CONCENTRATION INEQUALITIES FOR BOUNDED FUNCTIONALS VIA GENERALIZED LOG-SOBOLEV INEQUALITIES

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

Abstract. In this paper we prove multilevel concentration inequalities for bounded functionals of random variables $X_1, \ldots, X_n$ that satisfy certain logarithmic Sobolev inequalities (LSI). We show that in the independent case, an LSI holds with a constant 1, irrespective of the distribution of $X_1, \ldots, X_n$. The constants in the tail estimates depend on k-tensors of higher order differences of the functional f, with either the average Hilbert–Schmidt or operator-type norms.

As an application, we demonstrate how these LSI results lead to significant improvements of concentration properties of polynomials in Bernoulli random variables. We derive exponential tail estimates for the subgraph counts in Erdös–Rényi random graphs for k-cycles with an optimal rate in a p-range that is (up to logarithmic corrections) optimal. This leads to the solution of the "infamous" triangle problem via concentration inequalities for polynomials, and extends it to cycles of all orders.

Lastly, we show that the tail estimates can be further sharpened for certain (non-independent) random variables $X_1, \ldots, X_n$ with values in finite spaces, if these satisfy an even stronger logarithmic Sobolev inequality. We apply the theory to suprema of homogeneous chaos in bounded random variables in the Banach space case. This generalizes earlier results of Talagrand and Boucheron–Bousquet–Lugosi–Massart.

1. Introduction

During the last forty years, the *concentration of measure phenomenon* has become an established part of probability theory with applications in numerous fields, see for example [MS86; Led01; BLM13; RS14; vH16]. One way to prove concentration of measure is by using functional inequalities, more specifically the *entropy method*. It has emerged as a way to prove several groundbreaking concentration inequalities in product spaces by Talagrand [Tal91; Tal96], mainly in the works of Ledoux [Led97] and Bobkov and Ledoux [BL97]. The connection between an LSI and subgaussian tail estimates has been known before (the so-called *Herbst argument*) and can also be found in [AS94]. The tensorization property of logarithmic Sobolev inequalities is well known and allowed to prove concentration properties for functionals of high-dimensional vectors of independent random variables $X_1, \ldots, X_n$. However, the starting point was a logarithmic Sobolev inequality for the individual distributions of $X_1, \ldots, X_n$.

In this work, we propose a generalized logarithmic Sobolev inequality that is valid for bounded functionals of independent random variables, irrespective of the underlying

[1]Faculty of Mathematics, Bielefeld University, Bielefeld, Germany

E-mail address: goetze@math.uni-bielefeld.de, hsambale@math.uni-bielefeld.de, asinulis@math.uni-bielefeld.de.

Date: December 3, 2018.

1991 *Mathematics Subject Classification.* Primary 60E15, 05C80.

Key words and phrases. concentration of measure, functional inequalities, logarithmic Sobolev inequality, product spaces, suprema of chaos, Erdös–Rényi graphs.

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*".

distributions. To convey the idea, let us recall the logarithmic Sobolev inequality for the standard Gaussian measure μ in $\mathbb{R}^n$ (see [Gro75]) states that for any $f \in C_c^\infty(\mathbb{R}^n)$ we have

$$\mathrm{Ent}_\mu(f^2) \le 2 \int |\nabla f|^2 d\mu, \tag{1.1}$$

where $\mathrm{Ent}_\mu(f^2) = \int f^2 \log f^2 d\mu - \int f^2 d\mu \log f^2 d\mu$ is the *entropy functional*. Informally, it bounds the disorder of a function f (under μ) by its average local fluctuations, measured in terms of the length of the gradient. It is also known that if μ is a measure on a discrete set $\mathcal{X}$ (or a more abstract set not allowing for a replacement for $|\nabla f|$), then there are several ways to reformulate equation (1.1), see e.g. [DS96] or [BT06]. We will continue these thoughts and work in the framework of *difference operators*. Given any probability space with measure μ, we call any operator $\Gamma : L^\infty(\mu) \to L^\infty(\mu)$ satisfying $|\Gamma(af+b)| = a\,|\Gamma f|$ for all $a > 0$, $b \in \mathbb{R}$ a difference operator. Accordingly, given Γ, we say that μ satisfies a $\Gamma-\mathrm{LSI}(\sigma^2)$, if for all bounded functionals we have

$$\mathrm{Ent}_\mu(f^2) \le 2\sigma^2 \int \Gamma(f)^2 d\mu. \tag{1.2}$$

By abuse of language, we say that $X = (X_1, \ldots, X_n)$ satisfies a $\Gamma-\mathrm{LSI}(\sigma^2)$, if its does so. Due to the shift-invariance and homogeneity of difference operators, this also implies a *Poincaré inequality*

$$\mathrm{Var}_\mu(f) \le \sigma^2 \int \Gamma(f)^2 d\mu. \tag{1.3}$$

where $\mathrm{Var}_\mu(f) = \int f^2 d\mu - (\int f d\mu)^2$ is the variance functional, which can be shown by a Taylor expansion of $x \log x$ and is by now classical.

Apart from the domain of Γ, it is clear that (1.2) can be seen as generalization to (1.1) by defining $\Gamma(f) = |\nabla f|$ on $\mathbb{R}^n$. The aim of this work is to show that the freedom of choosing a suitable difference operator Γ leads to interesting results in the setting of both independent and dependent random variables. A specific choice of Γ will lead to a universal inequality of the type (1.2).

Throughout this note, $X = (X_1, \ldots, X_n)$ is a random vector taking values in some product space $\mathcal{Y} = \otimes_{i=1}^n \mathcal{X}_i$ (equipped with the product σ-algebra) defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We denote the law of X by μ. In any finite-dimensional vector space, we let $|\cdot|$ be the Euclidean norm. Moreover, for any probability measure $\mathbb{P}$ and any random k-tensor A we write for any $p \in (0, \infty]$

$$\|A\|_{\mathrm{HS},p} = \left(\mathbb{E}\,|A|_{\mathrm{HS}}^p\right)^{1/p}, \tag{1.4}$$

$$\|A\|_{\mathrm{op},p} = \left(\mathbb{E}|A|_{\mathrm{op}}^p\right)^{1/p}. \tag{1.5}$$

Here, $|A|_{\mathrm{HS}}$ is the Hilbert–Schmidt and $|A|_{\mathrm{op}}$ the operator norm, see Section 3.

1.1. **Main results.** To present the general logarithmic Sobolev inequality, we need to introduce a specific difference operator $\mathfrak{h}$, which is frequently used in the method of bounded differences. Let $X' = (X'_1, \ldots, X'_n)$ be an independent copy of X, defined on the same probability space. Given $f(X) \in L^\infty(\mathbb{P})$, define for each $i \in \{1, \ldots, n\}$

$$T_i f(X) = T_i f = f(X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_n) =: f(\overline{X}_i, X'_i) \tag{1.6}$$

and

$$\mathfrak{h}_i f(X) = \|f(X) - T_i f(X)\|_{i,\infty}, \qquad \mathfrak{h} f = (\mathfrak{h}_1 f, \ldots, \mathfrak{h}_n f), \tag{1.7}$$

where $\|\cdot\|_{i,\infty}$ denotes the L^∞-norm with respect to (X_i, X'_i). The difference operator $|\mathfrak{h} f|$ is given as the Euclidean norm of the vector $\mathfrak{h} f$. Clearly, depending on the random vector $(X_j)_{j \neq i}$, $\mathfrak{h}_i f$ provides a uniform upper bound on the differences with respect to the i-th coordinate.

The first main result establishes that a $\mathfrak{h}-\mathrm{LSI}(1)$ holds for any product probability measure.

Theorem 1.1. *Let $X = (X_1, \ldots, X_n)$ be a random vector with independent components and values in any measurable space $\mathcal{Y} = \bigotimes_{i=1}^n \mathcal{X}_i$. For any functional $f : \mathcal{Y} \to \mathbb{R}$ with $f(X) \in L^\infty(\mathbb{P})$ we have*

$$\mathrm{Ent}(f(X)^2) \le 2\,\mathbb{E}|\mathfrak{h}f(X)|^2. \tag{1.8}$$

It is interesting to note that Theorem 1.1 can be generalized to allow for weak dependence. Since this requires more definitions and notations, we postpone the formulation and the proof to Section 2, see Theorem 2.1.

To the best of our knowledge, Theorem 1.1 is new. It might be compared to the Efron–Stein inequality (see e.g. [ES81; Ste86]) which is a counterpart of the tensorization property for the variance, and can be regarded as a universal Poincaré inequality for product measures (see e.g. [BGS18] for such an interpretation). A similar inequality with entropy replaced by variance can for example be found in [vH16, Lemma 2.1]. It can also be deduced from (1.8) in the usual way.

As an easy corollary, we obtain McDiarmid's inequality with a slightly worse constant.

Corollary 1.2. *Let $X_1, \ldots, X_n$ be independent random variables and $f = f(X)$ such that $\mathfrak{h}_i f(X) \le c_i$ for all $i = 1, \ldots, n$. Then*

$$\mathbb{P}\left(|f(X) - \mathbb{E}\, f(X)| \ge t\right) \le 2\exp\left(-\frac{t^2}{32e\sum_{i=1}^n c_i^2}\right). \tag{1.9}$$

Clearly, since Theorem 1.1 can be generalized to weakly dependent random variables, so does Corollary 1.2. However, Corollary 1.2 is a result based on concentration of first order only. Theorem 1.1 allows for extensions to arbitrary orders, which we develop next.

Starting from the $\mathfrak{h}-\mathrm{LSI}(\sigma^2)$, it is possible to prove tail estimates for the fluctuations of $f(X)$ around its mean. The next theorem is a significant improvement of [BGS18, Theorem 1.1], replacing the Hilbert-Schmidt norms appearing therein by operator norms in the independent case, as well as generalizing it to the non-independent setting. This leads to much sharper bounds and wider range of applications.

Theorem 1.3. *Let $X = (X_1, \ldots, X_n)$ be a random vector and $f : \mathcal{Y} \to \mathbb{R}$ a measurable function satisfying $f(X) \in L^\infty(\mathbb{P})$.*

If X satisfies an $\mathfrak{h}-\mathrm{LSI}(\sigma^2)$, then there is a constant $C > 0$ depending on d only, such that for any $t > 0$

$$\mathbb{P}\left(|f(X) - \mathbb{E}\, f(X)| \ge t\right) \le 2\exp\left(-\frac{1}{C\sigma^2}\min_{k=1,\ldots,d-1}\left(\frac{t}{\|\mathfrak{h}^{(k)}f\|_{\mathrm{HS},1}}\right)^{2/k} \wedge \left(\frac{t}{\|\mathfrak{h}^{(d)}f\|_{\mathrm{HS},\infty}}\right)^{2/d}\right). \tag{1.10}$$

If X has independent components, then there exists a constant $C > 0$ depending on d such that the estimate

$$\mathbb{P}\left(|f(X) - \mathbb{E}\, f(X)| \ge t\right) \le 2\exp\left(-\frac{1}{C}\min_{k=1,\ldots,d-1}\left(\frac{t}{\|\mathfrak{h}^{(k)}f\|_{\mathrm{op},1}}\right)^{2/k} \wedge \left(\frac{t}{\|\mathfrak{h}^{(d)}f\|_{\mathrm{op},\infty}}\right)^{2/d}\right) \tag{1.11}$$

holds.

Since defining the (iterated) difference operators $\mathfrak{h}^{(j)}$ and the operator norms for k-tensors is rather lengthy, these are given in Section 3. They can be thought of as analogues of the k-tensors of all partial derivatives of order k in the abstract setting.

At first sight, (1.11) may seem stronger than (1.10) due to the inequality $|A|_{\mathrm{op}} \le |A|_{\mathrm{HS}}$ for any k-tensor A. However, note that in (1.11) we do not have the parameter σ^2 anymore, so that (1.10) and (1.11) can be incomparable and each have their own merit. In particular, the presence of the logarithmic Sobolev constant σ^2 can lead to sharper results if σ^2 is small. For example, in the case of Erdös-Rényi random graphs, it is favorable to use (1.10), since σ^2 tends to zero, which leads to Corollaries 1.6 and 1.7.

For a class of random variables $X_1, \ldots, X_n$, we can sharpen the tail estimates (1.10) further. To this end, we will introduce another difference operator, which is more familiar in the context of logarithmic Sobolev inequalities for Markov chains, as developed in [DS96]. Assume that $\mathcal{Y} = \mathcal{X}^n$ for some finite set $\mathcal{X}$, and let X be a $\mathcal{Y}$-valued random vector. We define

$$
\begin{aligned}
|\mathfrak{d} f|^2 &:= \sum_{i=1}^n (\mathfrak{d}_i f)^2 := \sum_{i=1}^n \frac{1}{2} \iint (f(\overline{x}_i, y) - f(\overline{x}_i, y'))^2 d\mu(y \mid \overline{x}_i) d\mu(y' \mid \overline{x}_i) \\
&= \sum_{i=1}^n \iint (f(\overline{x}_i, y) - f(\overline{x}_i, y'))_+^2 d\mu(y \mid \overline{x}_i) d\mu(y' \mid \overline{x}_i),
\end{aligned} \tag{1.12}
$$

where we denote by $\mu(\cdot \mid \overline{x}_i)$ the conditional measure (interpreted as a measure on $\mathcal{X}$) and by $\overline{\mu}_i$ the marginal on $\mathcal{X}^{n-1}$. It appears naturally in the Dirichlet form associated to the Glauber dynamic of μ, which is given by

$$
\mathcal{E}(f,f) := \sum_{i=1}^n \int \mathrm{Var}_{\mu(\cdot \mid \overline{x}_i)}(f(\overline{x}_i, \cdot)) d\overline{\mu}_i(\overline{x}_i) = \int |\mathfrak{d} f|^2 d\mu. \tag{1.13}
$$

In [SS18], the authors show various examples of random variables satisfying a $\mathfrak{d}-\mathrm{LSI}(\sigma^2)$.

Theorem 1.4. *Let $X_1, \ldots, X_n$ be random variables satisfying a $\mathfrak{d}-\mathrm{LSI}(\sigma^2)$. There is a constant $C > 0$ depending on d such that for any $f(X) \in L^\infty(\mathbb{P})$ and $t > 0$*

$$
\mathbb{P}\left(|f(X) - \mathbb{E} f(X)| \geq t\right) \leq 2 \exp\left(-\frac{1}{C\sigma^2} \min_{k=1,\ldots,d-1} \left(\frac{t}{\|\mathfrak{h}^{(k)} f\|_{\mathrm{op},1}} \right)^{2/k} \wedge \left(\frac{t}{\|\mathfrak{h}^{(d)} f\|_{\mathrm{op},\infty}} \right)^{2/d} \right). \tag{1.14}
$$

Clearly, an inequality of the type (1.14) is stronger than (1.10). Therefore it would be desirable to prove an analogue of Theorem 1.1 with $\mathfrak{d}$ instead of $\mathfrak{h}$. Unfortunately the inequality (1.8) cannot hold in this generality with $\mathfrak{d}$. Indeed, in the context of logarithmic Sobolev inequalities, the behavior of $\mathfrak{h}$ and $\mathfrak{d}$ is fundamentally different. This can be illustrated by considering the case of independent random variables $X_1, \ldots, X_n$. Theorem 1.1 establishes an $\mathfrak{h}-\mathrm{LSI}(1)$ irrespective of the distribution, whereas the property of satisfying a logarithmic Sobolev inequality with respect to $\mathfrak{d}$ is quite restrictive. If $(\mathcal{Y}, \mathcal{A}, \mu)$ is a probability space with a sequence $A_n \in \mathcal{A}$ with $\mu(A_n) \to 0$, then choosing the sequence of functions $f_n := \mathbb{1}_{A_n} \in L^\infty(\mu)$, we have

$$
\mathrm{Ent}_\mu(f_n^2) = \mu(A_n) \log(1/\mu(A_n)).
$$

On the other hand, we have $\int (\mathfrak{d} f_n)^2 d\mu = \mu(A_n)(1 - \mu(A_n))$, so that a $\mathfrak{d}-\mathrm{LSI}(\sigma^2)$ cannot hold. Hence $\mathfrak{d}$ is a difference operator which essentially makes sense in finite situations only.

Moreover, this shows that one cannot expect a "entropy" version of the Efron–Stein inequality of the form $\mathrm{Ent}_\mu(f^2) \leq \mathbb{E}|\mathfrak{d} f|^2$. It seems that for the entropy, it is essential to change the difference operator.

The tail estimates of Theorem 1.3 can also be sharpened by considering special classes of functions. This includes so-called multilinear (or tetrahedral) polynomials and chaos-type functionals. In particular, for such functions it is possible to replace (1.10) by estimates which depend on possibly sharper norms, which will be discussed in the next subsections.

1.1.1. *Polynomials and subgraph counts in Erdös–Rényi graphs.* The first class of functions we consider are polynomials. Let $(X_1, \ldots, X_n)$ be a random vector in $\mathbb{R}^n$ with law μ

supported in $[a,b]^n, a < b$, satisfying an $\mathfrak{h}-\mathrm{LSI}(\sigma^2)$. Let $f_d : \mathbb{R}^n \to \mathbb{R}$ be a multilinear (also called tetrahedral) polynomial of degree d, i.e. a polynomial of the form

$$f_d(x) := \sum_{k=1}^{d} \sum_{\substack{I \subset \{1,\dots,n\} \\ |I|=k}} a_I^k x_I \tag{1.15}$$

for symmetric k-tensors a^k with vanishing diagonal. A k-tensor a^k is called symmetric, if $a^k_{i_1\dots i_k} = a^k_{\sigma(i_1)\dots\sigma(i_k)}$ for any permutation $\sigma \in \mathcal{S}_k$, and the diagonal is defined as $\Delta_k := \{(i_1,\dots,i_k) : |\{i_1,\dots,i_k\}| < k\}$. Denote by $\nabla^k f$ the k-tensor of all partial derivatives of order k of f. We have the following result.

Theorem 1.5. *Let $d \in \mathbb{N}$, X be a random vector with law μ supported in $[-1,+1]^n$ and satisfying an $\mathfrak{h}-\mathrm{LSI}(\sigma^2)$, and let f_d be as in* (1.15)*. There exists a constant $C > 0$ depending on d only such that for all $t > 0$*

$$\mathbb{P}\left(|f_d(X) - \mathbb{E}\, f_d(X)| \geq t\right) \leq 2\exp\left(-\frac{1}{C} \min_{k=1,\dots,d} \min_{\mathcal{I} \in P_k} \left(\frac{t}{\sigma^k \|\mathbb{E}\,\nabla^k f_d(X)\|_{\mathcal{I}}}\right)^{2/|\mathcal{I}|}\right). \tag{1.16}$$

The family of norms $\|\cdot\|_{\mathcal{I}}$ arises by different embeddings of copies of $\mathbb{R}^n$ into the space of all tensors and will be defined in Section 5. It has been first introduced in [Lat06].

Note that Theorem 1.5 applies in particular to independent, real-valued random variables $X_1,\dots,X_n$. Moreover, Theorem 1.5 generalizes [AKPS18, Theorem 2.2]. Although the approach is similar as in [AKPS18], our setup allows to apply it to *any* random vector satisfying an $\mathfrak{h}-\mathrm{LSI}(\sigma^2)$. The Ising model treated therein can be considered as a special case, since in [GSS18] we have shown that the Ising model satisfies an inequality which implies an $\mathfrak{h}-\mathrm{LSI}(\sigma^2)$.

An important application of concentration inequalities for polynomials are random graphs. Indeed, one can describe a random graph by a family of $\{0,1\}$-valued random variables X_j corresponding to the presence or absence of edges. One is often interested in the fluctuations of the number of a fixed graph H in the random graph, which is a multilinear polynomial of order $|E(H)|$. For example, the Erdös–Rényi graph $G(n,p)$ on n vertices is a random graph with independent random variables such that $\mathbb{P}(X_e = 1) = p$ for each edge e.

Corollary 1.6. *Let X be the Erdös–Rényi model with parameters $n, p = p(n)$, assume that $p \geq \frac{\log(n)^4}{n}$ and let $T_3(X) := \sum_{i\neq j\neq k} X_{ij}X_{jk}X_{ik}$ be the number of triangles. We have for some constant $C > 0$ and any $t > 0$*

$$\mathbb{P}\left(|T_3 - \mathbb{E}\,T_3| \geq t\right) \leq 2\exp\left(-\frac{1}{C}\min\left(\frac{t^2}{\sigma^6 n^3 + \sigma^4 p^2 n^3 + \sigma^2 p^4 n^4}, \frac{t}{\sigma^3 n^{1/2} + \sigma^2 pn}, \frac{t^{2/3}}{\sigma^2}\right)\right).$$

Consequently, for $t = \varepsilon\,\mathbb{E}\,T_3 = \varepsilon n(n-1)(n-2)p^3$

$$\mathbb{P}\left(|T_3 - \mathbb{E}\,T_3| \geq \varepsilon\,\mathbb{E}\,T_3\right) \leq 2\exp\Big(-C(\varepsilon) n^2 p^2 \log(1/p)\Big). \tag{1.17}$$

Actually, a careful inspection of the proof shows that we have

$$\mathbb{P}\left(|T_3 - \mathbb{E}\,T_3| \geq \varepsilon\,\mathbb{E}\,T_3\right) \leq 2\exp\left(-\frac{1}{C} n^2 p^2 \log(1/p) \min\left(\frac{\varepsilon^2}{3}, \frac{\varepsilon}{3}, \frac{\varepsilon^{2/3}}{2}\right)\right).$$

Corollary 1.6 gives an alternative solution to the upper tail problem of the triangle count in the range $p \geq \log(n)^4 n^{-1}$ and can be compared to the result in [Cha12, Theorem 1.1] and [DK12a, Theorem 1.1], where this upper tail bound was first proven. Thus it gives a solution to the classical question of the "infamous upper bound" ([JR02]) of the number of triangles.

It is possible to derive similar results as in Corollary 1.6 for other subgraphs. For example, we can sharpen [AW15, Proposition 5.6] for the subgraph counts of k-cycles ($k \geq 3$) to the (up to logarithmic corrections) optimal range n^{-1}.

Corollary 1.7. *For any $k \geq 3$ let C_k be the k-cycle and denote by T_k the subgraph count of C_k. If $p \geq n^{-1} \log^{(k+1)/(k-2)}(n)$, then*

$$\mathbb{P}\left(|T_k - \mathbb{E}\, T_k| \geq \varepsilon\, \mathbb{E}\, T_k\right) \leq 2 \exp\Big(-C_{k,\varepsilon} n^2 p^2 \log(1/p)\Big). \tag{1.18}$$

Note that this does not lead to precise large deviation results as given in [LZ17] or [CD16] (i.e. with a sharp constant C_ε), since the constant $\frac{1}{C}$ does not need to be sharp. On the other hand, this estimate is valid for the (almost) optimal region $p \geq \log(n)^4 n^{-1}$ (see the comment before Theorem 1.1 in [LZ17]), and it is a non-asymptotic result. It remains an open question whether one can derive optimal concentration inequalities for any graph H using this approach, since the norms $\|\cdot\|_{\mathcal{I}}$ are very hard to estimate precisely.

We compare our results to previously known estimates in subsection 1.2.

1.1.2. *Suprema of polynomial chaos.* Next let us treat the case of suprema of chaos-type functionals. The results will be based on a logarithmic Sobolev inequality with respect to our second difference operator $\mathfrak{d}$.

Let $X_1, \ldots, X_n$ be a sequence of independent $\{+1, -1\}$-valued random variables with $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$, $\mathcal{I}_{n,d}$ denote the family of subsets of $\{1, \ldots, n\}$ with d elements and $\mathcal{T}$ be a compact set of vectors in $\mathbb{R}^{\mathcal{I}_{n,d}}$ and write $X_I := \prod_{i \in I} X_i$.

In [BBLM05, Theorem 14, Corollary 4] the authors have proven that the random variable

$$f(X) := f_{\mathcal{T}}(X) := \sup_{t \in \mathcal{T}} \Big| \sum_{I \in \mathcal{I}_{n,d}} X_I t_I \Big| \tag{1.19}$$

satisfies for all $p \geq 2$

$$\|(f(X) - \mathbb{E}\, f(X))_+\|_p \leq \sum_{j=1}^{d} (4\kappa p)^{j/2}\, \mathbb{E}\, W_j, \tag{1.20}$$

where $\kappa \approx 1.27$ is a numerical constant (cf. Theorem 4.1) and

$$W_k := \sup_{t \in \mathcal{T}} \sup_{\substack{\alpha^1, \ldots, \alpha^k \in \mathbb{R}^n \\ \|\alpha^i\|_2 \leq 1}} \left| \sum_{\substack{i_1, \ldots, i_k \\ \text{distinct}}} \alpha^1_{i_1} \cdots \alpha^k_{i_k} \sum_{I \in \mathcal{I}_{n,d-k} : i_1, \ldots, i_k \notin I} X_I t_{I \cup \{i_1, \ldots, i_k\}} \right|. \tag{1.21}$$

For $k = d$, we use the convention $\mathcal{I}_{n,0} = \{\emptyset\}$ and $X_\emptyset := 1$. This led to one-sided deviations inequalities.

We strengthen the result in several ways. Firstly, we prove concentration inequalities rather than deviation inequalities for the upper tail. Secondly, the estimate will be valid for any Banach space. Thirdly, we remove the requirement of $X_1, \ldots, X_n$ being independent. We will first formulate the general theorem, and then show that (1.20) is a corollary.

Fix a Banach space $(\mathcal{B}, \|\cdot\|)$ with its dual space $(\mathcal{B}^*, \|\cdot\|_*)$, a compact subset $\mathcal{T} \subset \mathcal{B}^{\mathcal{I}_{n,d}}$ and let $\mathcal{B}^*_1$ be the 1-ball in $\mathcal{B}^*$ with respect to $\|\cdot\|_*$. Let $X_1, \ldots, X_n$ be real-valued random variables and define

$$f(X) := f_{\mathcal{T}}(X) := \sup_{t \in \mathcal{T}} \Big\| \sum_{I \in \mathcal{I}_{n,d}} X_I t_I \Big\|. \tag{1.22}$$

Since the theory applies to bounded functionals, we assume that $X_1, \ldots, X_n$ are $\mathbb{P}$-a.s. bounded by a common constant K, which due to the d-homogeneity of f we assume to be 1.

For any $k \in \{1, \ldots, n\}$ we define the generalization of (1.21) given by

$$W_k := \sup_{t \in \mathcal{T}} \sup_{v^* \in \mathcal{B}_1^*} \sup_{\substack{\alpha^1, \ldots, \alpha^k \in \mathbb{R}^n \\ \|\alpha^i\|_2 \le 1}} v^*\Big(\sum_{\substack{i_1, \ldots, i_k \\ \text{distinct}}} \alpha^1_{i_1} \cdots \alpha^k_{i_k} \sum_{\substack{I \in \mathcal{I}_{n,d-k} \\ i_1, \ldots, i_k \notin I}} X_I t_{I \cup \{i_1, \ldots, i_k\}} \Big) \tag{1.23}$$

$$= \sup_{t \in \mathcal{T}} \sup_{\substack{\alpha^1, \ldots, \alpha^k \in \mathbb{R}^n \\ \|\alpha^i\|_2 \le 1}} \Big\| \sum_{\substack{i_1, \ldots, i_k \\ \text{distinct}}} \alpha^1_{i_1} \cdots \alpha^k_{i_k} \sum_{\substack{I \in \mathcal{I}_{n,d-k} \\ i_1, \ldots, i_k \notin I}} X_I t_{I \cup \{i_1, \ldots, i_k\}} \Big\|.$$

Furthermore, in the case of independent random variables, we need the quantities

$$\widetilde{W}_k := \sup_{\substack{\alpha^1, \ldots, \alpha^k \in \mathbb{R}^n \\ \|\alpha^i\|_2 \le 1}} \sum_{\substack{i_1, \ldots, i_k \\ \text{distinct}}} \alpha^1_{i_1} \cdots \alpha^k_{i_k} \sup_{t \in \mathcal{T}} \sup_{v^* \in \mathcal{B}_1^*} v^*\Big(\sum_{\substack{I \in \mathcal{I}_{n,d-k} \\ i_1, \ldots, i_k \notin I}} X_I t_{I \cup \{i_1, \ldots, i_k\}} \Big) \tag{1.24}$$

$$= \sup_{\substack{\alpha^1, \ldots, \alpha^k \in \mathbb{R}^n \\ \|\alpha^i\|_2 \le 1}} \sum_{\substack{i_1, \ldots, i_k \\ \text{distinct}}} \alpha^1_{i_1} \cdots \alpha^k_{i_k} \sup_{t \in \mathcal{T}} \Big\| \sum_{\substack{I \in \mathcal{I}_{n,d-k} \\ i_1, \ldots, i_k \notin I}} X_I t_{I \cup \{i_1, \ldots, i_k\}} \Big\|.$$

Clearly $\widetilde{W}_k \ge W_k$ for all $k \in \{1, \ldots, d\}$. One can interpret the quantities W_k in the following way. If we denote by $f_t(x) = \sum_{I \in \mathcal{I}_{n,d}} x_I t_I$ the corresponding polynomial in n variables, and by $\nabla^k f_t(x)$ the k-linear tensor of all partial derivatives of order k, then $W_k = \sup_{t \in \mathcal{T}} \|\nabla^k f_t(X)\|_{\mathrm{op}}$.

Theorem 1.8. *Let $X = (X_1, \ldots, X_n)$ be a random vector in $\mathbb{R}^n$ with support in $[a,b]^n$ and satisfying a $\mathfrak{d}-\mathrm{LSI}(\sigma^2)$. For f as in* (1.22) *and all $p \ge 2$ we have*

$$\|f - \mathbb{E} f\|_p \le \sum_{j=1}^d (2\sigma^2 (b-a)^2 p)^{j/2} \, \mathbb{E} W_j, \tag{1.25}$$

and

$$\mathbb{P}\left(|f(X) - \mathbb{E} f(X)| \ge t\right) \le 2 \exp\left(-\frac{1}{C} \min_{k=1,\ldots,d} \left(\frac{t}{\mathbb{E} W_k} \right)^{2/k} \right). \tag{1.26}$$

On the other hand, if X consists of independent components and has support in $[a,b]^n$, then for f as in (1.19) *and all $p \ge 2$*

$$\|(f - \mathbb{E} f)_+\|_p \le \sum_{j=1}^d (2(b-a)^2 p)^{j/2} \, \mathbb{E} W_j, \tag{1.27}$$

$$\|f - \mathbb{E} f\|_p \le \sum_{j=1}^d (2(b-a)^2 p)^{j/2} \, \mathbb{E} \widetilde{W}_j. \tag{1.28}$$

Moreover, (1.27) *implies one-sided deviation inequalities and* (1.28) *implies tail estimates as in* (1.26).

The next corollary follows immediately, if one recalls that if $X_1, \ldots, X_n$ is a sequence of Rademacher random variables, then it satisfies a logarithmic Sobolev inequality with respect to $\mathfrak{d}$ with constant $\sigma^2 = 1$ (see e.g. [Gro75, Example 2.6], where it was proven for the first time, or [DS96, Example 3.1]).

Corollary 1.9. *Let $X_1, \ldots, X_n$ be independent Rademacher random variables and $f = f(X)$ as in* (1.19)*. We have*

$$\|f - \mathbb{E} f\|_p \le \sum_{j=1}^d (4p)^{j/2} \, \mathbb{E} W_j. \tag{1.29}$$

Consequently, there is a constant $C > 0$ depending on d, such that for all $t > 0$

$$\mathbb{P}\left(|f(X) - \mathbb{E}\, f(X)| \geq t\right) \leq 2\exp\left(-\frac{1}{C}\min_{k=1,\dots,d}\left(\frac{t}{\mathbb{E}\, W_j}\right)^{2/k}\right). \tag{1.30}$$

As a second corollary, Theorem 1.8 can be used to recover and strengthen a famous result by Talagrand [Tal96, Theorem 1.2] on concentration properties of quadratic forms in a Banach space. Considering the case $d = 2$, we can express the quantities $T_1 \coloneqq \mathbb{E}\, W_1$ and $T_2 \coloneqq \mathbb{E}\, W_2$ as

$$T_1 = \mathbb{E} \sup_{t \in \mathcal{T}} \sup_{v^* \in \mathcal{B}_1^*} \sup_{\substack{\alpha^{(1)} \in \mathbb{R} \\ \|\alpha^{(1)}\|_2 = 1}} v^*\left(\sum_{i=1}^n \alpha_i^{(1)} \sum_{j=1}^n X_j t_{ij}\right) = \mathbb{E} \sup_{t \in \mathcal{T}} \sup_{v^* \in \mathcal{B}_1^*} \left(\sum_{i=1}^n \left(\sum_{j=1}^n X_j v^*(t_{ij})\right)^2\right)^{1/2},$$

$$T_2 = \sup_{t \in \mathcal{T}} \sup_{v^* \in \mathcal{B}_1^*} \|(v^*(t_{ij}))_{i,j}\|_{\mathrm{op}}.$$

Corollary 1.10. *Assume that $X = (X_1, \dots, X_n)$ satisfies a $\mathfrak{d}-\mathrm{LSI}(\sigma^2)$ and is supported in $[a,b]^n$ and let $f_{\mathcal{T}}$ be as in 1.22 with $d = 2$. We have for some constant $C > 0$ and all $t \geq 0$*

$$\mathbb{P}\left(|f_{\mathcal{T}}(X) - \mathbb{E}\, f_{\mathcal{T}}(X)| \geq t\right) \leq 2\exp\left(-\frac{1}{C(b-a)^2\sigma^2}\min\left(\frac{t^2}{T_1^2}, \frac{t}{T_2}\right)\right). \tag{1.31}$$

Note that [Tal96, Theorem 1.2] can be retrieved by considering $\mathcal{T}$ consisting of a single element, since the Rademacher random variables satisfy a $\mathfrak{d}-\mathrm{LSI}(1)$ (although Talagrand considered fluctuations around the median instead of the mean). Moreover, Theorem 1.8 provides extensions to $d \geq 3$.

Theorem 1.8 can be applied to random variables other than Rademacher, for example to the spins in an Ising model on n sites in the Dobrushin uniqueness regime, given as follows. Let $\mathcal{Y} = \{-1, +1\}^n$ and $J = (J_{ij})_{1 \leq i,j \leq n}$ be a symmetric matrix with vanishing diagonal which satisfies $\|J\|_{1\to 1} = \max_{j=1,\dots,n} \sum_{i=1}^n |J_{ij}| = 1 - \rho$, $\rho > 0$. Define a probability measure $\mu = \mu_J$ on $\mathcal{Y}$

$$\mu(x) \coloneqq Z^{-1}\exp\left(\sum_{i,j} J_{ij} x_i x_j\right) \quad \text{where} \quad Z = \sum_{x \in \{-1,+1\}^n} \exp\left(\sum_{i,j} J_{ij} x_i x_j\right).$$

In [GSS18] it was shown that μ satisfies a $\mathfrak{d}-\mathrm{LSI}(\sigma^2)$ with a constant depending on ρ. Thus Theorem 1.8 can be applied to such Ising models. For example, the case $J_{ij} = \beta n^{-1}\delta_{i\neq j}$ (the Curie–Weiss model) satisfies the assumptions of Theorem 1.8 for $\beta < 1$.

1.2. **Related work.** The entropy method has been modified to the framework of so-called φ-entropies and developed in various works. The idea to replace the function $\varphi_0(x) \coloneqq x\log x$ in the definition of the entropy $\mathrm{Ent}_\mu^{\varphi_0}(f) = \mathbb{E}_\mu\, \varphi_0(f) - \varphi_0(\mathbb{E}_\mu\, f)$ by other functions φ is present in the works [BLM03; Cha04; BBLM05]. In the seminal work [BBLM05] the authors proved inequalities for φ-entropies for power functions $\varphi(x) = x^\alpha, \alpha \in (1,2]$, leading to moment inequalities for independent random variables.

The literature on concentration inequalities of polynomials in independent random variables is vast. It includes the works by Kim–Vu [KV00] for polynomials in $\{0,1\}$-valued random variables with positive coefficients, and Vu [Vu02] with generalization to different distributions and functions. Another approach was taken by Schudy–Sviridenko [SS12], who proved similar bounds for so-called moment bounded random variables. While we were preparing this work, in [AKPS18] the authors have proven similar concentration inequalities for tetrahedral polynomials in the Ising model in the high temperature regime, i.e. where the model satisfies a logarithmic Sobolev inequality with respect to $\mathfrak{d}$.

Additionally, note that in [MOO10] the authors have proven an invariance principle for multilinear polynomials in independent random variables, bounding the Kolmogorov distance of the distribution of $Q(X_1, \ldots, X_n)$ and $Q(G_1, \ldots, G_d)$, where $X_1, \ldots, X_n$ are independent random variables and $G_1, \ldots, G_n$ are Gaussian random variables.

To compare the results of Corollaries 1.6 and 1.7 with previous results, we have given a list of either concentration of measure inequalities or large deviation estimates for various subgraph counts (for the sparse regime, i.e. $p \to 0$).

subgraph	source	p-range ($p \geq \ldots$)	rate in the exponential
k-cycles	[Pan04]	$C \log(n) n^{-1}$	$n^2 p^2 (\log \log np)^{-1}$
any	[JOR04]	$n^{-1/\Delta}$	$n^2 p^\Delta$
triangle	[Cha12]	$C \log(n) n^{-1}$	$n^2 p^2 \log(1/p)$
triangle	[DK12a]	n^{-1}	$\min(n^2 p^2 \log(1/p), n^3 p^3)$
k-clique	[DK12b]	$n^{-2/(k-1)}$	$\min(n^2 p^{k-1} \log(1/p), n^k p^{\binom{k}{2}})$
triangle	[AW15]	$\log(n)^{-1/2} n^{-1/4}$	$n^2 p^2 \log(2/p)$
k-cycle	[AW15]	$\log(n)^{-1/2} n^{-(k-2)/(2k-2)}$	$n^2 p^2 \log(2/p)$
triangle	[LZ17]	$n^{-1/42} \log(n)$	$n^2 p^2 \log(1/p)$
any	[BGLZ17]	$n^{-\alpha_H}$	$n^2 p^\Delta \log(1/p)$
k-cycle	[CD18]	$\max(n^{(2-k)/k}, \log(n)^{\alpha_k} n^{-1/2})$	$n^2 p^2 \log(1/p)$

For a comparison of the upper tails prior to 2004, see also [JR02, Table 2]. The last three rows contain large deviation results, from which, under certain conditions, one can infer the existence of a constant C_ε such that

$$\lim_{n\to\infty} \frac{-\log \mathbb{P}(|T_G - \mathbb{E}\, T_G| \geq \varepsilon\, \mathbb{E}\, T_G)}{a_{n,p}} = C_\varepsilon, \tag{1.32}$$

where $a_{n,p}$ is the normalization given in the last column. In these results, the p-range was chosen such that we one can infer (1.32).

Given a graph G, let $\Delta = \Delta_G$ be the maximal degree and $m(G) = \max_{H \subset G : |V(H)| > 0} \frac{|E(H)|}{|V(H)|}$ the maximal density. It is known (see e.g. [JŁR00, Theorem 3.4]) that $n^{-1/m(G)}$ is an (asymptotic) threshold for the existence of a subgraph. Thus the natural boundary for the p-range should be $n^{-1/m(G)}$.

Moreover, the results by Janson, Oleszkiewicz and Ruciński show that the decay rate in the exponential depends on the rate of decay of p. Indeed, in [JOR04] the authors have identified the three regions $p \leq n^{-1/m(G)}, n^{-1/m(G)} \leq p \leq n^{-1/\Delta(G)}$ and $p \geq n^{-1/\Delta_G}$, in which the rates can differ, see e.g. [JOR04, Corollary 1.8, Corollary 1.9].

For all k-cycles C_k (including the triangular case $k = 3$) we have $m(C_k) = 1$, and for the k-cliques $m(K_k) = (k-1)/2$. In the regime $p \geq n^{-1/m(G)}$, the correct rate in the exponential should be $n^2 p^\Delta \log(1/p)$, see [JOR04, Theorem 1.5] and [BGLZ17, Corollary 1.6].

As for suprema of polynomial chaos, estimates of these functionals have been derived in various situations. A classical example (without the supremum), usually refered to as *Hanson–Wright inequalities*, has been studied in [HW71; Wri73] for subgaussian random variables and a real quadratic matrix by Hanson–Wright and Wright. A modern proof was given by Rudelson–Vershynin in [RV13]. It has been extended to non-independent random variables in [HKZ12] (for positive semi-definite matrices and X satisfying a uniform Subgaussianity property) and in [VW15] (with some logarithmic dependence on the dimension) and [Ada15] under the so-called convex concentration property. More recently, [CY18] have established the Hanson–Wright inequality for Hilbert spaces and [ALM18, Theorem 7] have proven it for Banach spaces and independent, subgaussian random variables.

Concentration properties for functionals as in (1.22) have been studied in the case of Rademacher random variables and in the real case in [BBLM05, Theorem 14] for all $d \geq 2$, and under certain technical assumptions in [Ada15].

1.3. **Outline.** In Section 2 we provide the general version of Theorem 1.1 and its proof, and we show how to deduce Corollary 1.2. Section 3 contains the definitions of the higher order difference operators that are required in Theorems 1.3 and 1.4, as well as a proposition providing a link between tail estimates in the formulations of Theorem 1.3 and L^p norm estimates. The concentration inequalities of Theorems 1.3 and 1.4 will be proven in Section 4.

Thereafter, in Section 5 we will prove the moment estimates for polynomials and its applications to the tails of the subgraph counts. Section 6 provides the proof of Theorem 1.8. The last Section 7 contains auxiliary results used frequently throughout this work.

2. Universal logarithmic Sobolev inequality: Proofs

As mentioned above, Theorem 1.1 admits a generalization to non-product measures. Indeed, a sufficient condition for the $\mathfrak{h}-\mathrm{LSI}(\sigma^2)$ property to hold is that the measure μ satisfies an approximate tensorization (AT) property.

To formulate the generalization, we will make use of the disintegration theorem on Polish spaces (see [DM78, Chapter III] and [AGS08, Theorem 5.3.1]): If μ is a measure on a product space $\otimes_{i=1}^n \mathcal{X}_i$, then for each $i \in \{1, \ldots, n\}$ we can decompose the measure using the marginal measure $\overline{\mu}_i$ (the measure on $\otimes_{j\neq i} \mathcal{X}_i$) and a conditional measure on $\mathcal{X}_i$, which we denote by $\mu(\cdot \mid \overline{x}_i)$. More precisely, for any Borel set $A \in \mathcal{B}(\otimes_{i=1}^n \mathcal{X}_i)$ we have $\mu(A) = \int_{\otimes_{j\neq i}\mathcal{X}_i} \int_{\mathcal{X}_i} \mathbb{1}_A(\overline{x}_i, x_i) d\mu(x_i \mid \overline{x}_i) d\overline{\mu}_i(\overline{x}_i)$.

Theorem 2.1. *Assume that $\mathcal{Y} = \otimes_{i=1}^n \mathcal{X}_i$ is a product of Polish spaces and $X = (X_1, \ldots, X_n)$ is a $\mathcal{Y}$-valued random vector with law μ. If μ satisfies an approximate tensorization property*

$$\mathrm{Ent}_\mu(f^2) \leq C \sum_{i=1}^n \int \mathrm{Ent}_{\mu(\cdot|\overline{x}_i)}(f^2(\overline{x}_i, \cdot)) d\overline{\mu}_i(\overline{x}_i), \tag{2.1}$$

then μ also satisfies an $\mathfrak{h}-\mathrm{LSI}(C)$.

The approximate tensorization property in Theorem 2.1 is interesting in its own right, but it is not yet well-studied. For finite spaces [Mar15] gives sufficient conditions for a measure μ to satisfy an approximate tensorization property. Similar results have been derived in [CMT15], which can be applied in discrete and continuous settings. For example, if one considers a measure of the form

$$\mu(x) = Z^{-1} \prod_{i=1}^n \mu_{0,i}(x_i) \exp\left(\sum_{i,j} J_{ij} w_{ij}(x_i, x_j) \right)$$

for some countable spaces Ω_i, $x_i \in \Omega_i$, measures $\mu_{0,i}$ on Ω_i and bounded functions w_{ij}, under certain technical conditions μ satisfies an approximate tensorization property. This does not require any functional inequality for $\mu_{0,i}$.

However, it requires a certain weak dependence assumption in general. For example, the push-forward of a random permutation to $\mathbb{N}^n$ cannot satisfy an approximate tensorization property, as the kernels $\mu(\cdot \mid \overline{x}_i)$ are Dirac measures. Since Theorem 2.1 relates the approximate tensorization property with concentration of measure results (although for bounded functions only), it is an interesting question to find necessary and sufficient conditions for the approximate tensorization property to hold.

Next, let us prove the universal logarithmic Sobolev inequality with respect to $\mathfrak{h}$ for random variables satisfying the tensorization property.

Proof of Theorem 2.1. First we consider the case $n = 1$. By homogeneity of both sides, we may and will assume that $\int f^2(X) d\,\mathbb{P} = 1$. Since f is bounded, we have $0 \le a \le |f(X)| \le b < \infty$ $\mathbb{P}$-a.s., where b is the essential supremum of $|f(X)|$ and a the essential infimum. Due to the constraints on the integral this leads to $a^2 \le 1 \le b^2$. (Actually the cases $b = 1$ or $a = 1$ are trivial, since then $f^2(X) = 1$ $\mathbb{P}$-a.s., but we will not make this distinction.)

Let $F(u) := \mathbb{P}(f^2(X) \ge u)$. In particular

$$F(u) = \begin{cases} 1 & u \le a^2, \\ 0 & u > b^2. \end{cases}$$

Using the partial integration formula (see e.g. [HS75, Theorem 21.67 and Remark 21.68]) in connection with [Bur07, Theorem 7.7.1] we have

$$\begin{aligned} \operatorname{Ent}(f^2(X)) &= \int_0^\infty u \log u d(-F(u)) = \int_0^{b^2} (\log u + 1) F(u) du \\ &= \int_0^{a^2} (\log u + 1) F(u) du + \int_{a^2}^{b^2} (\log u + 1) F(u) du \\ &= \int_0^{a^2} (\log u + 1) F(u) du + \int_{a^2}^{b^2} \log u F(u) du + (1 - a^2). \end{aligned}$$

The first integral can be calculated explicitly

$$\int_0^{a^2} (\log u + 1) F(u) du = u(\log u - 1) \mid_0^{a^2} = a^2 \log a^2,$$

and moreover we have due to $\log(u) \le \log(b^2)$ on $[a^2, b^2]$

$$\int_{a^2}^{b^2} \log u F(u) du \le \log(b^2)(1 - a^2).$$

Plugging in these two estimates yields

$$\operatorname{Ent}(f^2(X)) \le a^2 \log a^2 + (1 - a^2) + \log b^2 (1 - a^2) =: f(a, b).$$

Next, if we show that

$$f(a, b) \le 2(b - a)^2 \text{ on } G := \{(a, b) \in \mathbb{R}^2 : 0 \le a \le 1, 1 \le b < \infty\}, \tag{2.2}$$

we can further estimate

$$\operatorname{Ent}(f^2(X)) \le 2(b - a)^2 \le 2\,\mathbb{E}|\mathfrak{h} f|^2,$$

noting that $|\mathfrak{h} f|^2$ is a deterministic quantity in the case $n = 1$. To prove (2.2), define

$$g(a, b) := a^2 \log a^2 + (1 - a^2) + \log b^2 (1 - a^2) - 2(b - a)^2. \tag{2.3}$$

Now it is easy to see that

$$g(a, 1) = a^2 \log a^2 + (1 - a^2) - 2(1 - a)^2 \le 0,$$

since $\partial_a g(a, 1) \ge 0$ for $a \in [0, 1]$ and $g(1, 1) = 0$. Moreover

$$\partial_b g(a, b) = -\frac{2}{b} \left(b^2 - 1 + (a - b)^2 \right) \le 0,$$

so that g is decreasing on every strip $\{a_0\} \times [1, \infty)$, and thus $g(a, b) \le 0$ for all $a, b \in G$. This completes the proof for $n = 1$.

For arbitrary n, the proof is now easily completed. Assume that $f \in L^\infty(\mu)$, i.e. $\mu_i(\overline{x_i})$-a.s. we have $f(\overline{x_i}, \cdot) \in L^\infty(\mu(\cdot \mid \overline{x_i}))$. For these $\overline{x_i}$, by the $n = 1$ case we therefore obtain

$$\operatorname{Ent}_{\mu(\cdot|\overline{x}_i)}(f^2(\overline{x_i}, \cdot)) \le 2 \sup_{y_i', y_i''} |f(\overline{x}_i, y_i') - f(\overline{x}_i, y_i'')|^2. \tag{2.4}$$

Plugging this into the assumption leads to

$$\operatorname{Ent}_\mu(f^2) \le 2C \int \sum_{i=1}^n \sup_{y_i', y_i''} |f(\overline{x}_i, y_i') - f(\overline{x}_i, y_i'')|^2 d\overline{\mu}_i(\overline{x}_i) = 2C \int |\mathfrak{h} f|^2 d\mu. \tag{2.5}$$

□

Proof of Theorem 1.1. The fact that independent random variables satisfy the tensorization property (i.e. AT(1)) is by now classical and can be found in [Led01, Proposition 5.6], [BBLM05, Theorem 4.10] or [vH16, Theorem 3.14]. In the case of independent random variables, the assumption that $\mathcal{Y}$ is a product of Polish spaces can be dropped by simply defining $\mu(\cdot \mid \overline{x}_i) \coloneqq \mu_i = \mathbb{P} \circ X_i$. □

Proof of Corollary 1.2. From the $\mathfrak{h}-$LSI(1) we can deduce for all $p \in [2, \infty)$ (see Proposition 4.4)

$$\|f - \mathbb{E} f\|_p \le (4p)^{1/2} \|\mathfrak{h} f\|_p,$$

and using the trivial estimate $\|f\|_p \le \|f\|_q$ for $p \le q$ this leads to the inequality

$$\|f - \mathbb{E} f\|_p \le (8p)^{1/2} \|\mathfrak{h} f\|_p$$

for all $p \in [1, \infty)$. Since $\mathfrak{h}_i f \le c_i$, we have $|\mathfrak{h} f| \le |c|$. Combining these estimates gives

$$\sup_{p \ge 1} \frac{\|f - \mathbb{E} f\|_p}{p^{1/2}} \le (8|c|_2^2)^{1/2}.$$

However, it is known (see for example [Bob10, Section 8]) that the Orlicz-norm $\|\cdot\|_\psi$ associated to $\psi(x) \coloneqq \exp(x^2) - 1$ satisfies

$$\|f - \mathbb{E}\|_\psi \le (4e)^{1/2} \sup_{p \ge 1} \frac{\|f - \mathbb{E} f\|_p}{p^{1/2}} \le (32e|c|^2)^{1/2},$$

which leads to

$$\mathbb{E} \exp\left(\frac{|f - \mathbb{E} f|^2}{32e|c|^2} \right) \le 2.$$

Now Markov's inequality yields (1.9). □

3. Higher order difference operators and L^p norm estimates

In this section, we will first define the d-tensors $\mathfrak{h}^{(d)} f$ for $d \ge 2$ that were used in Theorem 1.3. The basis of the *higher order differences* will be the difference operator $\mathfrak{h}$. Secondly, since the tail estimates will depend on general L^p norm inequalities of the form

$$\|f(X) - \mathbb{E} f(X)\|_p \le \sum_{k=1}^d (C_k p)^{k/2} \quad \text{for all} \quad p \ge 2,$$

we will give a proposition that translates such inequalities into tail estimates.

We define the k-tensor $\mathfrak{h}^{(d)} f$ by specifying it on its "coordinates". Given distinct indices $i_1, \ldots, i_d$ we let

$$\begin{aligned}\mathfrak{h}_{i_1\ldots i_d} f(X) &= \Big\| \prod_{s=1}^{d} (\mathrm{Id} - T_{i_s}) f(X) \Big\|_{i_1,\ldots,i_d,\infty} \\ &= \Big\| f(X) + \sum_{k=1}^{d} (-1)^k \sum_{1 \le s_1 < \ldots < s_k \le d} T_{i_{s_1}\ldots i_{s_k}} f(X) \Big\|_{i_1,\ldots,i_d,\infty},\end{aligned} \tag{3.1}$$

where $T_{i_1\ldots i_d} = T_{i_1} \circ \ldots \circ T_{i_d}$ exchanges the random variables $X_{i_1}, \ldots, X_{i_d}$, and $\|\cdot\|_{i_1,\ldots,i_d,\infty}$ denotes the L^∞-norm with respect to $X_{i_1}, \ldots, X_{i_d}$ and $X'_{i_1}, \ldots, X'_{i_d}$. For instance, for $i \neq j$,

$$\mathfrak{h}_{ij} f = \| f - T_i f - T_j f + T_{ij} f \|_{i,j,\infty}.$$

Using the definition (3.1), we define tensors of d-th order differences as follows:

$$(\mathfrak{h}^{(d)} f(X))_{i_1\ldots i_d} = \begin{cases} \mathfrak{h}_{i_1\ldots i_d} f(X), & \text{if } i_1, \ldots, i_d \text{ are distinct,} \\ 0, & \text{else.} \end{cases} \tag{3.2}$$

Whenever no confusion is possible, we omit writing the random vector X, i.e. we freely write f instead of $f(X)$ and $\mathfrak{h}^{(d)} f$ instead of $\mathfrak{h}^{(d)} f(X)$. Note that $|\mathfrak{h}^{(d)} f|$ are again difference operators.

We will need another, closely related difference operator. For $i \in \{1, \ldots, n\}$ introduce

$$\mathfrak{h}_i^+ f(X) = \|(f(X) - T_i f(X))_+\|_{X'_i,\infty}, \qquad \mathfrak{h}^+ f = (\mathfrak{h}_1^+ f, \ldots, \mathfrak{h}_n^+ f), \tag{3.3}$$

$$\mathfrak{h}_i^- f(X) = \|(f(X) - T_i f(X))_-\|_{X'_i,\infty}, \qquad \mathfrak{h}^- f = (\mathfrak{h}_1^- f, \ldots, \mathfrak{h}_n^- f), \tag{3.4}$$

where $\|f\|_{X'_i,\infty}$ shall denote the L^∞ norm with respect to X'_i.

Lastly, we ignore any measurability issues that may arise. They may be dealt with by restricting oneself to appropriately well-behaved spaces (such as finite or more generally Polish spaces). Alternatively, we could define the difference operators in terms of a majorizing measurable function. Thus we assume that $\mathfrak{h}_{i_1\ldots i_d} f(X)$ (and $\mathfrak{h}_i^+ f(X)$) are measurable for any $d \in \mathbb{N}$ and $i_1, \ldots, i_d$.

The vector $\mathfrak{h}^+ f$ and the tensors $\mathfrak{h}^{(d)} f$ can be regarded as elements of the Euclidean spaces $\mathbb{R}^n$ and $\mathbb{R}^{n^d}$ respectively, and we let $|\mathfrak{h}^{(d)} f|_{\mathrm{HS}}$ be the Euclidean norm of $\mathfrak{h}^{(d)} f$ regarded as an element of $\mathbb{R}^{n^d}$, and $|\mathfrak{h}^+ f|_{\mathrm{HS}}$ as the Euclidean norm of the vector $\mathfrak{h}^+ f$. Furthermore, for any $d \in \mathbb{N}$ and any d-tensor $A = (A_{i_1\ldots i_d})_{1 \le i_1,\ldots,i_d \le n}$ we define the operator norm

$$|A|_{\mathrm{op}} := \sup_{\substack{v^1,\ldots,v^d \\ \|v^j\|_2 \le 1}} \langle v^1 \cdots v^d, A \rangle = \sup_{\substack{v^1,\ldots,v^d \\ \|v^j\|_2 \le 1}} \sum_{i_1,\ldots,i_d} v^1_{i_1} \cdots v^d_{i_d} A_{i_1\ldots i_d}, \tag{3.5}$$

using the vector product

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

It is easily seen that for a 1-tensor (i.e. a vector) we have $|A|_{\mathrm{op}} = |A|_{\mathrm{HS}}$ and for any $d \ge 2$ and any d-tensor A

$$|A|_{\mathrm{op}} \le |A|_{\mathrm{HS}}.$$

Moreover, note that the supremum is attained, and if A is a nonnegative tensor (i.e. $A_{i_1\ldots i_d} \ge 0$ for all $i_1, \ldots, i_d$), the maximizing vectors $\widetilde{v}^1, \ldots, \widetilde{v}^d$ can be chosen to have all positive entries. Indeed, since $\widetilde{v}^1_{i_1} \cdots \widetilde{v}^d_{i_d} \le |\widetilde{v}^1_{i_1} \cdots \widetilde{v}^d_{i_d}|$, we can define $|\widetilde{v}|^j$ by taking the absolute value elementwise.

In the proofs of Theorems 1.3 we will establish a growth rate on the L^p norms of $f(X) - \mathbb{E} f(X)$. The following proposition establishes the connection between such norm estimates and the concentration inequalities (1.10) and (1.11). It was proven in [Ada06, Theorem 7] and [AW15, Theorem 3.3]. We state it in the form given in [SS18, Proof of Theorem 1.1].

Proposition 3.1. *Assume that a random vector X satisfies for any $p \geq 2$*

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

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

$$\mathbb{P}(|f(X) - \mathbb{E} f(X)| \geq t) \leq e^2 \exp\left(-\frac{1}{(Le)^{2/r}} \min_{k=1,\ldots,d} \left\{ \frac{t^{2/k}}{C_k} \right\} \right). \tag{3.6}$$

4. Concentration inequalities under logarithmic Sobolev inequalities: Proofs

To prove Theorem 1.3 we recall the following L^p norm inequalities. These results (with a different choice of normalization for $\mathfrak{h}^{\pm}$ leading to slightly different constants) can be found in [BGS18, Lemma 2.2, Theorem 2.3] (building upon the earlier results in [BBLM05]), and we skip the proofs.

Theorem 4.1. *If $X_1, \ldots, X_n$ are independent random variables and $f = f(X)$, with the constant $\kappa = \frac{\sqrt{e}}{2(\sqrt{e}-1)}$, for any real $p \geq 2$,*

$$\|(f - \mathbb{E}f)_+\|_p \leq \sqrt{2\kappa p}\, \|\mathfrak{h}^+ f\|_p, \tag{4.1}$$

$$\|(f - \mathbb{E}f)_-\|_p \leq \sqrt{2\kappa p}\, \|\mathfrak{h}^- f\|_p. \tag{4.2}$$

Lemma 4.2. *For any $d \geq 2$,*

$$|\mathfrak{h}|\mathfrak{h}^{(d-1)} f(X)|_{\mathrm{HS}}| \leq |\mathfrak{h}^{(d)} f(X)|_{\mathrm{HS}}. \tag{4.3}$$

Moreover, we need the following auxiliary statements. The proofs are postponed to Section 7.

Lemma 4.3. *For any $d \geq 2$*

$$|\mathfrak{h}^+|\mathfrak{h}^{(d-1)} f(X)|_{\mathrm{op}}| \leq |\mathfrak{h}^{(d)} f(X)|_{\mathrm{op}}.$$

Proposition 4.4. *Let μ be a measure on a product of Polish spaces satisfying $\mathfrak{d}-\mathrm{LSI}(\sigma^2)$. Then, for any $f \in L^\infty(\mu)$ and any $p \geq 2$, we have*

$$\|f\|_p^2 - \|f\|_2^2 \leq 2\sigma^2(p-2)\|\mathfrak{d} f\|_p^2 \leq \sigma^2(p-2)\|\mathfrak{h} f\|_p^2 \tag{4.4}$$

as well as

$$\|f\|_p^2 - \|f\|_2^2 \leq \sigma^2(p-2)\|\mathfrak{h}^+|f|\|_p^2. \tag{4.5}$$

If μ satisfies an $\mathfrak{h}-\mathrm{LSI}(\sigma^2)$, we have

$$\|f\|_p^2 - \|f\|_2^2 \leq 4\sigma^2(p-2)\|\mathfrak{h} f\|_p^2. \tag{4.6}$$

Proof of Theorem 1.3. (1): From (4.6) and the Poincaré inequality we obtain

$$\|f - \mathbb{E} f\|_p \leq (4\sigma^2 p)^{1/2} \|\mathfrak{h} f\|_{\mathrm{HS},p} \leq (4\sigma^2 p)^{1/2} \left(\|\mathfrak{h} f\|_{\mathrm{HS},1} + \||\mathfrak{h} f|_{\mathrm{HS}} - \mathbb{E}|\mathfrak{h} f|_{\mathrm{HS}}\|_p \right). \tag{4.7}$$

Now an iteration and the pointwise inequality from Lemma 4.2 yields the result.

(2): Since $X_1, \ldots, X_n$ are independent, the vector X satisfies an $\mathfrak{h}-\mathrm{LSI}(1)$, and from the proof of the first part we have

$$\|f - \mathbb{E} f\|_p \leq (2p)^{1/2} \|\mathfrak{h} f\|_p \leq (2p)^{1/2}\, \mathbb{E}|\mathfrak{h} f| + (2p)^{1/2} \|(|\mathfrak{h} f| - \mathbb{E}|\mathfrak{h} f|)_+\|_p.$$

The second summand on the right hand side can now be estimated using Theorem 4.1, which in combination with Lemma 4.3 gives

$$\|(|\mathfrak{h}f| - \mathbb{E}|\mathfrak{h}f|)_+\|_p \le \sqrt{2\kappa p}\|\mathfrak{h}^+|\mathfrak{h}f|\|_p \le \sqrt{2\kappa p}\||\mathfrak{h}^{(2)}f|\|_{\mathrm{op},p}.$$

This can be easily iterated to yield

$$\|f - \mathbb{E}\,f\|_p \le \sum_{j=1}^{d-1}(\sigma^2 p)^{j/2}\,\mathbb{E}|\mathfrak{h}^{(j)}f|_{\mathrm{op}} + (\sigma^2 p)^{d/2}\|\mathfrak{h}^{(d)}f\|_{\mathrm{op},p}, \tag{4.8}$$

where $\sigma^2 = 4\kappa \le 6$. □

Proof of Theorem 1.4. The proof is very similar to the proof of Theorem 1.3. We will state it in the more general assumption of a $\mathfrak{h}^+$−LSI(σ^2), which is implied by a $\mathfrak{d}$−LSI(σ^2).

The important difference is that a $\mathfrak{h}^+$−LSI(σ^2) implies the two inequalities

$$\|f - \mathbb{E}\,f\|_p \le (\sigma^2 p)^{1/2}\|\mathfrak{h}^+|f|\|_p \tag{4.9}$$

$$\|f - \mathbb{E}\,f\|_p \le (\sigma^2 p)^{1/2}\|\mathfrak{h}f\|_p, \tag{4.10}$$

which follow by a combination of Proposition 4.4 and the Poincaré inequality. Now the first step is to apply equation (4.9) and the fact that for any positive random variable W we have

$$\|W\|_p \le \mathbb{E}\,W + \|(W - \mathbb{E}\,W)_+\|_p,$$

which leads to

$$\|f - \mathbb{E}\,f\|_p \le (\sigma^2 p)^{1/2}\|\mathfrak{h}f\|_{\mathrm{HS},1} + (\sigma^2 p)^{1/2}\|(|\mathfrak{h}f|_{\mathrm{HS}} - \mathbb{E}|\mathfrak{h}f|_{\mathrm{HS}})_+\|_p.$$

Now the proof is identical to the one of the first part, apart from replacing Lemma 4.2 by Lemma 4.3, noting that $|v|_2 = |v|_{\mathrm{HS}} = |v|_{\mathrm{op}}$ for all $v \in \mathbb{R}^n$. □

5. Concentration of polynomials: Proofs

Using the simple observation

$$\mathfrak{h}_i f_d(X)^2 \le \sup_{x_i', x_i'' \in [a,b]} |f_d(\overline{X}_i, x_i') - f_d(\overline{X}_i, x_i'')|^2 \le (b-a)^2 \partial_i f_d(X)^2, \tag{5.1}$$

where ∂_i is the i-th partial derivative of f, from (4.7) we obtain the inequality

$$\|f_d(X) - \mathbb{E}\,f_d(X)\|_p \le ((b-a)^2\sigma^2 p)^{1/2}\|\nabla f_d(X)\|_p, \tag{5.2}$$

where ∇ is the "ordinary" gradient. Thus, for polynomials we obtain a family of Sobolev-type inequalities. This leads to the following theorem.

The next result will crucially depend on [AW15, Theorem 1.2], and it requires some additional notation. Let P_d denote the set of all partitions of $\{1,\dots,d\}$. Any partition $\mathcal{I} = \{I_1,\dots,I_k\} \in P_d$ induces a partition of the space of d-tensors as follows. Identify the space of all d-tensors with $\mathbb{R}^{n^d}$ and decompose

$$\mathbb{R}^{n^d} \simeq \bigotimes_{j=1}^{k} \mathbb{R}^{n^{I_j}} \simeq \bigotimes_{j=1}^{k}\bigotimes_{i\in I_j} \mathbb{R}^n\,.$$

On this space, define a norm

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

Given $x = x^{(1)} \otimes \ldots \otimes x^{(k)}$, the identification with a d-tensor is given via $x_{(J_1,\dots,J_d)} = \prod_{l=1}^k x^{(l)}_{J_{I_l}}$. For example, for $d = 4$ and $\mathcal{I} = \{\{1,4\},\{2,3\}\}$ we have two matrices x, y and $x_{(J_1,J_2,J_3,J_4)} = x_{J_1J_4}y_{J_2J_3}$.

With this identification, any d-tensor A can be trivially identified with a linear functional on $\mathbb{R}^{n^d}$ via the standard scalar product, i.e.

$$Ax = A\left(x^{(1)} \otimes \ldots \otimes x^{(k)}\right) = \langle A, x^{(1)} \otimes \ldots \otimes x^{(k)} \rangle = \sum_{J \in \{1,\ldots,n\}^d} A_J \prod_{l=1}^d x^{(l)}_{J_{I_l}},$$

and we denote $\|A\|_{\mathcal{I}}$ the operator norm with respect to $\|\cdot\|_{\mathcal{I}}$:

$$\|A\|_{\mathcal{I}} = \sup_{\|x\|_{\mathcal{I}} \le 1} |Ax|. \tag{5.3}$$

These definitions agree with [AW15] and [AKPS18]. Moreover, we denote by $\nabla^k f$ the k-tensor comprising all partial derivatives of k-th order.

Proof of Theorem 1.5. We will give a sketch of the proof only and refer to [AKPS18, Proof of Theorem 2.2] for details. Recall that by (5.2) we have the inequality

$$\|f_d(X) - \mathbb{E}\, f_d(X)\|_p \le (4\sigma^2 p)^{1/2} \|\nabla f(X)\|_p.$$

Using the arguments in [AKPS18, Proof of Theorem 2.2], this leads to

$$\|f_d(X) - \mathbb{E}\, f_d(X)\|_p \le \sum_{k=1}^d (4\sigma^2 M^2)^{k/2} \|\langle \mathbb{E}_X \nabla^k f_d(X), G_1 \otimes \ldots \otimes G_k \rangle\|_p,$$

where M is an absolute constant and G_i is a sequence of independent standard Gaussian random variables, independent of X.

Furthermore, a result by Latała [Lat06] yields

$$\begin{aligned}\|f_d(X) - \mathbb{E}\, f_d(X)\|_p &\le \sum_{k=1}^d \sum_{\mathcal{I} \in P_k} (4\sigma^2 M^2 p)^{k/2} \|\mathbb{E}\, \nabla^k f(X)\|_{\mathcal{I}} \\ &\le \sum_{k=1}^d \sum_{\mathcal{I} \in P_k} (C\sigma^2 p)^{k/2} \|\mathbb{E}\, \nabla^k f(X)\|_{\mathcal{I}}.\end{aligned}$$

The rest now follows as in [AW15, Theorem 1.4] with L replaced by $C\sigma$. □

Proof of Corollary 1.6. The proof relies on [AW15, Proposition 5.5]. Define $\sigma^2 := p\log(1/p)$ for $p \in [0,1]$ and $\varphi(n,p) = n^2 p^2 \log(1/p)$. Then, by Theorem 1.5, we have for every $q \ge 2$

$$\|T_3(X) - \mathbb{E}\, T_3(X)\|_q \le \sum_{d=1}^3 (C\sigma)^d \sum_{\mathcal{I} \in \mathcal{P}_d} q^{|\mathcal{I}|/2} \|\mathbb{E}\, \nabla^d T_3(X)\|_{\mathcal{I}}. \tag{5.4}$$

The quantities $\|\mathbb{E}\, \nabla^d T_3(X)\|_{\mathcal{I}}$ have been estimated in [AW15, p. 571] in the following way:

$$\|\mathbb{E}\, \nabla f(X)\|_{\{1\}} \le n^2 p^2, \qquad \|\mathbb{E}\, \nabla^2 f(X)\|_{\{1\}\{2\}} \le 2pn$$

$$\|\mathbb{E}\, \nabla^3 f(X)\|_{\{1,2,3\}} \le n^{3/2} \qquad \|\mathbb{E}\, \nabla^3 f(X)\|_{\{1,2\}\{3\}} \le 2n^{1/2} \qquad \|\mathbb{E}\, \nabla^3 f(X)\|_{\{1\}\{2\}\{3\}} \le 2^{3/2}.$$

This leads to

$$\mathbb{P}\left(|T_3 - \mathbb{E}\, T_3| \ge \varepsilon\, \mathbb{E}\, T_3\right) \le 2\exp\left(-\frac{\varphi(n,p)}{C} \min\left(\frac{\varepsilon^2}{K_1(n,p)}, \frac{\varepsilon}{K_2(n,p)}, \varepsilon^{2/3}\right)\right), \tag{5.5}$$

where

$$\begin{aligned} K_1(n,p) &= (\sigma^6/(np^4) + \sigma^4/(p^2 n) + \sigma^2)\log(1/p) \\ K_2(n,p) &= (\sigma^3/(n^{1/2}p) + \sigma^2)\log(1/p). \end{aligned}$$

Thus to show the inequality, it remains to prove that $K_1(n,p), K_2(n,p)$ can be bounded from above by an absolute constant, i.e.

$$\frac{\log^4(1/p)}{np} + \frac{\log^3(1/p)}{n} + \log(1/p)^2 p \le c$$

$$\frac{\log(1/p)^{5/2}p}{n^{1/2}p^{1/2}} + p\log(1/p)^3 \le c.$$

As long as $p \ge \frac{\log(n)^4}{n}$, the first summand of K_1 can be bounded by 1, and the other two tend to zero, as long as p is bounded away from 1. Moreover, the summands in K_2 can be bounded as long as $np \to \infty$. □

Proof of Corollary 1.7. Again we use the result from [AW15]. For any $t > 0$ and any $k \ge 3$ we have

$$\mathbb{P}\left(|T_k - \mathbb{E}\,T_k| \ge \varepsilon\,\mathbb{E}\,T_k)\right) \le 2\exp\left(-\frac{1}{C_k}\left(\frac{\varepsilon^2 n^{2k}p^{2k}}{\sigma^{2k}n^k} \wedge \min_{\substack{1\le l\le d\le k\\ d<k \text{ or } l>1}}\left(\frac{\varepsilon n^k p^k}{\sigma^d p^{k-d} n^{k-d/2-l/2}}\right)^{2/l}\right)\right)$$
$$=: 2\exp\left(-\frac{1}{C_k}\left(Q(\varepsilon,n,p) \wedge \min_{\substack{1\le l\le d\le k\\ d<k \text{ or } l>1}} H_{l,d}(\varepsilon,n,p)^{2/l}\right)\right).$$

It remains to show that in the range $p \ge n^{-1}\log^{(k+1)/(k-2)}(n)$ the quadratic term Q as well as all the other terms $H_{l,d}(\varepsilon,n,p)$ can be bounded from below by $C_{\varepsilon,k,l}n^2p^2\log(1/p)$. The quadratic term is easy to handle, as by the assumption we easily obtain

$$Q(\varepsilon,n,p) = n^2p^2\log(1/p)\varepsilon^2\frac{n^{k-2}p^{k-2}}{\log^{k+1}(1/p)} \ge n^2p^2\log(1/p)\varepsilon^2.$$

Now fix any pair (l,d) with $1 \le l \le d \le k$, $d < k$ or $l > 1$. Again we write

$$H_{l,d}(\varepsilon,n,p) = \varepsilon\frac{n^{(d+l)/2}p^{d/2}}{\log^{d/2}(1/p)} = \varepsilon n^l p^l \log(1/p)^{l/2}\frac{n^{l/2}(np)^{d/2-l}}{\log(1/p)^{(d+l)/2}}$$
$$\ge \varepsilon n^l p^l \log(1/p)^{l/2} C_{d,l},$$

where the inequality follows from $np \to \infty$ and $n^\alpha/(\log^\beta(1/p)) \to \infty$ for all $\alpha,\beta > 0$.

Plugging in these estimates, we obtain

$$\mathbb{P}\left(|T_k - \mathbb{E}\,T_k| \ge \varepsilon\,\mathbb{E}\,T_k)\right) \le 2\exp\left(-C_{k,\varepsilon}n^2p^2\log(1/p)\right). \tag{5.6}$$

□

6. Suprema of chaos: Proofs

Proof of Theorem 1.8. Let us first consider the case that X satisfies a $\mathfrak{d}-\mathrm{LSI}(\sigma^2)$. From (4.4) (and the Poincaré inequality to remove the L^2 norm) we obtain

$$\|f - \mathbb{E}\,f\|_p \le (2\sigma^2 p)^{1/2}\|\mathfrak{d}f\|_p.$$

We shall make use of the pointwise inequality

$$|\mathfrak{d}f| \le (b-a)W_1. \tag{6.1}$$

To prove (6.1), first note that

$$|\mathfrak{d}f(X)|^2 = \sum_{i=1}^n \iint (f(\overline{X}_i,x_i') - f(\overline{X}_i,x_i''))_+^2 d\mu(x_i' \mid \overline{x}_i)d\mu(x_i'' \mid \overline{x}_i) \tag{6.2}$$
$$\le \int |\mathfrak{h}^+ f(\overline{X}_i,x_i')|^2 d\mu(x_i' \mid \overline{x}_i),$$

so that it suffices to prove $|\mathfrak{h}^+ f(X)| \le (b-a)W_1$. To see this, let $(\widetilde{t},\widetilde{v}^*)$ be the tuple satisfying $\sup_{t\in\mathcal{T}}\sup_{v^*\in\mathcal{B}_1^*} v^*(\sum_{I\in\mathcal{I}_{n,d}} X_I t_I) = \widetilde{v}^*(\sum_{I\in\mathcal{I}_{n,d}} X_I \widetilde{t}_I)$. We have

$$|\mathfrak{h}^+ f(X)|^2 = \sum_{i=1}^n \sup_{x_i'}\Big(\sup_{t\in\mathcal{T}}\sup_{v^*\in\mathcal{B}_1^*} v^*\Big(\sum_{I\in\mathcal{I}_{n,d}} X_I t_I\Big) - \sup_{t\in\mathcal{T}}\sup_{v^*\in\mathcal{B}_1^*} v^*\Big(\sum_{I\in\mathcal{I}_{n,d}} (\overline{X}_i,x_i')_I t_I\Big)\Big)_+^2$$

$$
\begin{aligned}
&\leq \sum_{i=1}^n \sup_{x_i'} \Big((X_i - x_i') \sum_{\substack{I \in \mathcal{I}_{n,d-1} \\ i \notin I}} \widetilde{v}^*(X_I \widetilde{t}_{I \cup \{i\}}) \Big)^2 \\
&\leq (b-a)^2 \sum_{i=1}^n \Big(\widetilde{v}^* \Big(\sum_{\substack{I \in \mathcal{I}_{n,d-1} \\ i \notin I}} X_I \widetilde{t}_{I \cup \{i\}} \Big) \Big)^2 \\
&= (b-a)^2 \sup_{\alpha^{(1)} : \|\alpha^{(1)}\| \leq 1} \widetilde{v}^* \Big(\sum_{i=1}^n \alpha_i^{(1)} \sum_{I \in \mathcal{I}_{n,d-1} : i \notin I} X_I \widetilde{t}_{I \cup \{i\}} \Big)^2 \\
&\leq (b-a)^2 \sup_{t \in \mathcal{T}} \sup_{v^* \in \mathcal{B}_1^*} \sup_{\alpha^{(1)} : \|\alpha^{(1)}\| \leq 1} v^* \Big(\sum_{i=1}^n \alpha_i^{(1)} \sum_{I \in \mathcal{I}_{n,d} : i \notin I} X_I t_{I \cup \{i\}} \Big) \\
&= (b-a)^2 W_1^2,
\end{aligned}
$$

which proves (6.1). Consequently,

$$
\|f - \mathbb{E} f\|_p \leq (2\sigma^2 (b-a)^2 p)^{1/2} \|W_1\|_p \leq (2\sigma^2 (b-a)^2 p)^{1/2} \left(\mathbb{E} W_1 + \|W_1 - \mathbb{E} W_1\|_p \right).
$$

As in [BBLM05], this can now be iterated, i.e. we have for any $k \in \{1, \ldots, d-1\}$ $|\mathfrak{h}^+ W_k| \leq (b-a) W_{k+1}$. Here we may argue as above, where the only difference is to choose $(\widetilde{t}, \widetilde{v}^*)$ and $\widetilde{\alpha}^{(1)}, \ldots, \widetilde{\alpha}^{(k)}$ which maximize W_k. Finally we obtain

$$
\|f - \mathbb{E} f\|_p \leq \sum_{j=1}^d (2\sigma^2 (b-a)^2 p)^{j/2} \, \mathbb{E} W_j, \tag{6.3}
$$

where we have used that W_d is constant, i.e. $\|W_d\|_p = \mathbb{E} W_d = W_d$. This proves (1.25).

The same arguments are also valid without a $\mathfrak{d}-\mathrm{LSI}(\sigma^2)$ property, if one considers $\|(f - \mathbb{E} f)_+\|_p$ and applies Theorem 4.1 instead. This leads to equation (1.27).

Lastly, to prove (1.28), let us first consider why we cannot argue as in the first two parts. Note that the argument heavily relies on the positive part of the difference operator $\mathfrak{h}^+$, which allows us to choose the maximizers independently of $i \in \{1, \ldots, n\}$. This is no longer possible in the case of independent random variables. Here, by a combination of Theorem 1.1, Proposition 4.4 and Theorem 4.1 we obtain

$$
\|f - \mathbb{E} f\|_p \leq (4p)^{1/2} \|\mathfrak{h} f\|_{\mathrm{HS},p} \tag{6.4}
$$

$$
\|(f - \mathbb{E} f)_+\|_p \leq (4p)^{1/2} \|\mathfrak{h}^+ f\|_{\mathrm{HS},p}. \tag{6.5}
$$

Thus this argument fails if we try to use (6.4). However, we can rewrite

$$
\mathfrak{h}_i f(x) = \sup_{x_i', x_i''} |f(\overline{x}_i, x_i') - f(\overline{x}_i, x_i'')| = \sup_{x_i', x_i''} (f(\overline{x}_i, x_i') - f(\overline{x}_i, x_i''))_+ = \sup_{x_i'} \mathfrak{h}_i^+ f(\overline{x}_i, x_i'),
$$

where the sup is to be understood as an $L^\infty(\mu)$ norm. As a consequence, we have for each fixed $i \in \{1, \ldots, n\}$ (again choosing $\widetilde{t}$ by maximizing the first summand in the brackets)

$$
\begin{aligned}
\mathfrak{h}_i f(x)^2 &= \sup_{x_i'} \sup_{x_i''} \Bigg(\sup_{t \in \mathcal{T}} \Big\| \sum_{I \in \mathcal{I}_{n,d}} (\overline{X}_i, x_i')_I t_I \Big\| - \sup_{t \in \mathcal{T}} \Big\| \sum_{I \in \mathcal{I}_{n,d}} (\overline{X}_i, x_i'')_I t_I \Big\| \Bigg)_+^2 \\
&\leq \sup_{x_i'} \sup_{x_i''} \Big\| (x_i' - x_i'') \sum_{I \in \mathcal{I}_{n,d-1} : i \notin I} X_I \widetilde{t}_{I \cup \{i\}} \Big\|^2 \\
&\leq \sup_{x_i', x_i''} |x_i' - x_i''|^2 \sup_{t \in \mathcal{T}} \Big\| \sum_{I \in \mathcal{I}_{n,d-1} : i \notin I} X_I t_{I \cup \{i\}} \Big\|^2
\end{aligned}
$$

$$\leq (b-a)^2 \sup_{t \in \mathcal{T}} \Big\| \sum_{I \in \mathcal{I}_{n,d-1} : i \notin I} X_I t_{I \cup \{i\}} \Big\|^2,$$

which implies

$$|\mathfrak{h} f|^2(x) \leq (b-a)^2 \sup_{\substack{\alpha^1 \in \mathbb{R}^n \\ \|\alpha^1\|_2 \leq 1}} \sum_{i=1}^n \alpha_i^1 \sup_{t \in \mathcal{T}} \Big\| \sum_{\substack{I \in \mathcal{I}_{n,d-1} \\ i \notin I}} X_I t_{I \cup \{i\}} \Big\| = (b-a)^2 \widetilde{W}_1^2. \tag{6.6}$$

The proof is now completed as using the same arguments as in the first part, however with W_k replaced by $\widetilde{W}_k$. □

Proof of Corollary 1.9. Since the uniform distribution on $\{-1,+1\}^n$ satisfies a $\mathfrak{d}$-LSI(1), an inequality with 4 replaced by 8 in (1.29) follows immediately from Theorem 1.8. To obtain the constant 4, we need to use the (sharper) inequality $|\mathfrak{d} f(X)|^2 \leq \frac{1}{2}(b-a)^2 W_1^2$, valid for Rademacher random variables, which can be seen by analyzing (6.2) and the estimate thereafter. □

7. Auxiliary results: Proofs

This section contains the proofs of the auxiliary statements used in Section 4. Recall the (formal) operator $T_i(X) = (\overline{X}_i, X_i')$, where X' is an independent copy of X.

Proof of Lemma 4.3. We have

$$\begin{aligned}
|\mathfrak{h}^+ |\mathfrak{h}^{(d-1)} f|_{\mathrm{op}}(X)|^2 &= \sum_{i=1}^n \Big\| \Big(|\mathfrak{h}^{(d-1)} f(X)|_{\mathrm{op}} - |\mathfrak{h}^{(d-1)} f(T_i X)|_{\mathrm{op}} \Big)_+^2 \Big\|_{i,\infty} \\
&= \sum_{i=1}^n \Big\| \Big(\sup_{v^1,\ldots,v^{d-1}} \langle v^1 \cdots v^{d-1}, \mathfrak{h}^{(d-1)} f(X) \rangle - \sup_{v^1,\ldots,v^{d-1}} \langle v^1 \cdots v^{d-1}, \mathfrak{h}^{(d-1)} f(T_i X) \rangle \Big)_+^2 \Big\|_{i,\infty} \\
&\leq \sum_{i=1}^n \Big\| \Big(\langle \widetilde{v}^1 \cdots \widetilde{v}^{d-1}, \mathfrak{h}^{(d-1)} f(X) - \mathfrak{h}^{(d-1)} f(T_i X) \rangle \Big)_+^2 \Big\|_{i,\infty} \\
&\leq \sum_{i=1}^n \Big\| \Big(\sum_{i_1 \ldots i_{d-1}} \widetilde{v}_{i_1}^1 \cdots \widetilde{v}_{i_{d-1}}^{d-1} \Big(\Big\| \prod_{j=1}^{d-1} (\mathrm{Id} - T_{i_s}) f(X) \Big\|_{i_1 \cdots i_{d-1},\infty} - \Big\| \prod_{j=1}^{d-1} (\mathrm{Id} - T_{i_s}) f(T_i X) \Big\|_{i_1 \cdots i_{d-1},\infty} \Big) \Big)_+^2 \Big\|_{i,\infty} \\
&\leq \sum_{i=1}^n \Big\| \sum_{i_1,\ldots,i_d} \widetilde{v}_{i_1}^1 \cdots \widetilde{v}_{i_{d-1}}^{d-1} \Big\| (\mathrm{Id} - T_i) \prod_{j=1}^{d-1} (\mathrm{Id} - T_{i_s}) f(X) \Big\|_{i,i_1 \cdots i_{d-1},\infty}^2 \Big\|_{i,\infty} \\
&= \sum_{i=1}^n \Big(\sum_{i_1,\ldots,i_{d-1}} \widetilde{v}_{i_1}^1 \cdots \widetilde{v}_{i_1}^{d-1} \mathfrak{h}_{i i_1 \cdots i_{d-1}} f(X) \Big)^2 \\
&= \Big(\sup_{v^d : \|v^d\|_2 \leq 1} \sum_{i_d=1}^n \sum_{i_1,\ldots,i_{d-1}} \widetilde{v}_{i_1}^1 \cdots \widetilde{v}_{i_{d-1}}^{d-1} v_{i_d}^d \mathfrak{h}_{i_1 \cdots i_d} f(X) \Big)^2 \\
&\leq \Big(\sup_{v^1,\ldots,v^d : \|v^j\|_2 \leq 1} \sum_{i_1,\ldots,i_d} v_{i_1}^1 \cdots v_{i_d}^d \mathfrak{h}_{i_1 \cdots i_d} f(X) \Big)^2 \\
&= |\mathfrak{h}^{(d)} f(X)|_{\mathrm{op}}^2,
\end{aligned}$$

where in the first inequality we insert the vectors $\widetilde{v}^1, \ldots, \widetilde{v}^{d-1}$ maximizing the supremum. Taking the square root yields the claim. □

Proof of Proposition 4.4. Let $p > 0$, and let f be any measurable function on an arbitrary probability space such that $0 < \|f\|_{p+\varepsilon} < \infty$ for some $\varepsilon > 0$. Then, we have the general formula

$$\frac{d}{dp} \|f\|_p = \frac{1}{p^2} \|f\|_p^{1-p} \operatorname{Ent}(|f|^p). \tag{7.1}$$

In particular, it follows that

$$\frac{d}{dp}\|f\|_p^2 = \frac{2}{p^2}\|f\|_p^{2-p}\,\mathrm{Ent}(|f|^p). \tag{7.2}$$

Moreover, note that for any $i \in \{1, \ldots, n\}$

$$\begin{aligned}\mathbb{E}_\mu(\mathfrak{d}_i f)^2 &= \frac{1}{2}\iint (f(x) - f(\overline{x_i}, y_i))^2 d\mu(y_i \mid \overline{x}_i) d\mu(x) \\ &= \iiint (f(\overline{x}_i, x_i') - f(\overline{x}_i, x_i''))_+^2 d\mu(x_i' \mid \overline{x}_i) d\mu(x_i'' \mid \overline{x}_i) d\overline{\mu}_i(\overline{x}_i) \\ &= \iint (f(x) - f(\overline{x}_i, x_i'))_+^2 d\mu(x_i' \mid \overline{x}_i) d\mu(x).\end{aligned}$$

Therefore, it follows that

$$\mathbb{E}_\mu|\mathfrak{d}f|^2 = \sum_{i=1}^n \iint (f(x) - f(\overline{x}_i, x_i'))_+^2\, d\mu(x_i' \mid \overline{x}_i) d\mu(x). \tag{7.3}$$

Now let $p > 2$ and f be non-constant. (The assumption $\|f\|_{p+\varepsilon} < \infty$ is always true since $f \in L^\infty(\mu)$.) Applying the logarithmic Sobolev inequality to the function $|f|^{p/2}$ and rewriting this in terms of (7.3) yields

$$\mathrm{Ent}(|f|^p) \le 2\sigma^2 \sum_{i=1}^n \iint \left(|f|^{p/2}(x) - |f|^{p/2}(\overline{x}_i, x_i')\right)_+^2 d\mu(x_i' \mid \overline{x}_i) d\mu(x) \tag{7.4}$$

Using the inequality

$$(|a|^{p/2} - |b|^{p/2})_+^2 \le \frac{p^2}{4}|a|^{p-2}(|a| - |b|)_+^2$$

which is valid for all $a, b \in \mathbb{R}$ and $p \ge 2$, we obtain

$$(|f|^{p/2} - |f|^{p/2}(\overline{x}_i, x_i'))_+^2 \le \frac{p^2}{4}(|f| - |f|(\overline{x}_i, x_i'))_+^2 |f|^{p-2}$$

from which it follows in combination with (7.4) that

$$\mathrm{Ent}(|f|^p) \le p^2\sigma^2 \int |f|^{p-2} \sum_{i=1}^n (\mathfrak{d}_i f)^2 d\mu = p^2\sigma^2\, \mathbb{E}_\mu |f|^{p-2} |\mathfrak{d} f|^2$$

and

$$\mathrm{Ent}(|f|^p) \le \frac{p^2\sigma^2}{2} \mathbb{E}_\mu |f|^{p-2} |\mathfrak{h}^+|f||^2.$$

Hölder's inequality with exponents $\frac{p}{2}$ and $\frac{p}{p-2}$ applied to the last integral yields

$$\mathrm{Ent}(|f|^p) \le p^2\sigma^2 \|\mathfrak{d}f\|_p^2 \|f\|_p^{p-2}$$

and

$$\mathrm{Ent}(|f|^p) \le \frac{p^2\sigma^2}{2} \|\mathfrak{h}^+|f|\|_p^2 \|f\|_p^{p-2}$$

respectively. Combining this with (7.2), we arrive at the differential inequalities $\frac{d}{dp}\|f\|_p^2 \le 2\sigma^2 \|\mathfrak{d}f\|_p^2$ and $\frac{d}{dp}\|f\|_p^2 \le \sigma^2 \|\mathfrak{h}^+|f|\|_p^2$ respectively, which after integration gives (4.4) and (4.5).

Next, if μ satisfies an $\mathfrak{h}-\mathrm{LSI}(\sigma^2)$, we have

$$\begin{aligned}\mathrm{Ent}_\mu(f^2) &\le 2\sigma^2 \sum_{i=1}^n \int \sup_{x_i', x_i''} |f(\overline{x}_i, x_i') - f(\overline{x}_i, x_i'')|^2 d\mu(x) \\ &\le 8\sigma^2 \sum_{i=1}^n \int \sup_{x_i'} |f(x) - f(\overline{x}_i, x_i')|^2 d\mu(x).\end{aligned}$$

As above, this leads to

$$\mathrm{Ent}_\mu(|f|^p) \le 2\sigma^2 p^2 \int |f|^{p-2} |\mathfrak{h} f|^2 d\mu \le 2\sigma^2 p^2 \|f\|_p^{p-2} \|\mathfrak{h} f\|_p^2,$$

i.e.

$$\frac{d}{dp} \|f\|_p^2 \le 4\sigma^2 \|\mathfrak{h} f\|_p^2.$$

□

References

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

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

[AKPS18] Radosław Adamczak, Michał Kotowski, Bartłomiej Polaczyk, and Michał Strzelecki. "A note on concentration for polynomials in the Ising model". In: *arXiv preprint* (2018). arXiv: 1809.03187.

[ALM18] Radosław Adamczak, Rafał Latała, and Rafał Meller. "Hanson-Wright inequality in Banach spaces". In: *arXiv preprint* (2018). arXiv: 1811.00353.

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

[AS94] Shigeki Aida and Daniel W. Stroock. "Moment estimates derived from Poincaré and logarithmic Sobolev inequalities". In: *Math. Res. Lett.* 1.1 (1994), pp. 75–86. DOI: 10.4310/MRL.1994.v1.n1.a9.

[AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Second. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2008, pp. x+334. ISBN: 978-3-7643-8721-1.

[BGLZ17] Bhaswar B. Bhattacharya, Shirshendu Ganguly, Eyal Lubetzky, and Yufei Zhao. "Upper tails and independence polynomials in random graphs". In: *Adv. Math.* 319 (2017), pp. 313–347. DOI: 10.1016/j.aim.2017.08.003.

[Bob10] Sergey G. Bobkov. "The growth of L^p-norms in presence of logarithmic Sobolev inequalities". In: *Vestnik Syktyvkar Univ.* 11.2 (2010), pp. 92–111.

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

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

[BT06] Sergey G. Bobkov and Prasad Tetali. "Modified logarithmic Sobolev inequalities in discrete settings". In: *J. Theoret. Probab.* 19.2 (2006), pp. 289–336. DOI: 10.1007/s10959-006-0016-3.

[BBLM05] Stéphane Boucheron, Olivier Bousquet, Gábor Lugosi, and Pascal Massart. "Moment inequalities for functions of independent random variables". In: *Ann. Probab.* 33.2 (2005), pp. 514–560. DOI: 10.1214/009117904000000856.

[BLM03] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. "Concentration inequalities using the entropy method". In: *Ann. Probab.* 31.3 (2003), pp. 1583–1614. DOI: 10.1214/aop/1055425791.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities*. A nonasymptotic theory of independence, With a foreword by Michel Ledoux. Oxford University Press, Oxford, 2013, pp. x+481. ISBN: 978-0-19-953525-5.

[Bur07] Frank E. Burk. *A garden of integrals*. Vol. 31. The Dolciani Mathematical Expositions. Mathematical Association of America, Washington, DC, 2007, pp. xiv+281. ISBN: 978-0-88385-337-5.

[CMT15] Pietro Caputo, Georg Menz, and Prasad Tetali. "Approximate tensorization of entropy at high temperature". In: *Ann. Fac. Sci. Toulouse Math. (6)* 24.4 (2015), pp. 691–716. DOI: 10.5802/afst.1460.

[Cha04] Djalil Chafaï. "Entropies, convexity, and functional inequalities: on Φ-entropies and Φ-Sobolev inequalities". In: *J. Math. Kyoto Univ.* 44.2 (2004), pp. 325–363. DOI: 10.1215/kjm/1250283556.

[Cha12] Sourav Chatterjee. "The missing log in large deviations for triangle counts". In: *Random Structures Algorithms* 40.4 (2012), pp. 437–451. DOI: 10.1002/rsa.20381.

[CD16] Sourav Chatterjee and Amir Dembo. "Nonlinear large deviations". In: *Adv. Math.* 299 (2016), pp. 396–450. DOI: 10.1016/j.aim.2016.05.017.

[CY18] Xiaohui Chen and Yun Yang. "Hanson-Wright inequality in Hilbert spaces with application to K-means clustering for non-Euclidean data". In: *arXiv preprint* (2018). arXiv: 1810.11180.

[CD18] Nicholas A. Cook and Amir Dembo. "Large deviations of subgraph counts for sparse Erdös–Rényi graphs". In: *arXiv preprint* (2018). arXiv: 1809.11148.

[DM78] Claude Dellacherie and Paul-André Meyer. *Probabilities and potential*. Vol. 29. North-Holland Mathematics Studies. North-Holland Publishing Co., Amsterdam-New York; North-Holland Publishing Co., Amsterdam-New York, 1978, pp. viii+189. ISBN: 0-7204-0701-X.

[DK12b] Bobby DeMarco and Jef Kahn. "Tight upper tail bounds for cliques". In: *Random Structures Algorithms* 41.4 (2012), pp. 469–487. DOI: 10.1002/rsa.20440.

[DK12a] Bobby DeMarco and Jeff Kahn. "Upper tails for triangles". In: *Random Structures Algorithms* 40.4 (2012), pp. 452–459. DOI: 10.1002/rsa.20382.

[DS96] Persi Diaconis and Laurent Saloff-Coste. "Logarithmic Sobolev inequalities for finite Markov chains". In: *Ann. Appl. Probab.* 6.3 (1996), pp. 695–750. DOI: 10.1214/aoap/1034968224.

[ES81] B. Efron and C. Stein. "The jackknife estimate of variance". In: *Ann. Statist.* 9.3 (1981), pp. 586–596. DOI: 10.1214/aos/1176345462.

[GSS18] Friedrich Götze, Holger Sambale, and Arthur Sinulis. "Higher order concentration for functions of weakly dependent random variables". In: *arXiv preprint* (2018). arXiv: 1801.06348.

[Gro75] Leonard Gross. "Logarithmic Sobolev inequalities". In: *Amer. J. Math.* 97.4 (1975), pp. 1061–1083. DOI: 10.2307/2373688.

[vH16] Ramon van Handel. *Probability in High Dimension*. APC 550 Lecture Notes, Princeton University. 2016.

[HW71] David L. Hanson and F. T. Wright. "A bound on tail probabilities for quadratic forms in independent random variables". In: *Ann. Math. Statist.* 42 (1971), pp. 1079–1083. DOI: 10.1214/aoms/1177693335.

[HS75] Edwin Hewitt and Karl Stromberg. *Real and abstract analysis*. A modern treatment of the theory of functions of a real variable, Third printing, Graduate Texts in Mathematics, No. 25. Springer-Verlag, New York-Heidelberg, 1975, pp. x+476.

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

[JŁR00] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000, pp. xii+333. ISBN: 0-471-17541-2. DOI: 10.1002/9781118032718.

[JOR04] Svante Janson, Krzysztof Oleszkiewicz, and Andrzej Ruciński. "Upper tails for subgraph counts in random graphs". In: *Israel J. Math.* 142 (2004), pp. 61–92. DOI: 10.1007/BF02771528.

[JR02] Svante Janson and Andrzej Ruciński. "The infamous upper tail". In: *Random Structures Algorithms* 20.3 (2002). Probabilistic methods in combinatorial optimization, pp. 317–342. DOI: 10.1002/rsa.10031.

[KV00] Jeong H. Kim and Van H. Vu. "Concentration of multivariate polynomials and its applications". In: *Combinatorica* 20.3 (2000), pp. 417–434. DOI: 10.1007/s004930070014.

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

[Led97] Michel Ledoux. "On Talagrand's deviation inequalities for product measures". In: *ESAIM Probab. Statist.* 1 (1997), pp. 63–87. DOI: 10.1051/ps:1997103.

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

[LZ17] Eyal Lubetzky and Yufei Zhao. "On the variational problem for upper tails in sparse random graphs". In: *Random Structures Algorithms* 50.3 (2017), pp. 420–436. DOI: 10.1002/rsa.20658.

[Mar15] Katalin Marton. "Logarithmic Sobolev inequalities in discrete product spaces: a proof by a transportation cost distance". In: *arXiv preprint* (2015). arXiv: 1507.02803.

[MS86] Vitali D. Milman and Gideon Schechtman. *Asymptotic theory of finite-dimensional normed spaces.* Vol. 1200. Lecture Notes in Mathematics. With an appendix by M. Gromov. Springer-Verlag, Berlin, 1986, pp. viii+156. ISBN: 3-540-16769-2.

[MOO10] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. "Noise stability of functions with low influences: invariance and optimality". In: *Ann. of Math. (2)* 171.1 (2010), pp. 295–341. DOI: 10.4007/annals.2010.171.295.

[Pan04] Dmitry Panchenko. "Deviation inequality for monotonic Boolean functions with application to the number of k-cycles in a random graph". In: *Random Structures Algorithms* 24.1 (2004), pp. 65–74. DOI: 10.1002/rsa.10105.

[RS14] Maxim Raginsky and Igal Sason. *Concentration of measure inequalities in information theory, communications, and coding.* Now Publishers Inc., 2014, pp. 1–260. ISBN: 978-1-60198-906-2.

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

[SS18] Holger Sambale and Arthur Sinulis. "Concentration of measure for finite spin systems". In: *arXiv preprint* (2018). arXiv: 1807.07765.

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

[Ste86] J. Michael Steele. "An Efron-Stein inequality for nonsymmetric statistics". In: *Ann. Statist.* 14.2 (1986), pp. 753–758. DOI: 10.1214/aos/1176349952.

[Tal91] Michel Talagrand. "A new isoperimetric inequality and the concentration of measure phenomenon". In: *Geometric aspects of functional analysis (1989–90).* Vol. 1469. Lecture Notes in Math. Springer, Berlin, 1991, pp. 94–124. DOI: 10.1007/BFb0089217.

[Tal96] Michel Talagrand. "New concentration inequalities in product spaces". In: *Invent. Math.* 126.3 (1996), pp. 505–563. DOI: 10.1007/s002220050108.

[Vu02] Van H. Vu. "Concentration of non-Lipschitz functions and applications". In: *Random Structures Algorithms* 20.3 (2002). Probabilistic methods in combinatorial optimization, pp. 262–316. DOI: 10.1002/rsa.10032.

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

[Wri73] F. T. Wright. "A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric". In: *Ann. Probability* 1.6 (1973), pp. 1068–1070. DOI: 10.1214/aop/1176996815.