{c}_\alpha := 2^{\alpha-1} \vee 1$. Indeed, if $\alpha \le 1$, using the concavity of the function $x \mapsto x^\alpha$, it follows by standard arguments that $2^{\alpha-1}(x^\alpha + y^\alpha) \le (x+y)^\alpha \le x^\alpha + y^\alpha$. Likewise, for $\alpha \ge 1$, using the convexity of $x \mapsto x^\alpha$, we obtain $x^\alpha + y^\alpha \le (x + y)^\alpha \le 2^{\alpha-1}(x^\alpha + y^\alpha)$.

Lemma A.1 *Let $X_1, \dots, X_n$ be independent, centered random variables such that $\|X_i\|_{\Psi_\alpha} \le 1$ for some $\alpha > 0$. Then, if $Y := \max_i |X_i|$ and $c := (c_\alpha^{-1} \log n)^{1/\alpha}$, we have*

$$\mathbb{P}(Y \ge c + t) \le 2\exp(-c_\alpha t^\alpha)$$

with c_α as in (A.3).

Proof We have

$$\begin{aligned}\mathbb{P}(Y \ge c + t) &\le n\mathbb{P}(|X_i| \ge c + t) \le 2n\exp(-(c + t)^\alpha)\\ &\le 2n\exp(-c_\alpha(t^\alpha + c^\alpha) = 2\exp(-c_\alpha t^\alpha),\end{aligned}$$

where we have used (A.3) in the next-to-last step. □

Lemma A.2 *Let $Y \ge 0$ be a random variable that satisfies*

$$\mathbb{P}(Y \ge c + t) \le 2\exp(-t^\alpha)$$

for some $c \ge 0$ and any $t \ge 0$. Then,

$$\|Y\|_{\Psi_\alpha} \le \widetilde{c}_\alpha^{1/\alpha} \max\left\{\left(\frac{\sqrt{2}+1}{\sqrt{2}-1}\right)^{1/\alpha}, c\left(\frac{2}{\log 2}\right)^{1/\alpha}\right\}$$

with $\widetilde{c}_\alpha$ as in (A.3).

Proof By (A.3) and monotonicity, we have $Y^\alpha \le \widetilde{c}_\alpha((Y - c)_+^\alpha + c^\alpha)$, where $x_+ := \max(x, 0)$. Thus,

$$\begin{aligned}\mathbb{E}\exp\left(\frac{Y^\alpha}{s^\alpha}\right) &\le \exp\left(\frac{\widetilde{c}_\alpha c^\alpha}{s^\alpha}\right)\mathbb{E}\exp\left(\frac{\widetilde{c}_\alpha(Y - c)_+^\alpha}{s^\alpha}\right)\\ &= \exp\left(\frac{c^\alpha}{t^\alpha}\right)\mathbb{E}\exp\left(\frac{(Y - c)_+^\alpha}{t^\alpha}\right) =: I_1 \cdot I_2,\end{aligned}$$

where we have set $t := s\widetilde{c}_\alpha^{-1/\alpha}$. Obviously, $I_1 \le \sqrt{2}$ if $t \ge c(1/\log\sqrt{2})^{1/\alpha}$. As for I_2, we have

$$I_2 = 1 + \int_1^\infty \mathbb{P}((Y-c)_+ \geq t(\log y)^{1/\alpha})dy$$
$$\leq 1 + 2\int_1^\infty \exp(-t^\alpha \log y)dy = 1 + 2\int_1^\infty \frac{1}{y^{t^\alpha}}dy \leq \sqrt{2}$$

if $t \geq ((\sqrt{2}+1)/(\sqrt{2}-1))^{1/\alpha}$. Therefore, $I_1 I_2 \leq 2$ if $t \geq \max\{((\sqrt{2}+1)/(\sqrt{2}-1))^{1/\alpha}, c(2/\log 2)^{1/\alpha}\}$, which finishes the proof. □

Having these lemmas at hand, the proof of Lemma 5.6 is easily completed.

Proof of Lemma 5.6 The random variables $\hat{X}_i := X_i/K$ obviously satisfy the assumptions of Lemma A.1. Hence, setting $Y := \max_i |\hat{X}_i| = K^{-1}\max_i |X_i|$,

$$\mathbb{P}(c_\alpha^{1/\alpha} Y \geq (\log n)^{1/\alpha} + t) \leq 2\exp(-t^\alpha).$$

Therefore, we may apply Lemma A.2 to $\hat{Y} := c_\alpha^{1/\alpha} K^{-1} \max_i |X_i|$. This yields

$$\|\hat{Y}\|_{\Psi_\alpha} \leq \widetilde{c}_\alpha^{1/\alpha} \max\left\{\left(\frac{\sqrt{2}+1}{\sqrt{2}-1}\right)^{1/\alpha}, (\log n)^{1/\alpha}\left(\frac{2}{\log 2}\right)^{1/\alpha}\right\},$$

i. e., the claim of Lemma 5.6, where we have set $C := (\widetilde{c}_\alpha c_\alpha^{-1})^{1/\alpha}$. □

Acknowledgments This work 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*." The author would moreover like to thank Arthur Sinulis for carefully reading this paper and many fruitful discussions and suggestions.

References

1. R. Adamczak, A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. Electron. J. Probab. **13**(34), 1000–1034 (2008)
2. R. Adamczak, A note on the Hanson-Wright inequality for random vectors with dependencies. Electron. Commun. Probab. **20**(72), 13 (2015)
3. R. Adamczak, R. Latała, Tail and moment estimates for chaoses generated by symmetric random variables with logarithmically concave tails. Ann. Inst. Henri Poincaré Probab. Stat. **48**(4), 1103–1136 (2012)
4. R. Adamczak, R. Latała, R. Meller, Hanson-Wright inequality in Banach spaces. Ann. Inst. Henri Poincaré Probab. Stat. **56**(4), 2356–2376 (2020)
5. R. Adamczak, M. Strzelecki, On the convex Poincaré inequality and weak transportation inequalities. Bernoulli **25**(1), 341–374 (2019)
6. R. Adamczak, P. Wolff, Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. Probab. Theory Relat. Fields **162**(3–4), 531–586 (2015)
7. S. Boucheron, G. Lugosi, P. Massart, *Concentration Inequalities* (Oxford University Press, Oxford, 2013). A nonasymptotic theory of independence, With a foreword by Michel Ledoux

8. V.H. de la Peña, E. Giné, *Decoupling*. Probability and Its Applications (New York) (Springer, New York, 1999)
9. L.H. Dicker, M.A. Erdogdu, Flexible results for quadratic forms with applications to variance components estimation. Ann. Stat. **45**(1), 386–414 (2017)
10. E.D. Gluskin, S. Kwapień, Tail and moment estimates for sums of independent random variables with logarithmically concave tails. Studia Math. **114**(3), 303–309 (1995)
11. F. Götze, H. Sambale, A. Sinulis, Concentration inequalities for bounded functionals via log-Sobolev-type inequalities. J. Theor. Probab. **34**(3), 1623–1652 (2021)
12. F. Götze, H. Sambale, A. Sinulis, Concentration inequalities for polynomials in α-sub-exponential random variables. Electron. J. Probab. **26**(48), 22 (2021)
13. N. Gozlan, C. Roberto, P.-M. Samson, From dimension free concentration to the Poincaré inequality. Calc. Var. Partial Differ. Equ. **52**(3–4), 899–925 (2015)
14. N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, P. Tetali, Characterization of a class of weak transport-entropy inequalities on the line. Ann. Inst. Henri Poincaré Probab. Stat. **54**(3), 1667–1693 (2018)
15. N. Gozlan, C. Roberto, P.-M. Samson, P. Tetali, Kantorovich duality for general transport costs and applications. J. Funct. Anal. **273**(11), 3327–3405 (2017)
16. D.L. Hanson, F.T. Wright, A bound on tail probabilities for quadratic forms in independent random variables. Ann. Math. Stat. **42**, 1079–1083 (1971)
17. P. Hitczenko, S.J. Montgomery-Smith, K. Oleszkiewicz, Moment inequalities for sums of certain independent symmetric random variables. Studia Math. **123**(1), 15–42 (1997)
18. D. Hsu, S.M. Kakade, T. Zhang, A tail inequality for quadratic forms of sub-Gaussian random vectors. Electron. Commun. Probab. **17**(52), 6 (2012)
19. W.B. Johnson, G. Schechtman, Remarks on Talagrand's deviation inequality for Rademacher functions, in *Functional Analysis (Austin, TX, 1987/1989)*. Lecture Notes in Mathematics, vol. 1470 (Springer, Berlin, 1991), pp. 72–77
20. Y. Klochkov, N. Zhivotovskiy, Uniform Hanson–Wright type concentration inequalities for unbounded entries via the entropy method. Electron. J. Probab. **25**(22), 30 (2020)
21. K. Kolesko, R. Latała, Moment estimates for chaoses generated by symmetric random variables with logarithmically convex tails. Stat. Probab. Lett. **107**, 210–214 (2015)
22. F. Krahmer, S. Mendelson, H. Rauhut, Suprema of chaos processes and the restricted isometry property. Commun. Pure Appl. Math. **67**(11), 1877–1904 (2014)
23. A.K. Kuchibhotla, A. Chakrabortty, Moving Beyond Sub-Gaussianity in High-Dimensional Statistics: Applications in Covariance Estimation and Linear Regression. arXiv preprint, 2018
24. R. Latała, Tail and moment estimates for sums of independent random vectors with logarithmically concave tails. Studia Math. **118**(3), 301–304 (1996)
25. R. Latała, Tail and moment estimates for some types of chaos. Studia Math. **135**(1), 39–53 (1999)
26. R. Latała, Estimates of moments and tails of Gaussian chaoses. Ann. Probab. **24**(6), 2315–2331 (2006)
27. R. Latała, R. Łochowski, Moment and tail estimates for multidimensional chaos generated by positive random variables with logarithmically concave tails, in *Stochastic Inequalities and Applications*. Progress in Probability, vol. 56 (Birkhäuser, Basel, 2003), pp. 77–92
28. J. Lederer, S. van de Geer, New concentration inequalities for suprema of empirical processes. Bernoulli **20**(4), 2020–2038 (2014)
29. M. Ledoux, On Talagrand's deviation inequalities for product measures. ESAIM Probab. Stat. **1**, 63–87 (1995/1997)
30. M. Ledoux, M. Talagrand, *Probability in Banach Spaces: Isoperimetry and Processes*. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 23 (Springer, Berlin, 1991)
31. A. Marchina, Concentration inequalities for separately convex functions. Bernoulli **24**(4A), 2906–2933 (2018)
32. P. Massart, Some applications of concentration inequalities to statistics. Ann. Fac. Sci. Toulouse Math. (6) **9**(2), 245–303 (2000)

33. M. Rudelson, R. Vershynin, Hanson-Wright inequality and sub-Gaussian concentration. Electron. Commun. Probab. **18**(82), 9 (2013)
34. H. Sambale, A. Sinulis, Logarithmic Sobolev inequalities for finite spin systems and applications. Bernoulli **26**(3), 1863–1890 (2020)
35. H. Sambale, A. Sinulis, Modified log-Sobolev inequalities and two-level concentration. ALEA Lat. Am. J. Probab. Math. Stat. **18**, 855–885 (2021)
36. P.-M. Samson, Concentration of measure inequalities for Markov chains and Φ-mixing processes. Ann. Probab. **28**(1), 416–461 (2000)
37. P.-M. Samson, Concentration inequalities for convex functions on product spaces, in *Stochastic Inequalities and Applications*. Progress in Probability, vol. 56 (Birkhäuser, Basel, 2003), pp. 33–52
38. M. Talagrand, An isoperimetric theorem on the cube and the Khintchine-Kahane inequalities. Proc. Am. Math. Soc. **104**(3), 905–909 (1988)
39. M. Talagrand, New concentration inequalities in product spaces. Invent. Math. **126**(3), 505–563 (1996)
40. S. van de Geer, J. Lederer, The Bernstein-Orlicz norm and deviation inequalities. Probab. Theory Relat. Fields **157**(1–2), 225–250 (2013)
41. R. Vershynin, *High-Dimensional Probability*. Cambridge Series in Statistical and Probabilistic Mathematics, vol. 47 (Cambridge University Press, Cambridge, 2018)
42. R. Vershynin, Concentration inequalities for random tensors. Bernoulli **26**(4), 3139–3162 (2020)
43. V.H. Vu, K. Wang, Random weighted projections, random quadratic forms and random eigenvectors. Random Struct. Algorithm **47**(4), 792–821 (2015)