



# Concentration Inequalities for Bounded Functionals via Log-Sobolev-Type Inequalities

Friedrich Götze<sup>1</sup> · Holger Sambale<sup>1</sup> · Arthur Sinulis<sup>1</sup>

Received: 13 February 2020 / Revised: 17 May 2020 / Published online: 12 June 2020  
© The Author(s) 2020

## Abstract

In this paper, we prove multilevel concentration inequalities for bounded functionals  $f = f(X_1, \dots, X_n)$  of random variables  $X_1, \dots, X_n$  that are either independent or satisfy certain logarithmic Sobolev inequalities. The constants in the tail estimates depend on the operator norms of  $k$ -tensors of higher order differences of  $f$ . We provide applications for both dependent and independent random variables. This includes deviation inequalities for empirical processes  $f(X) = \sup_{g \in \mathcal{F}} |g(X)|$  and suprema of homogeneous chaos in bounded random variables in the Banach space case  $f(X) = \sup_t \|\sum_{i_1 \neq \dots \neq i_d} t_{i_1 \dots i_d} X_{i_1} \cdots X_{i_d}\|_{\mathcal{B}}$ . The latter application is comparable to earlier results of Boucheron, Bousquet, Lugosi, and Massart and provides the upper tail bounds of Talagrand. In the case of Rademacher random variables, we give an interpretation of the results in terms of quantities familiar in Boolean analysis. Further applications are concentration inequalities for  $U$ -statistics with bounded kernels  $h$  and for the number of triangles in an exponential random graph model.

**Keywords** Concentration of measure · Empirical processes · Functional inequalities · Hamming cube · Logarithmic Sobolev inequality · Product spaces · Suprema of chaos · Weakly dependent random variables

**Mathematics Subject Classification (2010)** 60E15 · 05C80

---

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*”.

---

✉ Arthur Sinulis  
asinulis@math.uni-bielefeld.de  
Friedrich Götze  
goetze@math.uni-bielefeld.de  
Holger Sambale  
hsambale@math.uni-bielefeld.de

<sup>1</sup> Fakultät für Mathematik, Universität Bielefeld, Postfach 10 01 31, 33501 Bielefeld, Germany

## 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, as is witnessed by the monographs [18,38,42,45,54]. 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 [51,52], mainly in the works [11,37], and further developed in [41].

To convey the idea, let us recall that the *logarithmic Sobolev inequality* for the standard Gaussian measure  $\nu$  in  $\mathbb{R}^n$  (see [29]) states that for any  $f \in C_c^\infty(\mathbb{R}^n)$  we have

$$\text{Ent}_\nu(f^2) \leq 2 \int |\nabla f|^2 d\nu, \quad (1)$$

where  $\text{Ent}_\nu(f^2) = \int f^2 \log f^2 d\nu - \int f^2 d\nu \log \int f^2 d\nu$  is the *entropy functional*. Informally, it bounds the disorder of a function  $f$  (under  $\nu$ ) by its average local fluctuations, measured in terms of the length of the gradient. It is by now standard that (1) implies subgaussian tail decay for Lipschitz functions (e. g. by means of the Herbst argument). In particular, if  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is a  $C^1$  function such that  $|\nabla f| \leq L$  a.s., we have  $\nu(|f - \int f d\nu| \geq t) \leq 2 \exp(-t^2/(2L^2))$  for any  $t \geq 0$ .

If  $\mu$  is a probability measure on a discrete set  $\mathcal{X}$  (or a more abstract set not allowing for an immediate replacement for  $|\nabla f|$ ), then there are several ways to reformulate Eq. (1), see e. g. [12,26]. We continue these ideas by working in the framework of *difference operators*. Given a probability space  $(\mathcal{Y}, \mathcal{A}, \mu)$ , we call any operator  $\Gamma: L^\infty(\mu) \rightarrow L^\infty(\mu)$  satisfying  $|\Gamma(af + b)| = a |\Gamma f|$  for all  $a > 0, b \in \mathbb{R}$  a difference operator. Accordingly, we say that  $\mu$  satisfies a  $\Gamma$ -LSI( $\sigma^2$ ), if for all bounded measurable functions  $f$  we have

$$\text{Ent}_\mu(f^2) \leq 2\sigma^2 \int \Gamma(f)^2 d\mu. \quad (2)$$

Apart from the domain of  $\Gamma$ , it is clear that (2) can be seen as generalization of (1) by defining  $\Gamma(f) = |\nabla f|$  on  $\mathbb{R}^n$ .

Another route to obtain concentration inequalities is to modify the entropy method, which was done in the framework of so-called  $\varphi$ -entropies. The idea is to replace the function  $\varphi_0(x) := x \log x$  in the definition of the entropy  $\text{Ent}_\mu^{\varphi_0}(f) = \mathbb{E}_\mu \varphi_0(f) - \varphi_0(\mathbb{E}_\mu f)$  by other functions  $\varphi$ . This has been studied in [17,22,36]. In the seminal work [16] the authors proved inequalities for  $\varphi$ -entropies for power functions  $\varphi(x) = |x|^\alpha, \alpha \in (1, 2]$ , leading to moment inequalities for independent random variables.

Originally, the entropy method was primarily used to prove sub-Gaussian concentration inequalities for Lipschitz-type functions. However, there are many situations of interest in which the functions under consideration are not Lipschitz or have Lipschitz constants which grow as the dimension increases even after a renormalization which asymptotically stabilizes the variance. Among the simplest examples are polynomial-

type functions. Here, the boundedness of the gradient typically has to be replaced by more elaborate conditions on higher order derivatives (up to some order  $d$ ). Moreover, we cannot have subgaussian tail decay anymore. This is already obvious if we consider the product of two independent standard normal random variables, which leads to subexponential tails. We refer to this topic as *higher order concentration*.

The earliest higher order concentration results date back to the late 1960s. Already in [13,14,43], the growth of  $L^p$  norms and hypercontractive estimates of polynomial-type functions in Rademacher or Gaussian random variables, respectively, have been studied. The question of estimating the growth of  $L^p$  norms of multilinear polynomials in Gaussian random variables was considered in [8,15,35]. In the context of Erdős–Rényi graphs and the triangle problem, concentration inequalities for polynomials functions gained considerable attention, in papers such as [33].

More recently, *multilevel concentration inequalities* have been proven in [1,5,56] for many classes of functions. These included  $U$ -statistics in independent random variables, functions of random vectors satisfying Sobolev-type inequalities and polynomials in sub-Gaussian random variables, respectively. We refer to inequalities of the type

$$\mathbb{P}\left(|f(X) - \mathbb{E} f(X)| \geq t\right) \leq 2 \exp\left(-\frac{1}{C} \min_{k=1,\dots,d} f_k(t)\right) \quad (3)$$

as multilevel or *higher order ( $d$ -th order) concentration inequalities*. This means that the tails might have different decay properties in some regimes of  $[0, \infty)$ . Usually, we have  $f_k(t) = (t/C_k)^{2/k}$  for some constant  $C_k$  which typically depends on the  $k$ -th order derivatives.

To convey the basic idea of multilevel concentration inequalities, let us once again consider the case  $d = 2$ , e. g. a quadratic form of independent, say, Gaussian random variables. As sketched above, in this case the tails decay subexponentially in general. By means of a multilevel concentration inequality (the so-called Hanson–Wright inequality, which we address in more detail at a later point), we can show that while for  $t$  large, subexponential tail decay holds, for small  $t$  we even get subgaussian decay. In this sense, multilevel concentration inequalities provide refined tail estimates which do not only cover the behavior for large  $t$ .

Our own work started with a second-order concentration inequality on the sphere in [9] and was continued in [10] for bounded functionals of various classes of random variables (e. g. independent random variables or in presence of a logarithmic Sobolev inequality (1)), and in [28] for weakly dependent random variables (e. g. the Ising model). In these papers, we studied higher order concentration, arriving at multi-level tail inequalities of type (3). If the underlying measure  $\mu$  satisfies a logarithmic Sobolev inequality, [10, Corollary 1.11] yields  $f_k(t) = (t/C_k)^{2/k}$  with  $C_k = (\int |f^{(k)}|_{\text{op}}^2 d\mu)^{1/2}$  for  $k = 1, \dots, d - 1$  and  $C_d = \sup |f^{(d)}|_{\text{op}}$ , where  $|f^{(k)}|_{\text{op}}$  denotes the operator norm of the respective tensors of  $k$ -th order partial derivatives. A downside in both [10,28] is that for functions of independent or weakly dependent random variables, comparable estimates involve Hilbert–Schmidt instead of operator norms, leading to weaker estimates in general.

A central aspect of the present article is to fix this drawback by a slightly more elaborate approach. Here, we consider both independent and dependent random variables. In either case, we prove multilevel concentration inequalities of the same type, and apply them to different forms of functionals. We provide improvements of earlier higher order concentration results like [10, Theorem 1.1] or [28, Theorem 1.5], replacing the Hilbert–Schmidt norms appearing therein by operator norms. This leads to sharper bounds and a wider range of applicability.

A special emphasis is placed on providing *uniform versions* of the higher order concentration inequalities. By this, we mean that we consider functionals of supremum type  $f(X) = \sup_{f \in \mathcal{F}} |f(X)|$ , which includes suprema of polynomial chaoses, or empirical processes. Two more applications are given by  $U$ -statistics in independent and weakly dependent random variables as well as a triangle counting statistic in some models of random graphs, for which we prove concentration inequalities.

*Notations* Throughout this note,  $X = (X_1, \dots, X_n)$  is a random vector taking values in some product space  $\mathcal{Y} = \otimes_{i=1}^n \mathcal{X}_i$  (equipped with the product  $\sigma$ -algebra) with law  $\mu$ , defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . By abuse of language, we say that  $X$  satisfies a  $\Gamma$ -LSI( $\sigma^2$ ), if its distribution does. In any finite-dimensional vector space, we let  $|\cdot|$  be the Euclidean norm, and for brevity, we write  $[q] := \{1, \dots, q\}$  for any  $q \in \mathbb{N}$ . Given a vector  $x = (x_j)_{j=1, \dots, n}$  we write  $x_{i^c} = (x_j)_{j \neq i}$ . To any  $d$ -tensor  $A$  we define the Hilbert–Schmidt norm  $|A|_{\text{HS}} := (\sum_{i_1, \dots, i_d} A_{i_1 \dots i_d}^2)^{1/2}$  and the operator norm

$$|A|_{\text{op}} := \sup_{\substack{v^1, \dots, v^d \in \mathbb{R}^n \\ |v^j| \leq 1}} \langle v^1 \cdots v^d, A \rangle = \sup_{\substack{v^1, \dots, v^d \\ |v^j| \leq 1}} \sum_{i_1, \dots, i_d} v_{i_1}^1 \cdots v_{i_d}^d A_{i_1 \dots i_d},$$

using the outer product  $(v^1 \cdots v^d)_{i_1 \dots i_d} = \prod_{j=1}^d v_{i_j}^j$ . For brevity, for any random  $k$ -tensor  $A$  and any  $p \in (0, \infty]$  we abbreviate  $\|A\|_{\text{HS}, p} = (\mathbb{E} |A|_{\text{HS}}^p)^{1/p}$  as well as  $\|A\|_{\text{op}, p} = (\mathbb{E} |A|_{\text{op}}^p)^{1/p}$ . Lastly, we ignore any measurability issues that may arise. Thus, we assume that all the suprema used in this work are either countable or defined as  $\sup_{t \in T} = \sup_{F \subset T: F \text{ finite}} \sup_{t \in F}$ .

### 1.1 Main Results

To formulate our main results, we introduce a difference operator labeled  $\mathfrak{h}f$  which is frequently used in the method of bounded differences. Let  $X' = (X'_1, \dots, X'_n)$  be an independent copy of  $X$ , defined on the same probability space. Given  $f(X) \in L^\infty(\mathbb{P})$ , define for each  $i \in [n]$

$$T_i f := T_i f(X) := f(X_{i^c}, X'_i) = f(X_1, \dots, X_{i-1}, X'_i, [2]X_{i+1}, \dots, X_n)$$

and

$$\mathfrak{h}_i f(X) = \|f(X) - T_i f(X)\|_{i, \infty}, \quad \mathfrak{h}f(X) = (\mathfrak{h}_1 f(X), \dots, \mathfrak{h}_n f(X)),$$

where  $\|\cdot\|_{i,\infty}$  denotes the  $L^\infty$ -norm with respect to  $(X_i, X'_i)$ . The difference operator  $|\mathfrak{h}f|$  is given as the Euclidean norm of the vector  $\mathfrak{h}f$ .

We shall also need higher order versions of  $\mathfrak{h}$ , denoted by  $\mathfrak{h}^{(d)}f$ . They can be thought of as analogues of the  $d$ -tensors of all partial derivatives of order  $d$  in an abstract setting. To define the  $d$ -tensor  $\mathfrak{h}^{(d)}f$ , we specify it on its “coordinates”. That is, given distinct indices  $i_1, \dots, i_d$ , we set

$$\begin{aligned} \mathfrak{h}_{i_1\dots i_d}f(X) &= \left\| \prod_{s=1}^d (\text{Id} - T_{i_s})f(X) \right\|_{i_1,\dots,i_d,\infty} \\ &= \left\| f(X) + \sum_{k=1}^d (-1)^k \sum_{1 \leq s_1 < \dots < s_k \leq d} T_{i_{s_1}\dots i_{s_k}}f(X) \right\|_{i_1,\dots,i_d,\infty} \end{aligned} \tag{4}$$

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

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

Using the definition (4), we define tensors of  $d$ -th order differences as follows:

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

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  $\mathfrak{h}^{(d)}f(X)$ .

Our first main theorem is a concentration inequality for general, bounded functionals of independent random variables  $X_1, \dots, X_n$ .

**Theorem 1** *Let  $X$  be a random vector with independent components,  $f : \mathcal{Y} \rightarrow \mathbb{R}$  a measurable function satisfying  $f = f(X) \in L^\infty(\mathbb{P})$ ,  $d \in \mathbb{N}$  and define  $C := 217d^2$ . We have for any  $t \geq 0$*

$$\mathbb{P}(|f - \mathbb{E}f| \geq t) \leq 2 \exp \left( -\frac{1}{C} \min_{k=1,\dots,d-1} \left( \frac{t}{\|\mathfrak{h}^{(k)}f\|_{\text{op},1}} \right)^{2/k} \wedge \left( \frac{t}{\|\mathfrak{h}^{(d)}f\|_{\text{op},\infty}} \right)^{2/d} \right). \tag{5}$$

For the sake of illustration, let us consider the case of  $d = 2$ . Assuming that  $X_1, \dots, X_n$  satisfy  $\mathbb{E}X_i = 0$ ,  $\mathbb{E}X_i^2 = 1$  and  $|X_i| \leq M$  a.s., let  $f(X)$  be the quadratic form  $f(X) = \sum_{i < j} a_{ij} X_i X_j = X^T A X$ . Here,  $a_{ij} \in \mathbb{R}$  for all  $i < j$ , and  $A$  is the symmetric matrix with zero diagonal and entries  $A_{ij} = a_{ij}/2$  if  $i < j$ . In this case, it is easy to see that  $\|\mathfrak{h}f\|_{\text{op},1} \leq \|\mathfrak{h}f\|_{\text{op},2} \leq 4M|A|_{\text{HS}}$  and  $\|\mathfrak{h}^{(2)}f\|_{\text{op},\infty} \leq 8M^2|A^{\text{abs}}|_{\text{op}}$ , where  $A^{\text{abs}}$  is the matrix given by  $(A^{\text{abs}})_{ij} = |a_{ij}|$ . As a result,

$$\mathbb{P}(|f - \mathbb{E}f| \geq t) \leq 2 \exp \left( -\frac{1}{CM^2} \min \left( \frac{t^2}{|A|_{\text{HS}}^2}, \frac{t}{|A^{\text{abs}}|_{\text{op}}} \right) \right).$$

This is a version of the famous *Hanson–Wright inequality*. For the various forms of the Hanson–Wright inequality we refer to [2,4,30,32,47,55,57].

Note that by a modification of our proofs (using arguments especially adapted to polynomials), it is possible to replace  $|A^{\text{abs}}|_{\text{op}}$  by  $|A|_{\text{op}}$ , thus avoiding the drawback of switching to a matrix with a possibly larger operator norm. See Sects. 2.1 and 2.4 for details. On the other hand, Theorem 1 allows for *any* function  $f$ , not just quadratic forms, and the case of  $d = 2$  can in this sense be considered as generalization of the Hanson–Wright inequality.

For a certain class of weakly dependent random variables  $X_1, \dots, X_n$ , we can prove similar estimates as in Theorem 1. To this end, we introduce another difference operator, which is more familiar in the context of logarithmic Sobolev inequalities for Markov chains as developed in [26]. Assume that  $\mathcal{Y} = \otimes_{i=1}^n \mathcal{X}_i$  for some finite sets  $\mathcal{X}_1, \dots, \mathcal{X}_n$ , equipped with a probability measure  $\mu$  and let  $\mu(\cdot | x_{i^c})$  denote the conditional measure (interpreted as a measure on  $\mathcal{X}_i$ ) and  $\mu_{i^c}$  the marginal on  $\otimes_{j \neq i} \mathcal{X}_j$ . Finally, set

$$\begin{aligned} |\mathfrak{D}f|^2(x) &:= \sum_{i=1}^n (\mathfrak{D}_i f(x))^2 := \sum_{i=1}^n \text{Var}_{\mu(\cdot | x_{i^c})}(f(x_{i^c}, \cdot)) \\ &= \sum_{i=1}^n \frac{1}{2} \iint (f(x_{i^c}, y) - f(x_{i^c}, y'))^2 d\mu(y | x_{i^c}) d\mu(y' | x_{i^c}). \end{aligned}$$

This difference operator appears naturally in the Dirichlet form associated to the Glauber dynamic of  $\mu$ , given by

$$\mathcal{E}(f, f) := \sum_{i=1}^n \int \text{Var}_{\mu(\cdot | x_{i^c})}(f(x_{i^c}, \cdot)) d\mu_{i^c}(x_{i^c}) = \int |\mathfrak{D}f|^2 d\mu.$$

In the next theorem, we require a  $\mathfrak{D}$ -LSI for the underlying random variables  $X_1, \dots, X_n$ . A number of models which satisfy this assumption will be discussed below.

**Theorem 2** *Let  $X = (X_1, \dots, X_n)$  be a random vector satisfying a  $\mathfrak{D}$ -LSI( $\sigma^2$ ) and  $f : \mathcal{Y} \rightarrow \mathbb{R}$  a measurable function with  $f = f(X) \in L^\infty(\mathbb{P})$ . With the constant  $C = 15\sigma^2 d^2 > 0$  we have for any  $t \geq 0$*

$$\mathbb{P}(|f - \mathbb{E}f| \geq t) \leq 2 \exp \left( -\frac{1}{C} \min_{k=1, \dots, d-1} \left( \frac{t}{\|b^{(k)} f\|_{\text{op}, 1}} \right)^{2/k} \wedge \left( \frac{t}{\|b^{(d)} f\|_{\text{op}, \infty}} \right)^{2/d} \right). \tag{6}$$

Again, if  $d = 2$ , assuming that  $\mathbb{E}X_i = 0, \mathbb{E}X_i^2 = 1, |X_i| \leq M$  a.s. and  $\mathbb{E}X_i X_j = 0$  if  $i \neq j$ , we arrive at a Hanson–Wright type inequality, this time including dependent situations. Similar results still hold if we remove the uncorrelatedness condition.

Let us discuss the  $\mathfrak{D}$ -LSI condition in more detail. First, any collection of random independent variables  $X_1, \dots, X_n$  with finitely many values satisfies a  $\mathfrak{D}$ -LSI( $\sigma^2$ ) with  $\sigma^2$  depending on the minimal nonzero probability of the  $X_i$  (cf. Proposition 6). In this situation, Theorems 1 and 2 only differ by constants.

However, the  $\mathfrak{d}$ -LSI conditions also gives rise to numerous models of dependent random variables as in [28, Proposition 1.1] (the Ising model) or [48, Theorem 3.1] (various different models). Let us recall some of them. The *Ising model* is the probability measure on  $\{\pm 1\}^n$  defined by normalizing  $\pi(\sigma) = \exp(\frac{1}{2} \sum_{i,j} J_{ij} \sigma_i \sigma_j + \sum_{i=1}^n h_i \sigma_i)$  for a symmetric matrix  $J = (J_{ij})$  with zero diagonal and some  $h \in \mathbb{R}^n$ . In [28, Proposition 1.1], we have shown that if  $\max_{i=1, \dots, n} \sum_{j=1}^n |J_{ij}| \leq 1 - \alpha$  and  $\max_{i \in [n]} |h_i| \leq \tilde{\alpha}$ , the Ising model satisfies a  $\mathfrak{d}$ -LSI( $\sigma^2$ ) with  $\sigma^2$  depending on  $\alpha$  and  $\tilde{\alpha}$  only. For the special case of  $h = 0$  and  $J_{ij} = \beta$  for all  $i \neq j$ , we obtain the *Curie–Weiss model*. Here, the two conditions required above reduce to  $\beta < 1$ .

Another simple model in which a  $\mathfrak{d}$ -LSI holds is the *random coloring model*. If  $G = (V, E)$  is a finite graph and  $C = \{1, \dots, k\}$  is a set of colors, we denote by  $\Omega_0 \subset C^V$  the set of all proper coloring, i.e. the set of all  $\omega \in C^V$  such that  $\{v, w\} \in E \Rightarrow \omega_v \neq \omega_w$ . In [48, Theorem 3.1], we have shown that the uniform distribution on  $\Omega_0$  satisfies a  $\mathfrak{d}$ -LSI if the maximum degree  $\Delta$  is uniformly bounded and  $k \geq 2\Delta + 1$  (strictly speaking, we consider sequences of graphs here). In [48, Theorem 3.1], we moreover prove  $\mathfrak{d}$ -LSIs for the (vertex-weighted) *exponential random graph model* and the *hard-core model*. We will further discuss the exponential random graph model in Sect. 2.4.

The common feature in all these models is that the dependencies which appear can be controlled (e.g. by means of a coupling matrix which measures the interactions between the particles of the system under consideration, cf. [28, Theorem 4.2]) in such a way that the model is not “too far” from a product measure. For instance, in the Curie–Weiss model, this just translates to  $\beta < 1$ .

As a final remark, we discuss the LSI property with respect to various difference operators in Sect. 5. In particular, we show that the restriction to finite spaces which is implicit in Theorem 2 is natural since the  $\mathfrak{d}$ -LSI property requires the underlying space to be finite. By contrast, we prove that any set of independent random variables  $X_1, \dots, X_n$  satisfies an  $\mathfrak{h}$ -LSI(1). However, it seems that it is not possible to use the entropy method based on  $\mathfrak{h}$ -LSIs.

The upper bound in Theorem 2 admits a “uniform version”, i.e. we can prove deviation inequalities for suprema of functions, in the following sense. Let  $\mathcal{F}$  be a family of uniformly bounded, real-valued, measurable functions and set

$$g(X) := g_{\mathcal{F}}(X) := \sup_{f \in \mathcal{F}} |f(X)|. \tag{7}$$

For any  $d \in \mathbb{N}$  and  $j = 1, \dots, d$  let  $W_j = W_j(X) := \sup_{f \in \mathcal{F}} |\mathfrak{h}^{(j)} f(X)|_{\text{op}}$ .

**Theorem 3** *Assume that either  $X_1, \dots, X_n$  are independent or  $X$  satisfies a  $\mathfrak{d}$ -LSI( $\sigma^2$ ) and let  $g = g(X)$  be as in (7). With the same constant  $C$  as in Theorems 1 or 2, respectively, we have for any  $t \geq 0$  the deviation inequality*

$$\mathbb{P}(g - \mathbb{E} g \geq t) \leq 2 \exp \left( -\frac{1}{C} \min \left( \min_{j=1, \dots, d-1} \left( \frac{t}{\mathbb{E} W_j} \right)^{2/k}, \frac{t^{2/d}}{\|W_d\|_{\infty}} \right) \right)$$

As mentioned before, Theorem 3 yields bounds for the upper tail only. The background is that the entropy method has certain limitations when it is applied to suprema of functions, cf. also Proposition 1 or Theorem 4 below. Roughly sketched, the reason is that when evaluating difference operators of suprema, if a positive part is involved we may typically choose a coordinate-independent maximizer of the terms involved. Without a positive part, this is no longer possible. See in particular the proof of Theorem 4, where we provide some further details.

Functionals of the form (7) have been considered in various works, starting from the first results in [52, Theorem 1.4], and continued in [41, Theorem 3], [46, Théorème 1.1] and [19, Theorem 2.3] in the special case of

$$g(X) := \sup_{f \in \mathcal{F}} \left| \sum_{j=1}^n f(X_j) \right|. \quad (8)$$

Further research has been done in [34], [49, Sect. 3] and more recently [39, Proposition 5.4]. In these works, Bennett-type inequalities have been proven for general independent random variables. Furthermore, [16, Theorem 10] treats the case  $g(X) = \sup_{t \in \mathcal{T}} \sum_{i=1}^n t_i X_i$  for Rademacher random variables  $X_i$  and a compact set of vectors  $\mathcal{T} \subset \mathbb{R}^n$ . As a byproduct of our method, we prove a deviation inequality for  $g$  which can be regarded as a uniform bounded differences inequality.

**Proposition 1** *Assume that  $X = (X_1, \dots, X_n)$  satisfies a  $\vartheta$ -LSI( $\sigma^2$ ), let  $g = g(X)$  be as in (8), and let  $c(f)$  be such that  $|f(x) - f(y)| \leq c(f)$ . For any  $t \geq 0$  we have*

$$\mathbb{P} \left( g \geq \mathbb{E} g + t \right) \leq 2 \exp \left( - \frac{t^2}{15 \sigma^2 n \sup_{f \in \mathcal{F}} c(f)^2} \right).$$

Let us put Proposition 1 into context. In the above-mentioned works, the authors derive Bennett-type inequalities for independent random variables  $X_1, \dots, X_n$ , whereas in our case the concentration inequalities have sub-Gaussian tails. It might be compared to the sub-Gaussian tail estimates for Bernoulli processes, see e.g. [53, Theorem 5.3.2]. However, the  $\vartheta$ -LSI( $\sigma^2$ ) property is both more and less general. On the one hand, it is possible to include possibly dependent random vectors, but on the other hand for independent random variables it is only applicable if the  $X_i$  take finitely many values.

## 1.2 Outline

In Sect. 2, we present a number of applications and refinements of our main results. Section 3 contains the proofs of our main theorems. The proofs of the results from Sect. 2 is deferred to Sect. 4. We close out the paper by discussing different forms of logarithmic Sobolev inequalities with respect to various difference operators in the last Sect. 5.



## 2 Applications

In the sequel, we consider various situations in which our results can be applied. Some of them can be regarded as sharpenings of our main theorems for functions which have a special structure.

### 2.1 Uniform Bounds

If the functions under consideration are of polynomial type, we may somewhat refine the results from the previous section. Here, we focus on uniform bounds as discussed in Theorem 3.

Let  $\mathcal{I}_{n,d}$  denote the family of subsets of  $[n]$  with  $d$  elements, 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  $X = (X_1, \dots, X_n)$  be a random vector with support in  $[a, b]^n$  for some real numbers  $a < b$  and define

$$f(X) := f_{\mathcal{T}}(X) := \sup_{t \in \mathcal{T}} \left\| \sum_{I \in \mathcal{I}_{n,d}} X_I t_I \right\|, \tag{9}$$

where  $X_I := \prod_{i \in I} X_i$ . For any  $k \in [d]$  we let

$$\begin{aligned} W_k &:= \sup_{t \in \mathcal{T}} \sup_{v^* \in \mathcal{B}_1^*} \sup_{\substack{\alpha^1, \dots, \alpha^k \in \mathbb{R}^n \\ |\alpha^i| \leq 1}} v^* \left( \sum_{\substack{i_1, \dots, i_k \\ \text{distinct}}} \alpha_{i_1}^1 \cdots \alpha_{i_k}^k \sum_{\substack{I \in \mathcal{I}_{n,d-k} \\ i_1, \dots, i_k \notin I}} X_I t_{I \cup \{i_1, \dots, i_k\}} \right) \\ &= \sup_{t \in \mathcal{T}} \sup_{\substack{\alpha^1, \dots, \alpha^k \in \mathbb{R}^n \\ |\alpha^i| \leq 1}} \left\| \sum_{\substack{i_1, \dots, i_k \\ \text{distinct}}} \alpha_{i_1}^1 \cdots \alpha_{i_k}^k \sum_{\substack{I \in \mathcal{I}_{n,d-k} \\ i_1, \dots, i_k \notin I}} X_I t_{I \cup \{i_1, \dots, i_k\}} \right\|, \end{aligned} \tag{10}$$

where for  $k = d$  we use the convention  $\mathcal{I}_{n,0} = \{\emptyset\}$  and  $X_{\emptyset} := 1$ .

One can interpret the quantities  $W_k$  as follows: If  $f_t(x) = \sum_{I \in \mathcal{I}_{n,d}} x_I t_I$  is the corresponding polynomial in  $n$  variables, and  $\nabla^{(k)} f_t(x)$  is the  $k$ -tensor of all partial derivatives of order  $k$ , then  $W_k = \sup_{t \in \mathcal{T}} |\nabla^{(k)} f_t(X)|_{\text{op}}$ . In this sense, we are considering the same quantities as in Theorem 3 but replace the difference operator  $\mathfrak{h}$  by formal derivatives of the polynomial under consideration.

Furthermore, the concentration inequalities are phrased with the help of the quantities

$$\tilde{W}_k := \sup_{\substack{\alpha^1, \dots, \alpha^k \in \mathbb{R}^n \\ |\alpha^i| \leq 1}} \sum_{\substack{i_1, \dots, i_k \\ \text{distinct}}} \alpha_{i_1}^1 \cdots \alpha_{i_k}^k \sup_{t \in \mathcal{T}} \sup_{v^* \in \mathcal{B}_1^*} v^* \left( \sum_{\substack{I \in \mathcal{I}_{n,d-k} \\ i_1, \dots, i_k \notin I}} X_I t_{I \cup \{i_1, \dots, i_k\}} \right)$$

$$= \sup_{\substack{\alpha^1, \dots, \alpha^k \in \mathbb{R}^n \\ |\alpha^l| \leq 1}} \sum_{\substack{i_1, \dots, i_k \\ \text{distinct}}} \alpha_{i_1}^1 \cdots \alpha_{i_k}^k \sup_{t \in \mathcal{T}} \left\| \sum_{\substack{I \in \mathcal{I}_{n,d-k} \\ i_1, \dots, i_k \notin I}} X_I t_{I \cup \{i_1, \dots, i_k\}} \right\|.$$

Clearly  $\tilde{W}_k \geq W_k$  holds for all  $k \in [d]$ .

Concentration properties for functionals as in (9) have been studied for independent Rademacher variables  $X_1, \dots, X_n$  (i.e.  $\mathbb{P}(X_i = +1) = \mathbb{P}(X_i = -1) = 1/2$ ) and  $\mathcal{B} = \mathbb{R}$  in [16, Theorem 14] for all  $d \geq 2$ , and under certain technical assumptions in [2]. We prove deviation inequalities in the weakly dependent setting, and afterwards discuss how these compare to the particular result in [16]. It is easily possible to derive a similar result for functions of independent random variables (in the spirit of Theorem 1). As the corresponding proof is easily done by generalizing the proof of [16, Theorem 14], we omit it.

**Theorem 4** *Let  $X = (X_1, \dots, X_n)$  be a random vector in  $\mathbb{R}^n$  with support in  $[a, b]^n$  satisfying a  $\vartheta$ -LSI( $\sigma^2$ ). For  $f = f(X)$  as in (9) and all  $p \geq 2$  we have*

$$\|(f - \mathbb{E} f)_+\|_p \leq \sum_{j=1}^d (2\sigma^2(b-a)^2(p-3/2))^{j/2} \mathbb{E} W_j, \quad (11)$$

$$\|f - \mathbb{E} f\|_p \leq \sum_{j=1}^d (2(b-a)^2 p)^{j/2} \mathbb{E} \tilde{W}_j. \quad (12)$$

Consequently, for any  $t \geq 0$

$$\begin{aligned} \mathbb{P}(f - \mathbb{E} f \geq t) &\leq 2 \exp \left( -\frac{1}{2\sigma^2(b-a)^2} \min_{k=1, \dots, d} \left( \frac{t}{d \mathbb{E} W_k} \right)^{2/k} \right) \\ &\leq 2 \exp \left( -\frac{1}{2e^2 \sigma^2 (b-a)^2 d^2} \min_{k=1, \dots, d} \left( \frac{t}{\mathbb{E} W_k} \right)^{2/k} \right), \end{aligned} \quad (13)$$

and the same concentration inequalities hold with  $\mathbb{E} W_k$  replaced by  $\mathbb{E} \tilde{W}_k$ .

Note that independent Rademacher random variables satisfy a  $\vartheta$ -LSI(1) (see e.g. [26, Example 3.1] or [29, Theorem 3]). Therefore, we get back [16, Theorem 14] from Theorem 4 (with slightly different constants). However, Theorem 4 moreover includes many models with dependencies like those discussed in the introduction. Therefore, it may be considered as an extension of [16, Theorem 14] to dependent situations and moreover to coefficients from any Banach space  $\mathcal{B}$ . For instance, we may consider an Ising chaos as a natural generalization of a Rademacher chaos to a dependent situation. In this case, Theorem 4 yields that we still obtain basically the same concentration properties if the dependencies are sufficiently weak (which is guaranteed by the conditions outlined in the introduction).

To illustrate our results further, let us consider the case of  $d = 2$  separately. Here we write

$$T_1 := \mathbb{E} W_1 = \mathbb{E} \sup_{t \in \mathcal{T}} \sup_{v^* \in \mathcal{B}_1^*} \left( \sum_{i=1}^n \left( \sum_{j=1}^n X_j v^*(t_{ij}) \right)^2 \right)^{1/2}$$

$$T_2 := \mathbb{E} W_2 = \sup_{t \in \mathcal{T}} \sup_{v^* \in \mathcal{B}_1^*} \|(v^*(t_{ij}))_{i,j}\|_{\text{op}}.$$

The following corollary follows directly from Theorem 4.

**Corollary 1** *Assume that  $X = (X_1, \dots, X_n)$  satisfies a  $\mathfrak{d}$ -LSI( $\sigma^2$ ) and is supported in  $[a, b]^n$  and let  $f_{\mathcal{T}} = f_{\mathcal{T}}(X)$  be as in (9) with  $d = 2$ . We have for all  $t \geq 0$*

$$\mathbb{P}(f_{\mathcal{T}}(X) - \mathbb{E} f_{\mathcal{T}}(X) \geq t) \leq 2 \exp \left( -\frac{1}{60(b-a)^2 \sigma^2} \min \left( \frac{t^2}{T_1^2}, \frac{t}{T_2} \right) \right).$$

For the case of independent Rademacher variables, this recovers the upper tail in a famous result by Talagrand [52, Theorem 1.2] on concentration properties of quadratic forms in Banach spaces, which has also been done in [16]. Note that for  $\mathcal{B} = \mathbb{R}$ , we have

$$T_1 = \mathbb{E} \sup_{t \in \mathcal{T}} \left( \sum_{i=1}^n \left( \sum_{j=1}^n t_{ij} X_j \right)^2 \right)^{1/2}, \quad T_2 = \sup_{t \in \mathcal{T}} |T|_{\text{op}},$$

where  $T$  is the symmetric matrix with zero diagonal and entries  $T_{ij} = t_{ij}$  if  $i < j$ . If  $\mathcal{T}$  consists of a single element only, we have  $T_1 \leq |T|_{\text{HS}}$ . Hence, Corollary 1 can be regarded as a generalized Hanson–Wright inequality.

### 2.2 The Boolean Hypercube

The case of independent Rademacher random variables above can be interpreted in terms of quantities from Boolean analysis. Recall that any function  $f : \{-1, +1\}^n \rightarrow \mathbb{R}$  can be decomposed using the orthonormal *Fourier–Walsh basis* given by  $(x_S)_{S \subseteq [n]}$  for  $x_S := \prod_{i \in S} x_i$ . More precisely, we have

$$f(x) = \sum_{S \subseteq [n]} \hat{f}_S x_S = \sum_{j \in [n]} \sum_{S \subseteq [n]: |S|=j} \hat{f}_S x_S,$$

where the  $(\hat{f}_S)_{S \subseteq [n]}$  are given by  $\hat{f}_S = \int x_S f d\mu$  and are called the *Fourier coefficients* of  $f$ . For any  $j \in [n]$  we define the *Fourier weight of order  $j$*  as  $W_j(f) := \sum_{S \subseteq [n]: |S|=j} \hat{f}_S^2$ . It is clear that  $\|f\|_2^2 = \sum_{j=0}^n W_j(f)$ . The following multilevel concentration inequality can now be easily deduced.

**Proposition 2** *Let  $X_1, \dots, X_n$  be independent Rademacher random variables and let  $f : \{1, +1\}^n \rightarrow \mathbb{R}$  be a function given in the Fourier–Walsh basis as  $f(x) = \sum_{j=0}^d \hat{f}_j x_j$  for some  $d \in \mathbb{N}, d \leq n$ . For any  $t > 0$  we have*

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

*In other words, the event  $|f(X) - \mathbb{E} f(X)| \leq de \max_{j=1, \dots, d} (W_j(f)t^j)^{1/2}$  holds with probability at least  $1 - \exp(1 - t)$ .*

The literature on Boolean functions is vast, and a modern overview is given in [44]. Particularly for concentration results we may highlight [5, Theorem 1.4] (which in particular holds for Boolean functions), which we discuss further and partially generalize to dependent models in Sect. 2.4. Proposition 2 may be of interest due to the direct use of quantities from Fourier analysis. Finally, we should add that while many concentration results for Boolean functions like [5, Theorem 1.4] or also Proposition 2 are valid for functions whose Fourier–Walsh decomposition stops at some order  $d$ , Theorem 1 or Theorem 2 work for functions with Fourier–Walsh decomposition possibly up to order  $n$ .

### 2.3 Concentration Properties of $U$ -Statistics

Another application of Theorems 1 and 2 are concentration properties of so-called  $U$ -statistics which frequently arise in statistical theory. We refer to [24] for an excellent monograph. More recently, concentration inequalities for  $U$ -statistics have been considered in [1], [5, Sect. 3.1.2] and [10, Corollary 1.3].

Let  $\mathcal{Y} = \mathcal{X}^n$  and assume that  $X_1, \dots, X_n$  are either independent random variables, or the vector  $X = (X_1, \dots, X_n)$  satisfies a  $\vartheta$ -LSI( $\sigma^2$ ). Let  $h : \mathcal{X}^d \rightarrow \mathbb{R}$  be a measurable, symmetric function with  $h(X_{i_1}, \dots, X_{i_d}) \in L^\infty(\mathbb{P})$  for any  $i_1, \dots, i_d$ , and define  $B := \max_{i_1 \neq \dots \neq i_d} \|h(X_{i_1}, \dots, X_{i_d})\|_{L^\infty(\mathbb{P})}$ . We are interested in the concentration properties of the  $U$ -statistic with kernel  $h$ , i.e. of

$$f(X) = \sum_{i_1 \neq \dots \neq i_d} h(X_{i_1}, \dots, X_{i_d}). \tag{14}$$

**Proposition 3** *Let  $X = (X_1, \dots, X_n)$  be as above and  $f = f(X)$  be as in (14). There exists a constant  $C > 0$  (the same as in Theorems 1 and 2) such that for any  $t \geq 0$*

$$\mathbb{P}(|f - \mathbb{E} f| \geq Bt) \leq 2 \exp \left( -\frac{1}{C} \min_{k=1, \dots, d} \left( \frac{t}{\binom{d}{k} 2^k n^{d-k/2}} \right)^{2/k} \right)$$

and for some  $C = C(d)$

$$\mathbb{P}(n^{1/2-d} |f - \mathbb{E} f| \geq Bt) \leq 2 \exp \left( -\frac{1}{4C} \min \left( t^2, n^{1-1/d} t^{2/d} \right) \right). \tag{15}$$

The normalization  $n^{1/2-d}$  in (15) is of the right order for  $U$ -statistics generated by a non-degenerate kernel  $h$ , i.e.  $\text{Var}(\mathbb{E}_{X_1} h(X_1, \dots, X_d)) > 0$ , see [24, Remarks 4.2.5]. In the case of i.i.d. random variables  $X_1, \dots, X_n$  it states that

$$\frac{1}{n^{d-1/2}} \sum_{i_1 < \dots < i_d} h(X_{i_1}, \dots, X_{i_d}) \Rightarrow \mathcal{N}(0, d^2 \text{Var}(\mathbb{E}_{X_1} h(X_1, \dots, X_d)))$$

whenever  $\mathbb{E} h(X_1, \dots, X_d)^2 < \infty$ . Actually, (15) shows that for  $t \leq n^{1/2}$  we have sub-Gaussian tails for any finite  $n \in \mathbb{N}$  for bounded kernels  $h$ .

Proposition 3 improves upon our old result [10, Corollary 1.3] by providing multilevel tail bounds, thus yielding much finer estimates than the exponential moment bound given in the earlier paper. Moreover, it does not only address independent random variables but also weakly dependent models. As compared to the results from [1] and [5, Sect. 3.1.2], Proposition 3 covers different types of measures, since in [1] independent random variables were considered, while in [5] a Sobolev-type inequality was required, which does not include the various discrete models for which a  $\partial$ -LSI holds.

### 2.4 Polynomials and Subgraph Counts in Exponential Random Graph Models

Lastly, let us once again consider polynomial functions. The case of independent random variables has been treated in [5, Theorem 1.4] under more general conditions, so we omit it and concentrate on weakly dependent random variables.

Let  $f_d : \mathbb{R}^n \rightarrow \mathbb{R}$  be a multilinear (also called tetrahedral) polynomial of degree  $d$ , i.e. of the form

$$f_d(x) := \sum_{k=1}^d \sum_{1 \leq i_1 \neq \dots \neq i_k \leq n} a_{i_1 \dots i_k}^k x_{i_1} \cdots x_{i_k} \tag{16}$$

for symmetric  $k$ -tensors  $a^k$  with vanishing diagonal. Here, a  $k$ -tensor  $a^k$  is called symmetric, if  $a_{i_1 \dots i_k}^k = a_{\sigma(i_1) \dots \sigma(i_k)}^k$  for any permutation  $\sigma \in \mathcal{S}_k$ , and the (generalized) 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  $f$ .

For the next result, given some  $d \in \mathbb{N}$ , we recall a family of norms  $\|\cdot\|_{\mathcal{I}}$  on the space of  $d$ -tensors for each partition  $\mathcal{I} = \{I_1, \dots, I_k\}$  of  $\{1, \dots, d\}$ . The family  $\|\cdot\|_{\mathcal{I}}$  has been first introduced in [35], where it was used to prove two-sided estimates for  $L^p$  norms of Gaussian chaos, and the definitions given below agree with the ones from [35] as well as [3] and [5]. For brevity, write  $P_d$  for the set of all partitions of  $\{1, \dots, d\}$ . For each  $l = 1, \dots, k$  we denote by  $x^{(l)}$  a vector in  $\mathbb{R}^{n^{I_l}}$ , and for a  $d$ -tensor  $A = (a_{i_1, \dots, i_d})$  set

$$\|A\|_{\mathcal{I}} := \sup \left\{ \sum_{i_1 \dots i_d} a_{i_1 \dots i_d} \prod_{l=1}^k x_{i_{I_l}}^{(l)} : \sum_{i_{I_l}} (x_{i_{I_l}}^{(l)})^2 \leq 1 \text{ for all } l = 1, \dots, k \right\}.$$

We can regard the  $\|A\|_{\mathcal{I}}$  as a family of operator-type norms. In particular, it is easy to see that  $\|A\|_{\{1,\dots,d\}} = |A|_{\text{HS}}$  and  $\|A\|_{\{\{1\},\dots,\{d\}\}} = |A|_{\text{op}}$ .

The following result has been proven in the context of Ising models (in the Dobrushin uniqueness regime) in [3], and can easily be extended to any vector  $X$  satisfying a  $\vartheta$ -LSI( $\sigma^2$ ). By invoking the family of norms  $\|\cdot\|_{\mathcal{I}}$ , it provides a refinement of our general result for the special case of multilinear polynomials.

**Theorem 5** *Let  $X$  be a random vector supported in  $[-1, +1]^n$  and satisfying a  $\vartheta$ -LSI( $\sigma^2$ ), and  $f_d = f_d(X)$  be as in (16). There exists a constant  $C > 0$  depending on  $d$  only such that for all  $t \geq 0$*

$$\mathbb{P}(|*|f_d - \mathbb{E} f_d \geq t) \leq 2 \exp\left(-\frac{1}{C} \min_{k=1,\dots,d} \min_{\mathcal{I} \in P_k} \left(\frac{t}{\sigma^k \|\mathbb{E} \nabla^{(k)} f_d\|_{\mathcal{I}}}\right)^{2/|\mathcal{I}|}\right). \tag{17}$$

For illustration, let us once again consider the case of  $d = 2$ . In the notation of (16), we take  $a^1 = 0$  and  $a^2 = A$ , i.e.  $f_2(x) = x^T A x$  for a symmetric matrix  $A$  with vanishing diagonal. In this case, assuming the components of  $X$  to be centered (so the  $k = 1$  term vanishes), Theorem 5 reads

$$\mathbb{P}(|*|f_2 - \mathbb{E} f_2 \geq t) \leq 2 \exp\left(-\frac{1}{C} \min\left(\frac{t^2}{\sigma^4 |A|_{\text{HS}}^2}, \frac{t}{\sigma^2 |A|_{\text{op}}}\right)\right),$$

i.e. we obtain a Hanson–Wright inequality in this situation. For higher orders, we arrive at similar bounds. Altogether, for the class of multilinear polynomials, Theorem 5 yields finer bounds than Theorem 2 (by virtue of the large class of norms involved), though for  $d \geq 3$  explicit calculations of the norms involved can be difficult.

To point out one possible application, Theorem 5 can be used in the context of the exponential random graph model (ERGM). Let us briefly recall the definitions. Given  $s \in \mathbb{N}$  real numbers  $\beta_1, \dots, \beta_s$  and simple graphs  $G_1, \dots, G_s$  (with  $G_1$  being a single edge by convention), the ERGM with parameter  $\mathbf{fi} = (\beta_1, \dots, \beta_s, G_1, \dots, G_s)$  is a probability measure on the space of all graphs on  $n \in \mathbb{N}$  vertices given by the weight function  $\exp\left(\sum_{i=1}^s \beta_i n^{-|V_i|+2} N_{G_i}(x)\right)$ , where  $N_{G_i}(x)$  is the number of copies of  $G_i$  in the graph  $x$  and  $|V_i|$  is the number of vertices of  $G_i = (V_i, E_i)$ . For details, see [23,48]. One can think of the ERGM as an extension of the famous Erdős–Rényi model (which corresponds to the choice  $s = 1$ ) to account for dependencies between the edges.

By way of example we show concentration properties of the number of triangles  $T_3(X) = \sum_{\{e,f,g\} \in \mathcal{T}_3} X_e X_f X_g$  (where  $\mathcal{T}_3$  denotes the set of all three edges forming a triangle). To formulate our results, we need to recall the function  $\Phi_{\beta}(x) = \sum_{i=1}^s \beta_i |E_i| x^{|E_i|-1}$  which frequently appears in the discussion of the ERGM. Moreover, we set  $|\beta| := (|\beta_1|, \dots, |\beta_s|)$ . In the following corollary, the condition  $\frac{1}{2} \Phi'_{|\beta|}(1) < 1$  ensures weak dependence in the sense that a  $\vartheta$ -LSI holds. As outlined above, in comparison to earlier results like [48, Theorem 3.2], using Theorem 5 yields sharper tail estimates.

**Corollary 2** *Let  $X$  be an exponential random graph model with parameter  $\beta = (\beta_1, \dots, \beta_s, G_1, \dots, G_s)$  such that  $\frac{1}{2}\Phi'_{|\beta|}(1) < 1$ . There is a constant  $C(\beta)$  such that for all  $t \geq 0$*

$$\begin{aligned} & \mathbb{P}(|T_3 - \mathbb{E} T_3| \geq t) \\ & \leq 2 \exp\left(-\frac{1}{C(\beta)} \min\left(\frac{t^2}{\max(C_{S_2} n^4, C_{E} n^3, n^3)}, \frac{t}{\max(\sqrt{2}n, 2C_E n)}, \frac{t^{2/3}}{2}\right)\right). \end{aligned}$$

### 3 Concentration Inequalities Under Logarithmic Sobolev Inequalities: Proofs

In this section, we give the proofs of our main results. All of them work by first establishing a growth rate on the  $L^p$  norms of  $f - \mathbb{E} f$  which will then be iterated. For technical reasons, we need to introduce some auxiliary difference operators which are closely related to  $\mathfrak{h}$ . For  $i \in [n]$  let

$$\begin{aligned} \mathfrak{h}_i^+ f(X) &= \|(f(X) - T_i f(X))_+\|_{X'_i, \infty}, & \mathfrak{h}^+ f &= (\mathfrak{h}_1^+ f, \dots, \mathfrak{h}_n^+ f), \\ \mathfrak{h}_i^- f(X) &= \|(f(X) - T_i f(X))_-\|_{X'_i, \infty}, & \mathfrak{h}^- f &= (\mathfrak{h}_1^- f, \dots, \mathfrak{h}_n^- f), \end{aligned}$$

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

The  $L^p$  norm inequalities which form the core of our proofs can be found in [10, Theorem 2.3, Corollary 2.6] (building upon the earlier results in [16]). Note that as compared to [10], a different choice of normalization for  $\mathfrak{h}^\pm$  leads to slightly different constants.

**Theorem 6** *If  $X_1, \dots, X_n$  are independent random variables and  $f = f(X) \in L^\infty(\mathbb{P})$ , with the constant  $\kappa = \frac{\sqrt{e}}{2(\sqrt{e}-1)}$ , we have for any  $p \geq 2$ ,*

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

Consequently, this leads to

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

Furthermore, we need an auxiliary statement relating differences of consecutive order. In [10], we have proven that  $|\mathfrak{h}|\mathfrak{h}^{(d)} f|_{\text{HS}}| \leq |\mathfrak{h}^{(d+1)} f|_{\text{HS}}$ . Moreover, we explained that a similar estimate with the Hilbert–Schmidt replaced by operator norms cannot be true. As we will see next, the key step in order to be able to invoke operator norms nevertheless is to work with  $\mathfrak{h}^+$ .

Here, we need the following simple but crucial observation: if  $A$  is a  $d$ -tensor, the supremum in the definition of  $|A|_{\text{op}}$  is attained, and if  $A$  is a non-negative tensor (i. e.  $A_{i_1 \dots i_d} \geq 0$  for all  $i_1, \dots, i_d$ ), the maximizing vectors  $\tilde{v}^1, \dots, \tilde{v}^d$  can be chosen to have all positive entries. Indeed, since  $\tilde{v}_1^1 \dots \tilde{v}_{i_d}^d \leq |\tilde{v}_1^1 \dots \tilde{v}_{i_d}^d|$ , we can define  $|\tilde{v}|^j$  by taking the absolute value element-wise.

**Lemma 1** For any  $d \geq 2$

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

**Proof** We have

$$\begin{aligned} |\mathfrak{h}^+ |\mathfrak{h}^{(d-1)} f|_{op}|^2 &= \sum_{i=1}^n \left\| \left( |\mathfrak{h}^{(d-1)} f|_{op} - |\mathfrak{h}^{(d-1)} T_i f|_{op} \right)_+ \right\|_{i,\infty}^2 \\ &= \sum_{i=1}^n \left\| \left( \sup_{v^j} \langle v^1 \dots v^{d-1}, \mathfrak{h}^{(d-1)} f \rangle - \sup_{v^j} \langle v^1 \dots v^{d-1}, \mathfrak{h}^{(d-1)} T_i f \rangle \right)_+ \right\|_{i,\infty}^2 \\ &\leq \sum_{i=1}^n \left\| \left( \langle \tilde{v}^1 \dots \tilde{v}^{d-1}, \mathfrak{h}^{(d-1)} f - \mathfrak{h}^{(d-1)} T_i f \rangle \right)_+ \right\|_{i,\infty}^2 \\ &\leq \sum_{i=1}^n \left\| \sum_{i_1, \dots, i_{d-1}} \tilde{v}_{i_1}^1 \dots \tilde{v}_{i_{d-1}}^{d-1} \left\| (\text{Id} - T_i) \prod_{j=1}^{d-1} (\text{Id} - T_{i_j}) f \right\|_{i_1 \dots i_{d-1}, \infty} \right\|_{i,\infty}^2 \\ &\leq \sum_{i=1}^n \left( \sum_{i_1, \dots, i_{d-1}} \tilde{v}_{i_1}^1 \dots \tilde{v}_{i_{d-1}}^{d-1} \mathfrak{h}_{ii_1 \dots i_{d-1}} f \right)^2 \\ &= \left( \sup_{v^d: |v^d| \leq 1} \sum_{i_d=1}^n \sum_{i_1, \dots, i_{d-1}} \tilde{v}_{i_1}^1 \dots \tilde{v}_{i_{d-1}}^{d-1} v_{i_d}^d \mathfrak{h}_{i_1 \dots i_d} f \right)^2 \\ &\leq \left( \sup_{v^1, \dots, v^d: |v^j| \leq 1} \sum_{i_1, \dots, i_d} v_{i_1}^1 \dots v_{i_d}^d \mathfrak{h}_{i_1 \dots i_d} f \right)^2 \\ &= |\mathfrak{h}^{(d)} f|_{op}^2 \end{aligned}$$

Here, in the first inequality we insert the vectors  $\tilde{v}^1, \dots, \tilde{v}^{d-1}$  maximizing the supremum and use the monotonicity of  $x \mapsto x_+$ , and the second and third inequality follow from the triangle inequality. Taking the square root yields the claim.  $\square$

As a final step, we need to establish a connection between  $L^p$  norm estimates and multilevel concentration inequalities. This is given by the following proposition, which was proven in [1, Theorem 7] and [5, Theorem 3.3]. We state it in the form given in [48, Proof of Theorem 3.6] with slight modifications.

**Proposition 4** Assume that a random variable  $f$  satisfies for any  $p \geq 2$  and some constants  $C_1, \dots, C_d \geq 0$   $\|f - \mathbb{E} f\|_p \leq \sum_{k=1}^d C_k (p - s)^{k/2}$  for some  $s \in [0, 2)$ , and let  $L := |\{l : C_l > 0\}|$ . For any  $t \geq 0$  we have

$$\mathbb{P}(|f - \mathbb{E} f| \geq t) \leq 2 \exp \left( - \min \left( \frac{\log(2)}{2 - s}, 1 \right) \min_{k=1, \dots, d} \left( \frac{t}{L C_k} \right)^{2/k} \right).$$

We will not give a proof of Proposition 4 and refer to the aforementioned works. However, the proof is almost identical to the proof of Proposition 2. The two important



cases will be  $s = 0$  (for independent random variables) as well as  $s = 3/2$  (in the weakly dependent setting).

The proof of Theorem 1 is now easily completed.

**Proof of Theorem 1** Since  $X_1, \dots, X_n$  are independent, Theorem 6 yields

$$\|f - \mathbb{E} f\|_p \leq (8\kappa p)^{1/2} \|\mathfrak{h}f\|_p \leq (8\kappa p)^{1/2} \|\mathfrak{h}f\|_{\text{op},1} + (8\kappa p)^{1/2} \|(|\mathfrak{h}f| - \mathbb{E} |\mathfrak{h}f|)_+\|_p$$

where we have used that for any positive random variable  $W$

$$\|W\|_p \leq \mathbb{E} W + \|(W - \mathbb{E} W)_+\|_p. \tag{18}$$

The second term on the right hand side can now be estimated using Theorem 6 again, which in combination with Lemma 1 gives

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

This can be easily iterated to obtain for any  $d \in \mathbb{N}$

$$\|f - \mathbb{E} f\|_p \leq \sum_{j=1}^{d-1} (8\kappa p)^{j/2} \|\mathfrak{h}^{(j)} f\|_{\text{op},1} + (8\kappa p)^{d/2} \|\mathfrak{h}^{(d)} f\|_{\text{op},\infty}.$$

Now it remains to apply Proposition 4. □

To prove Theorem 2, we shall require the following proposition, which is proven in [28, Proposition 2.4], building upon arguments established in [6]. (Note that the definition of  $\mathfrak{h}$  there differed by a factor of  $\sqrt{2}$ .) The estimate (20) does not appear therein, but is an easy modification of the proof.

**Proposition 5** *Let  $\mu$  be a measure on a product of Polish spaces satisfying a  $\mathfrak{d}$ -LSI( $\sigma^2$ ). Then, for any  $f \in L^\infty(\mu)$  and any  $p \geq 2$  we have*

$$\|f - \mathbb{E} f\|_p \leq (2\sigma^2(p - 3/2))^{1/2} \|\mathfrak{d}f\|_p \leq (\sigma^2(p - 3/2)/2)^{1/2} \|\mathfrak{h}f\|_p \tag{19}$$

and

$$\|(f - \mathbb{E} f)_+\|_p \leq (2\sigma^2(p - 3/2))^{1/2} \|\mathfrak{h}^+ f\|_p. \tag{20}$$

**Proof of Theorem 2** The proof is very similar to the proof of Theorem 1. In the first step, using (19) leads to

$$\|f - \mathbb{E} f\|_p \leq (2\sigma^2(p - 3/2))^{1/2} \|\mathfrak{h}f\|_{\text{op},1} + (2\sigma^2(p - 3/2))^{1/2} \|(|\mathfrak{h}f| - \mathbb{E} |\mathfrak{h}f|)_+\|_p.$$

Equation (20) can be used to estimate the second term on the right-hand side. So, for any  $d \in \mathbb{N}$  we have by an iteration

$$\|f - \mathbb{E} f\|_p \leq \sum_{j=1}^{d-1} (2\sigma^2(p - 3/2))^{j/2} \|\mathfrak{h}^{(j)} f\|_{\text{op},1} + (2\sigma^2(p - 3/2))^{d/2} \|\mathfrak{h}^{(d)} f\|_{\text{op},\infty}$$

Again we can apply Proposition 4 to obtain the concentration inequality. □

To prove Theorem 3 we shall need the following lemma.

**Lemma 2** *Let  $(\mathcal{B}, \|\cdot\|)$  be a Banach space and  $\mathcal{F}$  a family of uniformly norm-bounded,  $\mathcal{B}$ -valued, measurable functions and set  $g(X) = \sup_{f \in \mathcal{F}} \|f(X)\|$ . We have*

$$|\mathfrak{h}^+ g(X)| \leq \sup_{f \in \mathcal{F}} |\mathfrak{h}^+ \|f\|(X)|.$$

**Proof** Fix an  $X \in \mathcal{Y}$  and choose for any  $\varepsilon > 0$  a function  $f_\varepsilon$  such that  $\|f_\varepsilon(X)\| \geq \sup_{f \in \mathcal{F}} \|f(X)\| - \varepsilon$ . This yields

$$\begin{aligned} \mathfrak{h}_i^+ g(X) &= \sup_{x'_i} (\sup_{f \in \mathcal{F}} \|f(X)\| - \sup_{f \in \mathcal{F}} \|f(X_{i^c}, x'_i)\|)_+ \leq \sup_{x'_i} (\|f_\varepsilon(X)\| + \varepsilon - \|f_\varepsilon(X_{i^c}, x'_i)\|)_+ \\ &\leq \sup_{x'_i} (\|f_\varepsilon(X)\| - \|f_\varepsilon(X_{i^c}, x'_i)\|)_+ + \varepsilon = \mathfrak{h}_i^+ \|f_\varepsilon\|(X) + \varepsilon, \end{aligned}$$

where the first inequality follows by monotonicity of  $x \mapsto x_+$  and the second one is a consequence of  $(a + b - c)_+ \leq (a - c)_+ + b$  for  $a, b, c \geq 0$ . Thus, we have

$$\begin{aligned} |\mathfrak{h}^+ g(X)| &= \left( \sum_{i=1}^n \mathfrak{h}_i^+ g(X)^2 \right)^{1/2} \leq \left( \sum_{i=1}^n (\mathfrak{h}_i^+ \|f_\varepsilon\|(X) + \varepsilon)^2 \right)^{1/2} = |\mathfrak{h}^+ \|f_\varepsilon\|(X) + \varepsilon(1, \dots, 1)| \\ &\leq |\mathfrak{h}^+ \|f_\varepsilon\|(X)| + \sqrt{n}\varepsilon \leq \sup_{f \in \mathcal{F}} |\mathfrak{h}^+ \|f\|(X)| + \sqrt{n}\varepsilon. \end{aligned}$$

Taking the limit  $\varepsilon \rightarrow 0$  yields the claim. □

**Proof of Theorem 3** Note that in the real-valued case, the estimate  $\mathfrak{h}_i^+ |f| \leq \mathfrak{h}_i f$  holds. For brevity, let  $s = 3/2$ . Using this in combination with Proposition 5 and Lemma 2 yields

$$\begin{aligned} \|(g - \mathbb{E} g)_+\|_p &\leq (2\sigma^2(p - s))^{1/2} \|\mathfrak{h}^+ g\|_p \leq (2\sigma^2(p - s))^{1/2} \sup_{f \in \mathcal{F}} |\mathfrak{h} f| \|g\|_p \\ &\leq (2\sigma^2(p - s))^{1/2} \mathbb{E} W_1 + (2\sigma^2(p - s))^{1/2} \|(W_1 - \mathbb{E} W_1)_+\|_p. \end{aligned}$$

We can apply Proposition 5 again on the right hand side, which gives

$$\|(g - \mathbb{E} g)_+\|_p \leq (2\sigma^2(p - s))^{1/2} \mathbb{E} W_1 + (2\sigma^2(p - s)) \|\mathfrak{h}^+ W_1\|_p.$$

A combination of Lemmas 1 and 2 shows that  $|\mathfrak{h}^+ W_j| \leq W_{j+1}$ , and so by an iteration we obtain

$$\|(g - \mathbb{E} g)_+\|_p \leq \sum_{j=1}^{d-1} (2\sigma^2(p - s))^{j/2} \mathbb{E} W_j + (2\sigma^2(p - s))^{d/2} \|W_d\|_\infty.$$

In the case of independent random variables, we replace the first step using Theorem 6. Here,  $2\sigma^2 = 2\kappa$  and  $s = 0$ . □

**Proof of Proposition 1** The proof shares some similarities with the proof of Lemma 2. Since  $X$  satisfies a  $\mathfrak{d}$ -LSI( $\sigma^2$ ), we have for any  $p \geq 2$

$$\|(g - \mathbb{E} g)_+\|_p \leq (2\sigma^2(p - 3/2))^{1/2} \|\mathfrak{h}^+ g\|_p.$$

Moreover, for any  $i \in [n]$  and  $x \in \mathcal{Y}$ , if a maximizer  $\tilde{f}$  of  $\sup_{f \in \mathcal{F}} |\sum_{j=1}^n f(x_j)|$  exists, we obtain

$$\begin{aligned} \mathfrak{h}_i^+ g(x)^2 &= \sup_{x'_i} \left( \sup_{f \in \mathcal{F}} |f(X)| - \sup_{f \in \mathcal{F}} |f(X_{i^c}, x'_i)| \right)_+^2 \\ &\leq \sup_{x'_i} \left( |\tilde{f}(X)| - |\tilde{f}(X_{i^c}, x'_i)| \right)_+^2 \leq c(\tilde{f})^2 \leq \sup_{f \in \mathcal{F}} c(f)^2. \end{aligned}$$

If a maximizer  $\tilde{f}$  does not exist, these estimates remain valid by an approximation argument as in the proof of Lemma 2. Consequently, we have  $\|(g - \mathbb{E} g)_+\|_p \leq (2\sigma^2(p - 3/2)n \sup_{f \in \mathcal{F}} c(f)^2)^{1/2}$ . The claim now follows from Proposition 4. □

### 4 Suprema of Chaos, U-statistics and Polynomials: Proofs

**Proof of Theorem 4** Let us first consider the case that  $X$  satisfies a  $\mathfrak{d}$ -LSI( $\sigma^2$ ). Recall that we have by (20)

$$\|(f - \mathbb{E} f)_+\|_p \leq (2\sigma^2(p - 3/2))^{1/2} \|\mathfrak{h}^+ f\|_p.$$

We shall make use of the pointwise inequality  $|\mathfrak{h}^+ f| \leq (b - a)W_1$ . To see this, let  $(\tilde{t}, \tilde{v}^*)$  be the tuple satisfying  $\sup_{t \in \mathcal{T}} \sup_{v^* \in \mathcal{B}_1^*} v^*(\sum_{I \in \mathcal{I}_{n,d}} X_I t_I) = \tilde{v}^*(\sum_{I \in \mathcal{I}_{n,d}} X_I \tilde{t}_I)$ . We have

$$\begin{aligned} |\mathfrak{h}^+ f(X)|^2 &= \sum_{i=1}^n \sup_{x'_i} \left( \sup_{t, v^*} v^* \left( \sum_{I \in \mathcal{I}_{n,d}} X_I t_I \right) - \sup_{t, v^*} v^* \left( \sum_{I \in \mathcal{I}_{n,d}} (X_{i^c}, x'_i)_I t_I \right) \right)_+^2 \\ &\leq \sum_{i=1}^n \sup_{x'_i} \left( (X_i - x'_i) \sum_{\substack{I \in \mathcal{I}_{n,d-1} \\ i \notin I}} \tilde{v}^*(X_I \tilde{t}_{I \cup \{i\}}) \right)^2 \end{aligned}$$

$$\begin{aligned}
 &\leq (b - a)^2 \sum_{i=1}^n \left( \tilde{v}^* \left( \sum_{\substack{I \in \mathcal{I}_{n,d-1} \\ i \notin I}} X_I \tilde{t}_{I \cup \{i\}} \right) \right)^2 \\
 &= (b - a)^2 \sup_{\alpha^{(1)}: \|\alpha^{(1)}\| \leq 1} \left( \tilde{v}^* \left( \sum_{i=1}^n \alpha_i^{(1)} \sum_{I \in \mathcal{I}_{n,d-1}: i \notin I} X_I \tilde{t}_{I \cup \{i\}} \right) \right)^2 \\
 &\leq (b - a)^2 \left( \sup_{t, v^*} \sup_{\alpha^{(1)}: \|\alpha^{(1)}\| \leq 1} v^* \left( \sum_{i=1}^n \alpha_i^{(1)} \sum_{I \in \mathcal{I}_{n,d}: i \notin I} X_I t_{I \cup \{i\}} \right) \right)^2 \\
 &= (b - a)^2 W_1^2,
 \end{aligned}$$

proving the first part. Consequently,

$$\| (f - \mathbb{E} f)_+ \|_p \leq (2\sigma^2(b - a)^2(p - 3/2))^{1/2} (\mathbb{E} W_1 + \|(W_1 - \mathbb{E} W_1)_+\|_p).$$

As in [16], this can now be iterated, i. e. we have for any  $k \in \{1, \dots, d - 1\}$   $|\mathfrak{h}^+ W_k| \leq (b - a)W_{k+1}$ . Here, we may argue as above, where the only difference is to choose  $(\tilde{t}, \tilde{v}^*)$  and  $\tilde{\alpha}^{(1)}, \dots, \tilde{\alpha}^{(k)}$  which maximize  $W_k$ . This finally leads to

$$\| f - \mathbb{E} f \|_p \leq \sum_{j=1}^d (2\sigma^2(b - a)^2(p - 3/2))^{j/2} \mathbb{E} W_j,$$

using that  $W_d$  is constant. This proves (11). The same arguments are also valid without a  $\mathfrak{d}$ -LSI( $\sigma^2$ ) property, if one considers  $\|(f - \mathbb{E} f)_+\|_p$  and applies Theorem 6 instead.

Lastly, to prove (12), let us first consider why we cannot argue as before. Note that the argument heavily relies on the positive part of the difference operator  $\mathfrak{h}^+$ , which allows us to choose the maximizers  $t_1, \dots, t_n$  independent of  $i \in [n]$ . This is no longer possible for the concentration inequality. Here, Theorem 6 yields

$$\begin{aligned}
 \| f - \mathbb{E} f \|_p &\leq (\sigma^2 p)^{1/2} \|\mathfrak{h} f\|_p \\
 \| (f - \mathbb{E} f)_+ \|_p &\leq (\sigma^2 p)^{1/2} \|\mathfrak{h}^+ f\|_p.
 \end{aligned}$$

Thus, this argument fails if we try to use these inequalities. However, we can rewrite  $\mathfrak{h}_i f(x) = \sup_{x'_i, x''_i} (f(x_{i^c}, x'_i) - f(x_{i^c}, x''_i))_+ = \sup_{x'_i} \mathfrak{h}_i^+ f(x_{i^c}, x'_i)$ , where the sup is to be understood with respect to the support of  $X'_i$ . As a consequence, we have for each fixed  $i \in [n]$  (again choosing  $\tilde{t}$  by maximizing the first summand in the brackets)

$$\begin{aligned}
 \mathfrak{h}_i f(x)^2 &= \sup_{x'_i} \sup_{x''_i} \left( \sup_{t \in \mathcal{T}} \left\| \sum_{I \in \mathcal{I}_{n,d}} (X_{i^c}, x'_i)_{I \cup \{i\}} \right\| - \sup_{t \in \mathcal{T}} \left\| \sum_{I \in \mathcal{I}_{n,d}} (X_{i^c}, x''_i)_{I \cup \{i\}} \right\| \right)_+^2 \\
 &\leq \sup_{x'_i} \sup_{x''_i} \left\| (x'_i - x''_i) \sum_{I \in \mathcal{I}_{n,d-1}: i \notin I} X_I \tilde{t}_{I \cup \{i\}} \right\|^2
 \end{aligned}$$

$$\begin{aligned} &\leq \sup_{x'_i, x''_i} |x'_i - x''_i|^2 \sup_{t \in \mathcal{T}} \left\| \sum_{I \in \mathcal{I}_{n,d-1}; i \notin I} X_I t_{I \cup \{i\}} \right\|^2 \\ &\leq (b - a)^2 \sup_{t \in \mathcal{T}} \left\| \sum_{I \in \mathcal{I}_{n,d-1}; i \notin I} X_I t_{I \cup \{i\}} \right\|^2. \end{aligned}$$

This implies

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

The proof is now completed as using the same arguments as in the first part, with  $W_k$  replaced by  $\widetilde{W}_k$ . The same argument is valid for  $X$  satisfying a  $\mathfrak{d}$ -LSI( $\sigma^2$ ).  $\square$

**Proof of Proposition 2** The proposition can be proven using a similar technique as before, since the Hilbert–Schmidt norms of higher order difference act as Fourier projections. We choose to take an alternate route as follows. The proof of [44, Theorem 9.21] shows that for any  $f$  with degree at most  $d$  and any  $p \geq 2$

$$\|f(X) - \mathbb{E} f(X)\|_p \leq \sum_{j=1}^d (p - 1)^{j/2} W_j(f)^{1/2}. \tag{21}$$

First off, by Chebyshev’s inequality we have for any  $p \geq 1$

$$\mathbb{P}(|f(X) - \mathbb{E} f(X)| \geq e \|f(X) - \mathbb{E} f(X)\|_p) \leq \exp(-p).$$

We want to apply this to a  $t$ -dependent parameter  $p$  given by the function

$$\eta_f(t) := 1 + \min_{j=1, \dots, d} \left( \frac{t}{deW_j(f)^{1/2}} \right)^{2/j}.$$

If  $\eta_f(t) \geq 2$ , (21) yields  $e \|f(X) - \mathbb{E} f(X)\|_{\eta_f(t)} \leq t$ , which combined with the trivial estimate  $\mathbb{P}(\cdot) \leq 1$  gives

$$\mathbb{P}(|f(X) - \mathbb{E} f(X)| \geq t) \leq e^2 \exp(-\eta_f(t)) = \exp\left(1 - \min_{j=1, \dots, d} \left( \frac{t}{deW_j(f)^{1/2}} \right)^{2/j}\right)$$

as claimed.  $\square$

**Proof of Proposition 3** We apply Theorems 1 and 2 in the respective cases. To this end, we make use of the general bound  $\|\mathfrak{h}^{(k)} f\|_{\text{op},1} \leq \|\mathfrak{h}^{(k)} f\|_{\text{HS},\infty}$  for  $k \in [d]$ . For any distinct  $j_1, \dots, j_k$  write  $\|\cdot\| = \|\cdot\|_{j_1, \dots, j_k, \infty}$ , so that

$$\mathfrak{h}_{j_1, \dots, j_k} f = \left\| f + \sum_{l=1}^k (-1)^l \sum_{1 \leq s_1 < \dots < s_l \leq k} T_{j_{s_1} \dots j_{s_l}} f \right\|$$

$$\begin{aligned}
 &= \left\| \sum_{i_1 \neq \dots \neq i_d} (h(X_{i_1}, \dots, X_{i_d}) + \sum_{l=1}^k (-1)^l \sum_{s_1 < \dots < s_l} T_{j_{s_1} \dots j_{s_l}} h(X_{i_1}, \dots, X_{i_d})) \right\| \\
 &=: \left\| \sum_{i_1 \neq \dots \neq i_d} S_{i_1, \dots, i_d}(h, X) \right\|.
 \end{aligned}$$

Now it is easy to see that  $S_{i_1, \dots, i_d}(h, X) = 0$  unless  $\{j_1, \dots, j_k\} \subset \{i_1, \dots, i_d\}$  (for example, this follows if one writes the sum inside the norm as  $\prod_{i=1}^k (\text{Id} - T_{j_i})f$ ), and in these cases one can upper bound the supremum by  $2^k B$ , from which we infer

$$\mathfrak{h}_{j_1, \dots, j_k} f \leq \binom{d}{k} 2^k B(n - k) \cdots (n - d + 1) \leq \binom{d}{k} 2^k B n^{d-k}.$$

Consequently, this leads to

$$\| \mathfrak{h}^{(k)} f \|_{\text{HS}, \infty} \leq \binom{d}{k} 2^k B n^{d-k} n^{k/2} = \binom{d}{k} 2^k B n^{d-k/2}.$$

Thus, an application of Theorems 1 or 2, respectively, yields for any  $t \geq 0$  and for  $C$  as given therein

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

For the second part, choose  $t = B n^{d-1/2} \tilde{t}$  for  $\tilde{t} > 0$  to obtain

$$\mathbb{P}\left(n^{1/2-d} B^{-1} |f(X) - \mathbb{E} f(X)| \geq \tilde{t}\right) \leq 2 \exp\left(-\frac{1}{4C_d} \min_{k=1, \dots, d} n^{\frac{k-1}{k}} \tilde{t}^{2/k}\right).$$

A short calculation shows that the minimum is attained for  $k = 1$  in the range  $t \leq n^{1/2}$  and for  $k = d$  otherwise, i. e.

$$\mathbb{P}\left(n^{1/2-d} B^{-1} |f(X) - \mathbb{E} f(X)| \geq t\right) \leq 2 \exp\left(-\frac{1}{4C_d} \min(t^2, n^{1-1/d} t^{2/d})\right). \tag{22}$$

□

**Proof of Theorem 5** We give a sketch of the proof only and refer to [3, Proof of Theorem 2.2] for details. Recall that by (19) we have the inequality

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

Using the arguments and notations from [3, Proof of Theorem 2.2] leads to

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

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 [35] yields

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

The rest now follows as in the previous proofs. □

**Proof of Corollary 2** In [48] the authors have proven that  $\frac{1}{2} \Phi'_{|\beta|}(1) < 1$  implies a  $\vartheta$ -LSI( $\sigma^2$ ) for  $\mu_\beta$  with a constant depending on the parameter  $\beta$  only. Thus, it remains to bound the norms in (17). Note that due to the structure of the exponential random graph model, the expectations of  $\mathbb{E} X_G$  and  $\mathbb{E} X_H$  are equal whenever  $G$  and  $H$  are isomorphic. Thus, we define  $C_{S_2} := \mathbb{E} X_{S_2}$  (where  $S_2$  is a 2-star) and  $C_E = \mathbb{E} X_e$ .

The Euclidean norms can be easily bounded:

$$\begin{aligned} |\mathbb{E} \nabla f(X)| &= \left( \sum_{e \in \mathcal{I}_n} ((n-2)C_{S_2})^2 \right)^{1/2} \leq C_{S_2} n^2 \\ |\mathbb{E} \nabla^{(2)} f(X)| &= \left( \sum_{e, f: e \cap f \neq \emptyset} (C_E)^2 \right)^{1/2} \leq C_E n^{3/2} \\ |\mathbb{E} \nabla^{(3)} f(X)| &= \left( \sum_{\{e, f, g\} \in \mathcal{T}_3} 1 \right)^{1/2} = n^{3/2}, \end{aligned}$$

and it remains to estimate the three remaining norms. However, in [5, Sect. 5.1], the authors given estimates for such norms in the Erdős–Rényi case, and it is easy to adapt these to any model with the property that  $\mathbb{E} X_G$  depends only on the isomorphism class of  $G$  (in the complete graph). Particularly, due to the structure of the exponential random graph models, this is true in this setting as well. This gives

$$|\mathbb{E} \nabla^{(3)} f(X)|_{\text{op}} \leq 2^{3/2}, \quad |\mathbb{E} \nabla^{(3)} f(X)|_{\{1,2\}\{3\}} \leq \sqrt{2n}, \quad |\mathbb{E} \nabla^{(2)} f(X)|_{\text{op}} \leq 2C_E n.$$

Inserting these estimates into (17) finishes the proof. □

### 5 Logarithmic Sobolev Inequalities and Difference Operators

To conclude this paper, we discuss the LSI property (2) for different choices of difference operators  $\Gamma$ . Here, we always assume that the probability measure  $\mu$  is defined on a product of Polish spaces  $\mathcal{Y} = \otimes_{i=1}^n \mathcal{X}_i$  with product Borel  $\sigma$ -algebra  $\mathcal{A} = \mathcal{B}(\otimes_{i=1}^n \mathcal{X}_i)$ .

In this situation, we can make use of the disintegration theorem on Polish spaces (see [7, Theorem 5.3.1] and [25, Chapter III]): If  $\mu$  is a measure on  $\mathcal{Y}$ , then for each  $i \in \{1, \dots, n\}$  we can decompose  $\mu$  using the marginal measure  $\mu_{i^c}$  (as a measure on  $\otimes_{j \neq i} \mathcal{X}_j$ ) and a conditional measure on  $\mathcal{X}_i$ , which we denote by  $\mu(\cdot \mid x_{i^c})$ . More precisely, for any  $A \in \mathcal{A}$  we have  $\mu(A) = \int_{\otimes_{j \neq i} \mathcal{X}_j} \int_{\mathcal{X}_i} \mathbb{1}_A(x_{i^c}, x_i) d\mu(x_i \mid x_{i^c}) d\mu_{i^c}(x_{i^c})$ .

For finite spaces,  $\mu(\cdot \mid x_{i^c})$  is just the ordinary conditional measure as used in the definition of the difference operator  $\mathfrak{d}$ . Note that the definition of  $\mathfrak{d}$  can in principle be rewritten for products of arbitrary Polish spaces. However, our first result shows that the  $\mathfrak{d}$ -LSI property in fact requires the underlying space to be finite. More precisely, we say that  $\mu$  has finite support if there is no sequence of sets  $A_n \in \mathcal{A}$  with  $\mu(A_n) > 0$  for any  $n$  and  $\mu(A_n) \rightarrow 0$ .

**Proposition 6** *Let  $\mathcal{Y} = \otimes_{i=1}^n \mathcal{X}_i$  be a product of Polish spaces, and let  $\mu$  be a probability measure on  $\mathcal{Y}$ . If  $\mu$  satisfies a  $\mathfrak{d}$ -LSI, then  $\mu$  has finite support. Moreover, if  $\mu$  is a product probability measure, then  $\mu$  satisfies a  $\mathfrak{d}$ -LSI iff  $\mu$  has finite support.*

**Proof** First assume  $\mu$  does not have finite support, i.e. there is a sequence  $A_n \in \mathcal{A}$  with  $\mu(A_n) \rightarrow 0$ . Choosing  $f_n := \mathbb{1}_{A_n} \in L^\infty(\mu)$  and assuming a  $\mathfrak{d}$ -LSI( $\sigma^2$ ) holds, we obtain

$$\mu(A_n) \log(1/\mu(A_n)) = \text{Ent}_\mu(f_n^2) \leq 2\sigma^2 \int (\mathfrak{d} f_n)^2 d\mu = 2\sigma^2 \mu(A_n)(1 - \mu(A_n)). \tag{23}$$

This easily leads to a contradiction.

On the other hand, let  $\mu$  be a product probability measure with finite support. By tensorization, it suffices to consider  $n = 1$ , and we may moreover assume  $\mathcal{Y}$  to have finitely many elements only. Then, by [12, Remark 6.6],  $\mu$  satisfies a  $\mathfrak{d}$ -LSI( $\sigma^2$ ) with  $\sigma^2 \leq C \log(1/\min_{y:\mu(y)>0} \mu(y))$ , which finishes the proof.  $\square$

In fact, Proposition 6 can be adapted to the difference operator  $\mathfrak{h}^+$  as well. To see this, note that that (23) can easily be rewritten for the difference operator  $\mathfrak{h}^+$  (with only minor changes) and  $\int |\mathfrak{d} f|^2 d\mu \leq \int |\mathfrak{h}^+ f|^2 d\mu$ . In particular, the  $\mathfrak{d}$ - and  $\mathfrak{h}^+$ -LSI properties are not essentially different.

The situation drastically changes if we consider  $\mathfrak{h}$ -LSIs instead. Here, a sufficient condition for the  $\mathfrak{h}$ -LSI property to hold is that the measure  $\mu$  satisfies an approximate tensorization (AT) property. As a consequence, for product probability measures, satisfying an  $\mathfrak{h}$ -LSI is in fact a universal property.

**Theorem 7** *Let  $\mathcal{Y} = \otimes_{i=1}^n \mathcal{X}_i$  be a product of Polish spaces, and let  $\mu$  be a probability measure on  $\mathcal{Y}$ . If  $\mu$  satisfies an approximate tensorization property*

$$\text{Ent}_\mu(f^2) \leq C \sum_{i=1}^n \int \text{Ent}_{\mu(\cdot \mid x_{i^c})}(f^2(x_{i^c}, \cdot)) d\mu_{i^c}(x_{i^c}), \tag{24}$$

*then  $\mu$  also satisfies an  $\mathfrak{h}$ -LSI( $C$ ). In particular, any product probability measure satisfies an  $\mathfrak{h}$ -LSI(1).*



To the best of our knowledge, Theorem 7 is new. For product measures, it might be compared to the Efron–Stein inequality (see e. g. [27,50]) which establishes the tensorization property for the variance, and can be regarded as a universal Poincaré inequality with respect to  $\mathfrak{d}$  (see e. g. [10] for such an interpretation). However, note that Theorem 7 (i. e. more precisely the  $\mathfrak{h}$ -LSI(1) for product measures) does not imply the Efron–Stein inequality, as the difference operator is  $\mathfrak{h}$  instead of  $\mathfrak{d}$ . Unfortunately, as Proposition 6 demonstrates, there is no “entropy version” of the Efron–Stein inequality of the form  $\text{Ent}_\mu(f^2) \leq C \mathbb{E}_\mu |\mathfrak{d}f|^2$  (for any product probability measure  $\mu$  and some universal constant  $C$ ).

As by Theorem 7, any set of independent random variables  $X_1, \dots, X_n$  satisfies an  $\mathfrak{h}$ -LSI(1), it might be tempting to regard Theorem 1 as an  $\mathfrak{h}$ -LSI analogue of Theorem 2. However, it seems that it is not possible to use the entropy method based on  $\mathfrak{h}$ -LSIs, so that this interpretation is not fully accurate. More precisely, Theorem 7 cannot be used to estimate the growth of  $L^p$  norms as in the setting of a  $\mathfrak{d}$ -LSI( $\sigma^2$ ). Indeed, it is impossible to prove the required moment inequalities

$$\|f - \mathbb{E} f\|_q \leq (\sigma^2 q)^{1/2} \|\mathfrak{h}f\|_q \tag{25}$$

under an  $\mathfrak{h}$ -LSI( $\sigma^2$ ). For example, the measure  $\mu_p = p\delta_1 + (1 - p)\delta_0$  satisfies  $\mathfrak{h}$ -LSI( $\sigma_p^2$ ) with  $\sigma_p^2 \sim p(1 - p) \log(1/p)$  (for  $p \rightarrow 0$ ), so that (25) would imply for  $f(x) = x$  an upper bound on the Orlicz norm associated to  $\Psi_2(x) = e^{x^2} - 1$

$$\|f - \mathbb{E} f\|_{\Psi_2} \leq 2e \sup_{q \geq 1} \frac{\|f - \mathbb{E} f\|_q}{q^{1/2}} \leq 4e\sigma_p.$$

However, a simple calculation shows that  $\mathbb{E} \exp\left(\frac{(f - \mathbb{E} f)^2}{16e^2\sigma_p^2}\right) \rightarrow \infty$  as  $p \rightarrow 0$ .

The approximate tensorization property in Theorem 7 is interesting in its own right, but it is not yet well-studied. For finite spaces [40] gives sufficient conditions for a measure  $\mu$  to satisfy an approximate tensorization property. Similar results have been derived in [21], 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  $\Omega_i$ ,  $x_i \in \Omega_i$ , measures  $\mu_{0,i}$  on  $\Omega_i$  and bounded functions  $w_{ij}$ , under certain technical conditions  $\mu$  satisfies an approximate tensorization property. This does not require any functional inequality for  $\mu_{0,i}$ . Very recently, in [3, Proposition 5.4] it has been shown that the AT( $C$ ) property implies dimension-free concentration inequalities for convex functions.

Note that the AT( $C$ ) property requires a certain weak dependence assumption in general. For example, the push-forward of a random permutation  $\pi$  of  $[n]$  to  $\mathbb{N}^n$  cannot satisfy an approximate tensorization property. It is an interesting question to

find necessary and sufficient conditions for the approximate tensorization property to hold.

**Proof of Theorem 7** Let  $X = (X_1, \dots, X_n)$  be a  $\mathcal{Y}$ -valued random vector with law  $\mu$ . First we consider the case  $n = 1$ . By homogeneity of both sides, we may assume  $\int f^2(X)d\mathbb{P} = 1$ . Since  $f$  is bounded, we have  $0 \leq a \leq |f(X)| \leq b < \infty$   $\mathbb{P}$ -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  $a^2 \leq 1 \leq b^2$ . (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  $F(u) := \mathbb{P}(f^2(X) \geq u)$ . In particular

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

Using the partial integration formula (see e. g. [31, Theorem 21.67 and Remark 21.68]) in connection with [20, Theorem 7.7.1] yields

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

The first integral can be calculated explicitly

$$\int_0^{a^2} (\log u + 1)F(u)du = u(\log u - 1) \Big|_0^{a^2} = a^2 \log a^2,$$

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

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

Plugging in these two estimates yields

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

Next, if we show that

$$f(a, b) \leq 2(b - a)^2 \text{ on } G := \{(a, b) \in \mathbb{R}^2 : 0 \leq a \leq 1, 1 \leq b < \infty\}, \quad (26)$$

we can further estimate (as  $|\mathfrak{h}f|^2$  is a deterministic quantity in the case  $n = 1$ )

$$\text{Ent}(f^2(X)) \leq 2(b - a)^2 = 2 \mathbb{E} |\mathfrak{h}f|^2.$$

To prove (26), define

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

Now it is easy to see that  $g(a, 1) = a^2 \log a^2 + (1 - a^2) - 2(1 - a)^2 \leq 0$ , since  $\partial_a g(a, 1) \geq 0$  for  $a \in [0, 1]$  and  $g(1, 1) = 0$ . Moreover

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

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

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

$$\text{Ent}_{\mu(\cdot | x_{i^c})} (f^2(x_{i^c}, \cdot)) \leq 2 \sup_{y'_i, y''_i} |f(x_{i^c}, y'_i) - f(x_{i^c}, y''_i)|^2.$$

Plugging this into the assumption leads to

$$\text{Ent}_\mu (f^2) \leq 2C \int \sum_{i=1}^n \sup_{y'_i, y''_i} |f(x_{i^c}, y'_i) - f(x_{i^c}, y''_i)|^2 d\mu_{i^c}(x_{i^c}) = 2C \int |\mathfrak{h}f|^2 d\mu.$$

As for the second part, it is a classical fact that independent random variables satisfy the tensorization property (i. e. AT(1)), see for example [38, Proposition 5.6], [16, Theorem 4.10] or [54, 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  $\mu(\cdot | x_{i^c}) := \mu_i = \mathbb{P} \circ X_i$ .  $\square$

**Acknowledgements** Open Access funding provided by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Adamczak, R.: Moment inequalities for  $U$ -statistics. *Ann. Probab.* **34**(6), 2288–2314 (2006). <https://doi.org/10.1214/009117906000000476>
2. Adamczak, R.: A note on the Hanson–Wright inequality for random vectors with dependencies. *Electron. Commun. Probab.* **20**(72), 13 (2015). <https://doi.org/10.1214/ECP.v20-3829>

3. Adamczak, R., Kotowski, M., Polaczyk, B., Strzelecki, M.: A note on concentration for polynomials in the Ising model. arXiv preprint (2018)
4. Adamczak, R., Latała, R., Meller, R.: Hanson–Wright inequality in Banach spaces. arXiv preprint (2018)
5. Adamczak, R., Wolff, P.: Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probab. Theory Relat. Fields* **162**(3–4), 531–586 (2015). <https://doi.org/10.1007/s00440-014-0579-3>
6. Aida, S., Stroock, D.W.: Moment estimates derived from Poincaré and logarithmic Sobolev inequalities. *Math. Res. Lett.* **1**(1), 75–86 (1994). <https://doi.org/10.4310/MRL.1994.v1.n1.a9>
7. Ambrosio, L., Gigli, N., Savaré, G.: *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics ETH Zürich, 2nd edn. Birkhäuser Verlag, Basel (2008)
8. Arcones, M.A., Giné, E.: On decoupling, series expansions, and tail behavior of chaos processes. *J. Theor. Probab.* **6**(1), 101–122 (1993). <https://doi.org/10.1007/BF01046771>
9. Bobkov, S.G., Chistyakov, G.P., Götze, F.: Second-order concentration on the sphere. *Commun. Contemp. Math.* **19**(5), 1650058 (2017). <https://doi.org/10.1142/S0219199716500589>
10. Bobkov, S.G., Götze, F., Sambale, H.: Higher order concentration of measure. *Commun. Contemp. Math.* **21**(3), 1850043 (2019). <https://doi.org/10.1142/S0219199718500438>
11. Bobkov, S.G., Ledoux, M.: Poincaré’s inequalities and Talagrand’s concentration phenomenon for the exponential distribution. *Probab. Theory Relat. Fields* **107**(3), 383–400 (1997). <https://doi.org/10.1007/s004400050090>
12. Bobkov, S.G., Tetali, P.: Modified logarithmic Sobolev inequalities in discrete settings. *J. Theor. Probab.* **19**(2), 289–336 (2006). <https://doi.org/10.1007/s10959-006-0016-3>
13. Bonami, A.: Ensembles  $\Lambda(p)$  dans le dual de  $D^\infty$ . *Ann. Inst. Fourier (Grenoble)* **18**(fasc. 2), 193–204 (1968). <https://doi.org/10.5802/aif.297>
14. Bonami, A.: Étude des coefficients de Fourier des fonctions de  $L^p(G)$ . *Ann. Inst. Fourier (Grenoble)* **20**(facs. 2), 335–402 (1970). <https://doi.org/10.5802/aif.357>
15. Borell, C.: On the Taylor series of a Wiener polynomial. In: *Seminar Notes on Multiple Stochastic Integration, Polynomial Chaos and their Integration*. Case Western Reserve University, Cleveland (1984)
16. Boucheron, S., Bousquet, O., Lugosi, G., Massart, P.: Moment inequalities for functions of independent random variables. *Ann. Probab.* **33**(2), 514–560 (2005). <https://doi.org/10.1214/009117904000000856>
17. Boucheron, S., Lugosi, G., Massart, P.: Concentration inequalities using the entropy method. *Ann. Probab.* **31**(3), 1583–1614 (2003). <https://doi.org/10.1214/aop/1055425791>
18. Boucheron, S., Lugosi, G., Massart, P.: *Concentration Inequalities*. Oxford University Press, Oxford (2013)
19. Bousquet, O.: A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris* **334**(6), 495–500 (2002). [https://doi.org/10.1016/S1631-073X\(02\)02292-6](https://doi.org/10.1016/S1631-073X(02)02292-6)
20. Burk, F.E.: *A Garden of Integrals, The Dolciani Mathematical Expositions*, vol. 31. Mathematical Association of America, Washington, DC (2007)
21. Caputo, P., Menz, G., Tetali, P.: Approximate tensorization of entropy at high temperature. *Ann. Fac. Sci. Toulouse Math.* (2015). <https://doi.org/10.5802/afst.1460>
22. Chafaï, D.: Entropies, convexity, and functional inequalities: on  $\Phi$ -entropies and  $\Phi$ -Sobolev inequalities. *J. Math. Kyoto Univ.* **44**(2), 325–363 (2004). <https://doi.org/10.1215/kjm/1250283556>
23. Chatterjee, S., Diaconis, P.: Estimating and understanding exponential random graph models. *Ann. Stat.* **41**(5), 2428–2461 (2013). <https://doi.org/10.1214/13-AOS1155>
24. de la Peña, V.H., Giné, E.: *Decoupling. Probability and its Applications*. Springer, New York (1999)
25. Dellacherie, C., Meyer, P.A.: *Probabilities and Potential*. North-Holland Mathematics Studies, vol. 29. North-Holland Publishing Co., Amsterdam (1978)
26. Diaconis, P., Saloff-Coste, L.: Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6**(3), 695–750 (1996). <https://doi.org/10.1214/aop/1034968224>
27. Efron, B., Stein, C.M.: The jackknife estimate of variance. *Ann. Stat.* **9**(3), 586–596 (1981). <https://doi.org/10.1214/aos/1176345462>
28. Götze, F., Sambale, H., Sinulis, A.: Higher order concentration for functions of weakly dependent random variables. *Electron. J. Probab.* **24**(85), 19 (2019)
29. Gross, L.: Logarithmic Sobolev inequalities. *Am. J. Math.* **97**(4), 1061–1083 (1975). <https://doi.org/10.2307/2373688>

30. Hanson, D.L., Wright, F.T.: A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Stat.* **42**, 1079–1083 (1971). <https://doi.org/10.1214/aoms/1177693335>
31. Hewitt, E., Stromberg, K.: *Real and Abstract Analysis*. Springer, New York (1975)
32. Hsu, D., Kakade, S.M., Zhang, T.: A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Probab.* **17**(52), 6 (2012). <https://doi.org/10.1214/ECP.v17-2079>
33. Kim, J.H., Vu, V.H.: Concentration of multivariate polynomials and its applications. *Combinatorica* **20**(3), 417–434 (2000). <https://doi.org/10.1007/s004930070014>
34. Klein, T., Rio, E.: Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33**(3), 1060–1077 (2005). <https://doi.org/10.1214/009117905000000044>
35. Latała, R.: Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.* **34**(6), 2315–2331 (2006). <https://doi.org/10.1214/009117906000000421>
36. Latała, R., Oleszkiewicz, K.: Between Sobolev and Poincaré. In: *Geometric Aspects of Functional Analysis. Lecture Notes in Mathematics*, vol. 1745, pp. 147–168. Springer, Berlin (2000). <https://doi.org/10.1007/BFb0107213>
37. Ledoux, M.: On Talagrand’s deviation inequalities for product measures. *ESAIM Probab. Stat.* **1**, 63–87 (1997). <https://doi.org/10.1051/ps:1997103>
38. Ledoux, M.: *The Concentration of Measure Phenomenon*, Mathematical Surveys and Monographs, vol. 89. American Mathematical Society, Providence (2001)
39. Marchina, A.: Concentration inequalities for separately convex functions. *Bernoulli* **24**(4A), 2906–2933 (2018). <https://doi.org/10.3150/17-BEJ949>
40. Marton, K.: *Logarithmic Sobolev Inequalities in Discrete Product Spaces: A Proof by a Transportation Cost Distance*. arXiv preprint (2015)
41. Massart, P.: About the constants in Talagrand’s concentration inequalities for empirical processes. *Ann. Probab.* **28**(2), 863–884 (2000). <https://doi.org/10.1214/aop/1019160263>
42. Milman, V.D., Schechtman, G.: *Asymptotic Theory of Finite-Dimensional Normed Spaces. Lecture Notes in Mathematics*, vol. 1200. Springer, Berlin (1986)
43. Nelson, E.: The free Markoff field. *J. Funct. Anal.* **12**, 211–227 (1973). [https://doi.org/10.1016/0022-1236\(73\)90025-6](https://doi.org/10.1016/0022-1236(73)90025-6)
44. O’Donnell, R.: *Analysis of Boolean Functions*. Cambridge University Press, New York (2014)
45. Raginsky, M., Sason, I.: *Concentration of Measure Inequalities in Information Theory, Communications, and Coding*. Now Publishers Inc., Delft (2014)
46. Rio, E.: Une inégalité de Bennett pour les maxima de processus empiriques. *Ann. Inst. Henri Poincaré Probab. Stat.* **38**(6), 1053–1057 (2002). [https://doi.org/10.1016/S0246-0203\(02\)01122-6](https://doi.org/10.1016/S0246-0203(02)01122-6)
47. Rudelson, M., Vershynin, R.: Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18**(82), 9 (2013). <https://doi.org/10.1214/ECP.v18-2865>
48. Sambale, H., Sinulis, A.: Logarithmic Sobolev inequalities for finite spin systems and applications. *Bernoulli* **26**(3), 1863–1890 (2020). <https://doi.org/10.3150/19-BEJ1172>
49. Samson, P.-M.: Infimum-convolution description of concentration properties of product probability measures, with applications. *Ann. Inst. Henri Poincaré Probab. Stat.* **43**(3), 321–338 (2007). <https://doi.org/10.1016/j.anihpb.2006.05.003>
50. Steele, J.M.: An Efron–Stein inequality for nonsymmetric statistics. *Ann. Stat.* **14**(2), 753–758 (1986). <https://doi.org/10.1214/aos/1176349952>
51. Talagrand, M.: A new isoperimetric inequality and the concentration of measure phenomenon. In: *Geometric Aspects of Functional Analysis. Lecture Notes in Mathematics*, vol. 1469, pp. 94–124. Springer, Berlin (1991). <https://doi.org/10.1007/BFb0089217>
52. Talagrand, M.: New concentration inequalities in product spaces. *Invent. Math.* **126**(3), 505–563 (1996). <https://doi.org/10.1007/s002220050108>
53. Talagrand, M.: Upper and lower bounds for stochastic processes. In: *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics*, vol. 60. Springer, Heidelberg (2014)
54. van Handel, R.: *Probability in high dimension. APC 550 Lecture Notes*, Princeton University (2016). <https://web.math.princeton.edu/~rvan/APC550.pdf>
55. Vu, V.H., Wang, K.: Random weighted projections, random quadratic forms and random eigenvectors. *Random Struct. Algorithms* **47**(4), 792–821 (2015). <https://doi.org/10.1002/rsa.20561>
56. Wolff, P.: On some Gaussian concentration inequality for non-Lipschitz functions. In: *High dimensional probability VI, Progress in Probability*, vol. 66, pp. 103–110. Birkhäuser/Springer, Basel (2013)

57. Wright, F.T.: A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Ann. Probability* **1**(6), 1068–1070 (1973). <https://projecteuclid.org/euclid.aop/1176996815>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.