Bernoulli **26**(1), 2020, 93–126
https://doi.org/10.3150/19-BEJ1118

Second order concentration via logarithmic Sobolev inequalities

FRIEDRICH GÖTZE* and HOLGER SAMBALE**

Department of Mathematics, Bielefeld University, Postbox 100131, 33501 Bielefeld, Germany.
E-mail: *goetze@math.uni-bielefeld.de; **hsambale@math.uni-bielefeld.de

We show sharpened forms of the concentration of measure phenomenon centered at first order stochastic expansions. The bound are based on second order difference operators and second order derivatives. Applications to functions on the discrete cube and stochastic Hoeffding type expansions in mathematical statistics are studied as well as linear eigenvalue statistics in random matrix theory.

Keywords: bootstrap approximation; concentration of measure phenomenon; functions on the discrete cube; Hoeffding decomposition; logarithmic Sobolev inequalities

1. Introduction

The concentration of measure phenomenon for product measures has been extensively studied in the past decades. It was established by M. Talagrand in the 1990s [31,32]. Further research was done by S. Bobkov, M. Ledoux and others [6,7,22]. For a comprehensive survey which summarizes the central concentration of measure results up to the end of the 1990s, see the monographs by M. Ledoux [23,24], for a more recent one see [13].

One of the basic results due to M. Talagrand are concentration inequalities for Lipschitz functions around their mean or median. For instance, in discrete probability models, the product probability space $(\Omega, \mathcal{A}, \mu) := \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ is typically equipped with the Hamming distance $d(x, y) := \operatorname{card}\{k = 1, \dots, n \colon x_k \neq y_k\}$. A related approach, which is essentially due to M. Ledoux [22], makes use of certain "difference operators". That is, for any function $f \colon \Omega \to \mathbb{R}$ in $L^2(\mu)$, set

$$\mathfrak{d}_i f(x) := \left(\frac{1}{2}\int_{\Omega_i} \big(f(x) - f(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_n)\big)^2 \mu_i(dy_i)\right)^{1/2} \tag{1.1}$$

and $\mathfrak{d} f := (\mathfrak{d}_1 f, \dots, \mathfrak{d}_n f)$. A slight modification of [7], Proposition 2.1, then yields

Proposition 1.1. *Let $(\Omega_i, \mathcal{A}_i, \mu_i)$ be probability spaces, and denote by $(\Omega, \mathcal{A}, \mu) := \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ their product. Moreover, let $f \colon \Omega \to \mathbb{R}$ be a bounded measurable function such that $\int f \, d\mu = 0$. Assume that $|\mathfrak{d} f| \leq 1$. Then, for any $t \geq 0$ we have*

$$\mu\big(|f| \geq t\big) \leq 2e^{-t^2/4}.$$

1350-7265

Note that the boundedness of f is in fact a consequence of the condition $|\mathfrak{d}f| \le 1$ (see Section 2). If we apply Proposition 1.1 to 1-Lipschitz functions with respect to the Hamming distance, we recover the classical concentration inequalities by M. Talagrand (cf. [22]). Similar results can be derived in the context of "penalties", which can be regarded as generalizations of the Hamming distance [22]. In [7], a generalized version of Proposition 1.1 is used for deriving concentration inequalities for randomized sums.

The tail bounds we deduce in the present article are motivated as follows: consider a suitably normalized non-linear statistic which is stochastically non-degenerate and bounded in the limit. By general principles, it will have the same limit distribution as a stochastically bounded Gaussian chaos functional. If it is non-degenerate of order 2 in the limit, it certainly has exponential tail decay.

As an example, consider a set of i.i.d. centered random variables $X_1, \ldots, X_n$ in L^∞ (e.g., Rademacher variables). Then, a particularly simple case where Proposition 1.1 applies is the function $g(X) := \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$. A natural second order analogue of g is the function $f(X) := \frac{1}{n} \sum_{i<j} X_i X_j$. However, in this case, evaluating the condition $|\mathfrak{d}f| \le 1$ and hence applying Proposition 1.1 does not lead to correct results. Indeed, f is not a function of a Lipschitz class bounded in n. This motivates the use of second order differences instead. A further aspect can be observed if we do not assume the X_i to be centered. In this case, we shall replace $f(X)$ by $Rf(X) := \frac{1}{n} \sum_{i<j} (X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)$. Comparing Rf to f, we see that we have not only removed the expected value of f but also a sort of "linear term".

Indeed, the notion of *second* order concentration has two aspects which generalize these observations. First, it refers to the use of difference operators of second order. Second, it means that instead of fluctuations of $f - \mathbb{E}f$ we will study fluctuations of $f - \mathbb{E}f - f_1$, where f_1 is the first order term in the Hoeffding decomposition of f. Let us briefly recall the notion of Hoeffding decomposition, which was introduced in [19]. Given a product probability space $(\Omega, \mathcal{A}, \mu) := \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ and some function $f \in L^1(\mu)$, the Hoeffding decomposition is the unique decomposition

$$\begin{aligned} f(x_1, \ldots, x_n) &= \int f\, d\mu + \sum_{i=1}^n h_i(x_i) + \sum_{i<j} h_{ij}(x_i, x_j) + \cdots \\ &= f_0 + f_1 + f_2 + \cdots + f_n \end{aligned} \tag{1.2}$$

such that $\int h_{i_1 \cdots i_k}(x_{i_1}, \ldots, x_{i_k}) \mu_{i_j}(dx_{i_j}) = 0$ for all $k = 1, \ldots, n$, $1 \le i_1 < \cdots < i_k \le n$ and $j \in \{1, \ldots, k\}$. The sum f_d is called the Hoeffding term of degree d or simply dth Hoeffding term of f. Note that for $f \in L^2(\mu)$ the f_j, $j \in \mathbb{N}_0$, form an orthogonal decomposition of f in $L^2(\mu)$.

We now formulate our main results. For that, we need to introduce a notion of second order differences based on $\mathfrak{d}$. Indeed, for any function $f\colon \Omega \to \mathbb{R}$ in $L^2(\mu)$ and any $i \ne j$, set

$$\begin{aligned} \mathfrak{d}_{ij} f(x) := \Big(\frac{1}{4} \int_{\Omega_i} \int_{\Omega_j} \big(& f(x) - f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, y_j, \ldots, x_n) \\ & + f(x_1, \ldots, y_i, \ldots, y_j, \ldots, x_n) \big)^2 \mu_i(dy_i) \mu_j(dy_j) \Big)^{1/2}. \end{aligned} \tag{1.3}$$

In particular, we consider the following modified "Hessian" with respect to $\mathfrak{d}$:

$$\big(\mathfrak{d}^{(2)} f(X)\big)_{ij} := \begin{cases} \mathfrak{d}_{ij} f(X), & i \neq j, \\ 0, & i = j. \end{cases} \tag{1.4}$$

For $x \in \mathbb{R}^n$ let $|x|$ denote its Euclidean norm, and for an $n \times n$ matrix $A = (a_{ij})_{ij}$ let $\|A\|_{\mathrm{HS}}$ denote its Hilbert–Schmidt norm given by $\|A\|_{\mathrm{HS}} = (\sum_{i,j=1}^n |a_{ij}|^2)^{1/2}$.

Theorem 1.2. *Let $(\Omega_i, \mathcal{A}_i, \mu_i)$ be probability spaces, and denote by $(\Omega, \mathcal{A}, \mu) := \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ their product. Moreover, let $f\colon \Omega \to \mathbb{R}$ be a bounded measurable function so that its Hoeffding decomposition with respect to μ is given by $f = \sum_{k=2}^n f_k$. Assume that the conditions*

$$\big|\mathfrak{d}|\mathfrak{d} f|\big| \le 1 \quad \textit{and} \quad \int \big\|\mathfrak{d}^{(2)} f\big\|_{\mathrm{HS}}^2 \, d\mu \le b^2 \tag{1.5}$$

are satisfied for some $b \ge 0$, where $\|\mathfrak{d}^{(2)} f\|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt norm of $\mathfrak{d}^{(2)} f$. Then, we have

$$\int \exp\bigg(\frac{1}{2(3+b^2)} |f|\bigg) \, d\mu \le 2.$$

Note that by Chebychev's inequality, Theorem 1.2 implies $\mu(|f| \ge t) \le 2e^{-ct}$ for all $t > 0$ and some constant $c = c(b^2)$ (in fact, $c = (2(3+b^2))^{-1}$). In other words, Theorem 1.2 yields subexponential tails with an optimal exponent for large t in accordance with the discussion above.

In Theorem 1.2, we have one condition assuming pointwise boundedness of second order-type differences and a second condition assuming boundedness in mean of the squared Hilbert–Schmidt norm of a suitable "Hessian". This mirrors the structure of Theorem 1.1 in S.G. Bobkov, G.P. Chistyakov and F. Götze [5]. Some ways of explicitly evaluating the pointwise condition $|\mathfrak{d}|\mathfrak{d} f|| \le 1$ are given in Section 6. In particular, if for large n, the much stronger Hilbert–Schmidt norm is used, this will bound the constant b in the second condition in (1.5). In other situations (especially in differentiable settings, cf. Section 1.1), $|\mathfrak{d}|\mathfrak{d} f||$ may be bounded by operator-type norms of second order differences or derivatives.

Note that in general, the two conditions are incomparable: though the pointwise condition will often dominate, sometimes it may happen that the second one is more restrictive. An elementary example is given by the function $f(X) := \frac{1}{\sqrt{n}}(X_1X_2 + X_3X_4 + \cdots)$ for a set of independent Rademacher variables and n even (here, we obviously have $|\mathfrak{d}|\mathfrak{d} f|| = 0$, and thus, the pointwise condition does not suffice in order to control the variance in the exponential estimates).

For applications, we formulate a convenient "hybrid" bound extending the results from Theorem 1.2 to functions with non-vanishing first order Hoeffding term (say, f_1). To this end, we need to provide that f_1 is of sufficiently small stochastic size. That is, in Theorem 1.2, let $f\colon \Omega \to \mathbb{R}$ be a function in $L^1(\mu)$ with Hoeffding decomposition $f = \sum_{k=0}^n f_k$. Then, we denote by

$$Rf := f - f_0 - f_1 = \sum_{k=2}^n f_k \tag{1.6}$$

the projection of f onto the space of the functions $f \in L^1(\mu)$ whose Hoeffding terms of orders 0 and 1 vanish. For convenience, we shall assume that the expected value f_0 of f vanishes. In order to obtain a result similar to Theorem 1.2, we add conditions ensuring $f_1 = \mathcal{O}_P(1)$ (cf. Proposition 1.1). The result is the following theorem.

Theorem 1.3. *Let $(\Omega_i, \mathcal{A}_i, \mu_i)$ be probability spaces, and denote by $(\Omega, \mathcal{A}, \mu) := \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ their product. Moreover, let $f : \Omega \to \mathbb{R}$ be a bounded measurable function such that its Hoeffding decomposition with respect to μ is given by $f = f_1 + \sum_{k=2}^n f_k = f_1 + Rf$. (In particular, we have $\mathbb{E} f = 0$.) Suppose that $|\mathfrak{d} f_1| \le b_0$ for some $b_0 \ge 0$ and that the conditions*

$$\big|\mathfrak{d}|\mathfrak{d} Rf|\big| \le 1 \quad \textit{and} \quad \int \|\mathfrak{d}^{(2)} f\|_{\mathrm{HS}}^2 \, d\mu \le b^2$$

for some $b \ge 0$ are satisfied. Then, we have

$$\int \exp\Big(\frac{1}{12 + 4b^2 + 7b_0} |f|\Big) \, d\mu \le 2.$$

Discussion of related inequalities. Hoeffding decompositions have been studied in particular in the context of U-statistics, that is, statistics of the form $U_n(h) = \frac{(n-m)!}{n!} \sum_{i_1 \neq \cdots \neq i_m} h(X_{i_1}, \ldots, X_{i_m})$ for a sequence of i.i.d. random variables $(X_i)_{i \in \mathbb{N}}$, a measurable kernel function h on $\mathbb{R}^m$ and natural numbers n, m such that $n \ge m$. A U-statistic is called completely degenerate (or canonical) if its Hoeffding decomposition consists of a single term only. There are a lot of results on the distributional properties of U-statistics. A partial overview is given in the monograph by V. de la Peña and E. Giné [15]. In particular, there are many inequalities describing their tail behavior starting with Hoeffding's inequalities. That is, for U-statistics like $U_n(h)$ introduced above, we have $P(U_n(h) > t) \le \exp(-[n/m]t^2/(2M^2))$ if the function $h : \mathbb{R}^m \to \mathbb{R}$ is bounded by some universal constant M and satisfies $\mathbb{E} h(X_1, \ldots, X_m) = 0$. Further exponential inequalities for completely degenerate U-statistics have been proved by M. Arcones and E. Giné [3] as well as P. Major [26]. These inequalities typically depend on the order m, the second moment σ^2 and some bound M of the kernel h only.

Finally, let us mention that in a subsequent paper together with S.G. Bobkov [9], we have extended some of the results of the present paper to arbitrary higher orders. However, the methodology is quite different. For instance, in the present paper our arguments are mainly based on modified logarithmic Sobolev inequalities and exponential inequalities which follow from them. By contrast, the main tool in [9] is a recursion inequality for the L^p-norms of f and its higher order differences (or derivatives) for any $p \ge 2$, which in turn does not appear in the present paper.

1.1. Differentiable functions

In arbitrary product spaces, the usual notion of differentiation is not available, which is why we need to work with difference operators as a kind of substitute. However, if we do consider differentiable settings, it seems natural to use the ordinary gradient ∇ instead. Therefore, we now complement our main theorems by results valid in differentiable settings.

Indeed, it is possible to formulate a result similar to Theorem 1.2 for probability measures on $\mathbb{R}^n$ which satisfy a logarithmic Sobolev inequality. Note that this situation has already been sketched in [5] (see Remark 5.3 there). In the present paper, we work out these ideas in detail and add some further material, including a version with an additional "linear" term similar to Theorem 1.3 and some applications. Let us first recall some basic notions.

Let $G \subset \mathbb{R}^n$ be some open set, and let μ be a probability measure on $(G, \mathcal{B}(G))$. Then, μ satisfies a *Poincaré inequality* with constant $\sigma^2 > 0$ if for all locally Lipschitz functions $f\colon G \to \mathbb{R}$

$$\mathrm{Var}_\mu(f) \le \sigma^2 \int_G |\nabla f|^2 \, d\mu, \tag{1.7}$$

where $\mathrm{Var}_\mu(f) = \int f^2 \, d\mu - (\int f \, d\mu)^2$ and $|\nabla f|$ denotes the Euclidean norm of the usual gradient. Another type of functional inequality for probability measures μ on $(G, \mathcal{B}(G))$ is given by the *logarithmic Sobolev inequality*. That is, μ satisfies a logarithmic Sobolev inequality with (Sobolev) constant $\sigma^2 > 0$ if for all locally Lipschitz functions $f\colon G \to \mathbb{R}$

$$\mathrm{Ent}_\mu\big(f^2\big) \le 2\sigma^2 \int_G |\nabla f|^2 \, d\mu, \tag{1.8}$$

where $\mathrm{Ent}_\mu(f^2) = \int f^2 \log f^2 \, d\mu - \int f^2 \, d\mu \log \int f^2 \, d\mu$ (see Section 3). Logarithmic Sobolev inequalities are stronger than Poincaré inequalities. For instance, if μ satisfies a logarithmic Sobolev inequality with constant σ^2, it also satisfies a Poincaré inequality with the same constant σ^2.

We now have the following result.

Theorem 1.4. *Let $G \subset \mathbb{R}^n$ be some open set, and let μ be a probability measure on $(G, \mathcal{B}(G))$ which satisfies a logarithmic Sobolev inequality with constant $\sigma^2 > 0$. Let $f\colon G \to \mathbb{R}$ be a $\mathcal{C}^2$-smooth function such that $f \in L^1(\mu)$ and $\partial_i f \in L^1(\mu)$ for all $i = 1, \dots, n$, where $\partial_i f$ denotes the ith partial derivative of f. Assume that $\int_G f \, d\mu = 0$ and $\int_G \partial_i f \, d\mu = 0$ for all $i = 1, \dots, n$. Moreover, assume that*

$$\big\| f''(x) \big\|_{\mathrm{Op}} \le 1 \quad \textit{for all } x \in G \quad \textit{and} \quad \int_G \big\| f'' \big\|_{\mathrm{HS}}^2 \, d\mu \le b^2$$

for some $b \ge 0$, where f'' denotes the Hessian of f and $\|f''\|_{\mathrm{Op}}$, $\|f''\|_{\mathrm{HS}}$ denote its operator and Hilbert–Schmidt norms, respectively. Then, the following inequality holds:

$$\int_G \exp\bigg(\frac{1}{2\sigma^2(1+b^2)} |f| \bigg) \, d\mu \le 2.$$

Note that unlike in Theorem 1.2, we do not need to require μ to be a product measure. Given any function $f \in \mathcal{C}^2(G)$ such that $f \in L^1(\mu)$ and $\partial_i f \in L^1(\mu)$ for all $i = 1, \dots, n$, we may modify f to remove a "linear" term by considering

$$Rf(x) = f(x) - \mu[f] - \sum_{i=1}^n \mu[\partial_i f]\big(x_i - \mu[x_i]\big), \tag{1.9}$$

where $\mu[h] = \int_G h\,d\mu$ for any function $h \in L^1(\mu)$. Rf represents a centered function with centered derivatives.

Similarly to Theorem 1.3, we may allow non-vanishing integrals $\mu[\partial_i f]$ in Theorem 1.4 if they are of sufficiently small size. This is the objective of the following theorem.

Theorem 1.5. *Let $G \subset \mathbb{R}^n$ be some open set, and let μ be a probability measure on $(G, \mathcal{B}(G))$ which satisfies a logarithmic Sobolev inequality with constant $\sigma^2 > 0$. Let $f : G \to \mathbb{R}$ be a $\mathcal{C}^2$-smooth function such that $f \in L^1(\mu)$ and $\partial_i f \in L^1(\mu)$ for all $i = 1, \ldots, n$, where $\partial_i f$ denotes the ith partial derivative of f. Assume that $\int_G f\,d\mu = 0$ and $\sum_{i=1}^n (\int_G \partial_i f\,d\mu)^2 \leq \sigma^2 b_0^2$ for some $b_0 \geq 0$. Moreover, assume that*

$$\|f''(x)\|_{\mathrm{Op}} \leq 1 \quad \textit{for all } x \in G \quad \textit{and} \quad \int_G \|f''(x)\|_{\mathrm{HS}}^2\,d\mu \leq b^2$$

for some $b \geq 0$, where f'' denotes the Hessian of f and $\|f''\|_{\mathrm{Op}}$, $\|f''\|_{\mathrm{HS}}$ denote its operator and Hilbert–Schmidt norms, respectively. Then, we have

$$\int_G \exp\left(\frac{1}{\sigma^2(4 + 4b^2 + 5b_0)}|f|\right) d\mu \leq 2.$$

Discussion of related inequalities. We shall compare our results to a measure concentration result for functions on the n-sphere which are orthogonal to linear functions, see S.G. Bobkov, G.P. Chistyakov and F. Götze [5]. In this context, Theorem 1.2 can be regarded as a "discrete" analogue of the latter result. Note that in particular, it covers the case of the discrete hypercube $\{\pm 1\}^n$ equipped with the uniform distribution. Theorem 1.4 may then be seen as an intermediate between Theorem 1.2 and the bounds in [5]. Indeed, if in Theorem 1.4 μ is the standard Gaussian measure, the condition $\int \partial_i f\,d\mu = 0$ for all i is satisfied if we require orthogonality to all linear functions (by partial integration). The idea of sharpening concentration inequalities for Gaussian and related measures by requiring orthogonality to linear functions also appears in [14].

We would moreover like to mention the results by R. Adamczak and P. Wolff [1]. They study the tail behavior of differentiable functions. Requiring certain Sobolev-type inequalities or subgaussian tail conditions, they derive exponential inequalities for functions with bounded higher-order derivatives (evaluated in terms of some tensor-product matrix norms). In comparison, our paper has a stronger emphasis on discrete models and difference operators with a focus on functions structured by Hoeffding expansions of vanishing first order or, in differentiable cases as in Theorem 1.4, functions from which we remove a kind of "linear term".

1.2. Outline

The main tools we use in this article will be introduced in Sections 2 and 3. This includes some basic facts about difference operators, Hoeffding decompositions and modified logarithmic Sobolev inequalities. The proofs of our main theorems for product measures will be given in Sections 4 and 5. Here, we will first derive exponential inequalities based on modified Sobolev inequalities. After that, second order differences will be invoked by making use of certain “harmonic analysis” arguments on the symmetric group established in Section 2. The proof of Theorem 1.2 then follows as an easy combination of both chains of arguments.

In Section 6, we discuss how to evaluate the second order conditions from Theorem 1.2. In particular, we give a reformulation of Theorem 1.2 which involves conditions which may be easier to apply. We also apply our results to functions of independent Rademacher variables.

The differentiable case will be discussed in Section 7. Here we need to modify some of the arguments from the proof of Theorems 1.2 and 1.3. Together with a simple application of the Poincaré inequality, this will lead us to the proof of Theorems 1.4 and 1.5.

Finally, Section 8 presents a number of examples for functions of independent random variables as well as in differentiable settings.

A prior version of these results is based on the Ph.D. thesis of the second author [29].

2. Difference operators

Let $(\Omega_1, \mathcal{A}_1), \dots, (\Omega_n, \mathcal{A}_n)$ be measurable spaces, and denote by $(\Omega, \mathcal{A})$ their product space. Similarly to [6], we study (difference) operators Γ on the space of the bounded measurable real-valued functions on $(\Omega, \mathcal{A})$ such that the following two conditions hold (in particular, no sort of “Leibniz rule” is required):

Conditions 2.1.

(i) For any bounded measurable function $f\colon \Omega \to \mathbb{R}$, $\Gamma f = (\Gamma_1 f, \dots, \Gamma_n f)\colon \Omega \to \mathbb{R}^n$ is a measurable function with values in $\mathbb{R}^n$. We often call Γ a *gradient operator* or simply *gradient*.

(ii) For all $i = 1, \dots, n$, all $a > 0$, $b \in \mathbb{R}$ and any bounded measurable real-valued function f, we have $|\Gamma_i(af + b)| = a|\Gamma_i f|$.

In addition to the “L^2 difference operator” $\mathfrak{d}$ in (1.1), we need a difference operator adapted to the Hoeffding decomposition. Indeed, for any function $f\colon \Omega \to \mathbb{R}$ in $L^1(\mu)$, let

$$\mathfrak{D}_i f(x) := f(x) - \int_{\Omega_i} f(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_n)\mu_i(dy_i) \tag{2.1}$$

and $\mathfrak{D} f := (\mathfrak{D}_1 f, \dots, \mathfrak{D}_n f)$. Higher order differences are defined by iteration, e.g. $\mathfrak{D}_{ij} f := \mathfrak{D}_i(\mathfrak{D}_j f)$ for $1 \le i, j \le n$. As in (1.4), we then define a modified “Hessian” with respect to $\mathfrak{D}$ by

$$\big(\mathfrak{D}^{(2)} f(x)\big)_{ij} := \begin{cases} \mathfrak{D}_{ij} f(x), & i \neq j, \\ 0, & i = j. \end{cases} \tag{2.2}$$

The difference operator $\mathfrak{D}$ is closely related to the Hoeffding decomposition (1.2). In essence, proving (1.2) is based on the identity $\mathbb{E}_i + \mathfrak{D}_i = \mathrm{Id}$ with $\mathfrak{D}_i$ as in (2.1). We finally get $h_{i_1\cdots i_k}(X_{i_1}, \dots, X_{i_k}) = (\prod_{j \notin \{i_1,\dots,i_k\}} \mathbb{E}_j \prod_{l \in \{i_1,\dots,i_k\}} \mathfrak{D}_l) f(X_1, \dots, X_n)$.

Let us collect some elementary facts about the difference operators $\mathfrak{d}$ and $\mathfrak{D}$. In the following assume that $X_1, \dots, X_n$ is a sequence of independent random variables on some probability space $(\Omega', \mathcal{A}', P)$ with distributions $\mu_1, \dots, \mu_n$ respectively. As we will see, introducing random variables sometimes facilitates notation:

Remark 2.2.

1. If $\mu_i = \frac{1}{2}\delta_{+1} + \frac{1}{2}\delta_{-1}$ for all $i = 1, \dots, n$, we have $\mathfrak{D}_i f(X) = \frac{1}{2}(f(X) - f(\sigma_i X))$, where $X = (X_1, \dots, X_n)$ and $\sigma_i X := (X_1, \dots, -X_i, \dots, X_n)$. Moreover, note that $\mathfrak{d}_i f = |\mathfrak{D}_i f|$.
2. For any function $f(X) \in L^1(P)$, we have $\mathfrak{D}_i f(X) = f(X) - \mathbb{E}_i f(X)$ or (in short) $\mathfrak{D}_i = \mathrm{Id} - \mathbb{E}_i$. Here, Id denotes the identity and $\mathbb{E}_i$ taking the expectation with respect to X_i.
3. Let $f(X) \in L^2(P)$, and let $\bar{X}_1, \dots, \bar{X}_n$ be a set of independent copies of the random variables $X_1, \dots, X_n$. Set $T_i f := f(X_1, \dots, X_{i-1}, \bar{X}_i, X_{i+1}, \dots, X_n)$ for any function $f(X_1, \dots, X_n)$. Then, we have

$$\mathfrak{d}_i f(X) = \Big(\frac{1}{2}\bar{\mathbb{E}}_i\big(f(X) - T_i f(X)\big)^2\Big)^{1/2}.$$

   Here, $\bar{\mathbb{E}}_i$ denotes the expectation with respect to $\bar{X}_i$. By independence, if $\mathbb{E}_i$ denotes the expectation with respect to X_i we can rewrite

$$\begin{aligned}\mathfrak{d}_i f(X) &= \Big(\frac{1}{2}\big(\big(f(X) - \mathbb{E}_i f(X)\big)^2 + \mathbb{E}_i\big(f(X) - \mathbb{E}_i f(X)\big)^2\big)\Big)^{1/2} \\ &= \Big(\frac{1}{2}\big(\big(\mathfrak{D}_i f(X)\big)^2 + \mathbb{E}_i\big(\mathfrak{D}_i f(X)\big)^2\big)\Big)^{1/2}. \qquad (2.3)\end{aligned}$$

4. Setting $T_{ij} = T_i \circ T_j$, second order analogues of the formulas for $\mathfrak{d}_i$ are given by

$$\mathfrak{d}_{ij} f(X) = \Big(\frac{1}{4}\bar{\mathbb{E}}_{ij}\big(f(X) - T_i f(X) - T_j f(X) + T_{ij} f(X)\big)^2\Big)^{1/2}, \qquad (2.4)$$

$$\mathfrak{d}_{ij} f = \Big(\frac{1}{4}\big((\mathfrak{D}_{ij} f)^2 + \mathbb{E}_i(\mathfrak{D}_{ij} f)^2 + \mathbb{E}_j(\mathfrak{D}_{ij} f)^2 + \mathbb{E}_{ij}(\mathfrak{D}_{ij} f)^2\big)\Big)^{1/2} \qquad (2.5)$$

   for any $i \neq j$. Here, $\bar{\mathbb{E}}_{ij}$ means taking the expectation with respect to $\bar{X}_i$ and $\bar{X}_j$, and $\mathbb{E}_{ij}$ means taking the expectation with respect to X_i and X_j.

By induction over n, f is bounded if and only if $|\mathfrak{D} f|$ is bounded. Using (2.3), the same holds for $|\mathfrak{d} f|$ instead of $|\mathfrak{D} f|$. Moreover, it follows immediately from (2.5) that

$$\int \|\mathfrak{d}^{(2)} f\|_{\mathrm{HS}}^2 \, d\mu = \int \|\mathfrak{D}^{(2)} f\|_{\mathrm{HS}}^2 \, d\mu, \qquad (2.6)$$

which will turn out to be an important identity in our proof.

For some kind of "harmonic" analysis arguments on the symmetric group, we shall need a specific second order operator we would call "Laplacian". Since in our discrete setting $\mathfrak{D}_{ii} = \mathfrak{D}_i$ for all i, this cannot be $\mathfrak{L} = \sum_i \mathfrak{D}_{ii}$. Instead, we define

$$\mathfrak{L} := \sum_{i \neq j} \mathfrak{D}_{ij}. \tag{2.7}$$

Calling (2.7) a Laplacian is justified for several reasons. First of all, (2.7) enjoys similar properties with respect to scalar products in function spaces (see Lemma 5.1 below) compared to the classical Euclidean or spherical Laplacian. Moreover, if we assume $\mu_i \equiv \mu_1$ for all i in Example 2.2, that is for functions of i.i.d. random variables, the Laplacian (2.7) is invariant under permutations, that is, $\mathfrak{L} f(x) = \mathfrak{L} f(\pi(x))$ for any μ-integrable function f on $\mathbb{R}^n$ and any permutation π of $\{1, 2, \ldots, n\}$. As usual, here we set $f(\pi(x)) = f(x_{\pi^{-1}(1)}, \ldots, x_{\pi^{-1}(n)})$. This may be regarded as a discrete analogue of the rotational invariance of the usual Laplacian.

Relating the Hoeffding decomposition to the Laplacian $\mathfrak{L}$ yields the following result.

Theorem 2.3. *Let $(\Omega_i, \mathcal{A}_i, \mu_i)$ be probability spaces, and denote by $(\Omega, \mathcal{A}, \mu) := \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ their product. Moreover, let f be some function in $L^1(\mu)$ with Hoeffding decomposition $f = \sum_{d=0}^n f_d$. Then, we have*

$$\mathfrak{L} f_d = (d)_2 f_d.$$

Here, $\mathfrak{L}$ is the Laplacian as introduced in (2.7), and we write $(d)_2 = d(d-1)$. Thus, the dth Hoeffding term is an eigenfunction of $\mathfrak{L}$ with eigenvalue $(d)_2$.

Consequently, there is an orthogonal decomposition of L^2-functions f on which the Laplacian operates diagonally.

Proof. Write $f_d(x_1, \ldots, x_n) = \sum_{i_1 < \cdots < i_d} h_{i_1 \cdots i_d}(x_{i_1}, \ldots, x_{i_d})$ as in (1.2). Fix $i_1 < \cdots < i_d$. Then, we get

$$\int h_{i_1 \cdots i_d}(x_{i_1}, \ldots, x_{i_d}) \mu_i(dx_i) = \begin{cases} 0, & i \in \{i_1, \ldots, i_d\}, \\ h_{i_1 \cdots i_d}(x_{i_1}, \ldots, x_{i_d}), & i \notin \{i_1, \ldots, i_d\}. \end{cases}$$

Therefore, we have

$$\mathfrak{D}_i f_d(x_1, \ldots, x_n) = \sum_{\substack{i_1 < \cdots < i_d \\ i \in \{i_1, \ldots, i_d\}}} h_{i_1 \cdots i_d}(x_{i_1}, \ldots, x_{i_d}), \tag{2.8}$$

$$\mathfrak{D}_{ij} f_d(x_1, \ldots, x_n) = \sum_{\substack{i_1 < \cdots < i_d \\ i, j \in \{i_1, \ldots, i_d\}}} h_{i_1 \cdots i_d}(x_{i_1}, \ldots, x_{i_d}). \tag{2.9}$$

Hence, it remains to check how often each term $h_{i_1 \cdots i_d}(x_{i_1}, \ldots, x_{i_d})$ appears in $\mathfrak{L} f_d = \sum_{i \neq j} \mathfrak{D}_{ij} f_d$. As we just saw, each pair $i \neq j$ such that $i, j \in \{i_1, \ldots, i_d\}$ replicates the summand $h_{i_1 \cdots i_d}(x_{i_1}, \ldots, x_{i_d})$ precisely once. As there are $d(d-1) = (d)_2$ such pairs, we arrive at the result. $\square$

In fact, there are at least two larger families of difference operators which satisfy similar "invariance properties" with respect to the symmetric group and the Hoeffding decomposition. One family of this type can be defined via $\mathfrak{L}_1 := \sum_i \mathfrak{D}_i$, $\mathfrak{L}_2 := \mathfrak{L}_1^2$ and more generally $\mathfrak{L}_k := \mathfrak{L}_1^k$ for any $k \in \{1, 2, \ldots, n\}$. Another one is given by $\mathfrak{L}_k^* := \sum_{i_1 \neq i_2 \neq \cdots \neq i_k} \mathfrak{D}_{i_1} \cdots \mathfrak{D}_{i_k}$ for any $k \in \{1, 2, \ldots, n\}$. It is possible to relate these two families to each other by representing the $\mathfrak{L}_k^*$ as polynomials in $\mathfrak{L}_1$, for example, we have $\mathfrak{L}_2^* = \mathfrak{L}_1^2 - \mathfrak{L}_1$.

As in the proof of Theorem 2.3, simple combinatorial arguments show that all the $\mathfrak{L}_k$ and $\mathfrak{L}_k^*$ operate diagonally on the Hoeffding decomposition. In case of the $\mathfrak{L}_k^*$, the eigenvalues of the Hoeffding terms of order up to $k - 1$ are 0.

In particular, with $\mathfrak{L}$ as in (2.7), we see that we have $\mathfrak{L} = \mathfrak{L}_2^*$. In other words, $\mathfrak{L}$ is the second order difference invariant operator which annihilates the Hoeffding terms up to first order. This is in accordance with our basic concept of second order concentration.

3. Modified logarithmic Sobolev inequalities and exponential inequalities

Let μ be a probability measure on some measurable space $(\Omega, \mathcal{A})$ and $g \colon \Omega \to [0, \infty)$ a measurable function. Then, we define the entropy of g with respect to μ by $\mathrm{Ent}(g) := \mathrm{Ent}_\mu(g) := \int g \log g \, d\mu - \int g \, d\mu \log \int g \, d\mu$. Here, we set $\mathrm{Ent}(g) := \infty$ if any of the integrals involved does not exist. A natural condition for existence of entropy is whether the integral of $g \log(1 + g)$ is finite or not. It is well known that by Jensen's inequality, we have $\mathrm{Ent}(g) \in [0, \infty]$. As a modification of the usual logarithmic Sobolev inequality, we now define the following.

Definition 3.1. Let μ be a probability measure on some measurable space $(\Omega, \mathcal{A})$, and let Γ be a difference operator on this space satisfying Conditions 2.1. Then, μ satisfies a modified logarithmic Sobolev inequality with constant $\sigma^2 > 0$ with respect to Γ if for any bounded measurable function $f \colon \Omega \to \mathbb{R}$

$$\mathrm{Ent}\big(e^f\big) \leq \frac{\sigma^2}{2} \int |\Gamma f|^2 e^f \, d\mu. \tag{3.1}$$

Here, $|\Gamma f|$ denotes the Euclidean norm of the gradient Γf.

This definition goes back to [6], where it is called LSI_{σ^2}. The term "modified logarithmic Sobolev inequality" is due to [24], Chapter 5.3, where other modifications of logarithmic Sobolev inequalities are discussed as well. The difference between the usual form of the LSI and the modified one in (3.1) is motivated by the fact that difference operators do not necessarily satisfy any sort of chain rule. The number $\sigma^2 > 0$ is also called *Sobolev constant*. When using σ instead of σ^2 itself, we will always assume it to be positive.

We will use Definition 3.1 with $\Gamma = \mathfrak{d}$. Note that setting $\Gamma = \mathfrak{D}$ would be too restrictive since in this case, only discrete probability measures with a finite number of atoms would have a chance to fulfill a modified LSI of type (3.1). By contrast, in case of $\mathfrak{d}$ we have the following proposition.

Proposition 3.2. *Let μ be any probability measure on some measurable space $(\Omega, \mathcal{A})$. Then, μ satisfies the modified LSI (3.1) with Sobolev constant $\sigma^2 = 2$ with respect to the gradient operator $\mathfrak{d}$ from (1.1).*

Proof. This is due to [7] and essentially based on [22]. For the reader's convenience, we include a sketch of its proof here. First, we apply Jensen's inequality to get

$$\begin{aligned}\mathrm{Ent}_\mu\big(e^g\big) \le \mathrm{Cov}_\mu\big(g, e^g\big) &= \frac{1}{2}\iint \big(g(x) - g(y)\big)\big(e^{g(x)} - e^{g(y)}\big)\mu(dx)\mu(dy)\\ &\le \frac{1}{4}\iint \big(g(x) - g(y)\big)^2\big(e^{g(x)} + e^{g(y)}\big)\mu(dx)\mu(dy) = \int |\mathfrak{d}g|^2 e^g \, d\mu.\end{aligned}$$

Here g is any real-valued measurable function on Ω such that the integrals involved are finite, and the next-to-last step uses the elementary estimate $(a-b)(e^a - e^b) \le \frac{1}{2}(a-b)^2(e^a + e^b)$ for all $a, b \in \mathbb{R}$. However, this means that μ satisfies the modified LSI (3.1) with Sobolev constant $\sigma^2 = 2$. □

If we especially consider two-point measures, the Sobolev constant can still be improved a little by the following.

Proposition 3.3. *Let $\mu = p\delta_{+1} + (1-p)\delta_{-1}$ for some $p \in (0, 1)$, where δ_x denotes the Dirac measure in $x \in \mathbb{R}$. Then, μ satisfies the modified LSI (3.1) with Sobolev constant $\sigma^2 = 1$ with respect to $\mathfrak{d}$ as in (1.1).*

This is again due to [7], and we omit the proof here. It is easy to verify that for instance, in case of $p = \frac{1}{2}$, this constant is optimal.

From Propositions 3.2 and 3.3, we can easily go on to product spaces by the following tensorization property which goes back to [22].

Lemma 3.4. *For all $i = 1, \dots, n$, let $(\Omega_i, \mathcal{A}_i)$ be measurable spaces equipped with probability measures μ_i each satisfying the modified LSI (3.1) with Sobolev constants $\sigma_i^2 > 0$ with respect to $\mathfrak{d}$ as in (1.1). Then, the product measure $\mu_1 \otimes \cdots \otimes \mu_n$ on $(\Omega_1 \times \cdots \times \Omega_n, \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n)$ also satisfies the modified LSI (3.1) with Sobolev constant $\sigma^2 = \max_{i=1,\dots,n} \sigma_i^2$ with respect to $\mathfrak{d}$.*

As in the case the usual logarithmic Sobolev inequality, this is a consequence of the subadditivity (or tensorization) property of the entropy functional together with the additivity property of the gradient operator $\mathfrak{d}$. Therefore, Propositions 3.2 and 3.3 naturally extend to product measures.

4. Exponential inequalities

In this section, we derive exponential moment inequalities for functions of independent random variables. Consider any probability measure on some measurable space $(\Omega, \mathcal{A})$ which satisfies the modified LSI (3.1) with Sobolev constant $\sigma^2 > 0$ with respect to $\mathfrak{d}$. In S.G. Bobkov and F. Götze [6], it was proved that for all bounded measurable functions $f\colon \Omega \to \mathbb{R}$ such that $\int f \, d\mu = 0$, we have

$$\int e^f \, d\mu \le \int e^{\sigma^2 |\mathfrak{d} f|^2} \, d\mu. \tag{4.1}$$

The proof of (4.1) is similar to the proof of inequality (4.5) which will be sketched in the proof of Lemma 4.2.

In addition to (4.1), we need a second inequality of the form

$$\int e^{tu^2} \, d\mu \le \exp\Big(c(t) \int u^2 \, d\mu \Big)$$

for small t and some constant c depending on t. An inequality of the desired form due to S. Aida, T. Masuda and I. Shikegawa [2] is known if the underlying gradient operator satisfies the chain rule (cf. (7.4) in Section 7). Here, the main argument for which the chain rule is needed is as follows: let ∇ denote the usual gradient and $|\nabla f|$ its Euclidean norm. Then, if we assume $|\nabla f| \le 1$, we immediately get $|\nabla f^2| = 2|f||\nabla f| \le 2|f|$. However, if we replace ∇ by the L^2-difference operator $\mathfrak{d}$ from (1.1), such an inequality does not hold.

This desirable property is restored by switching to yet another difference operator which we denote by $\mathfrak{d}^+$. In detail,

$$\mathfrak{d}_i^+ f(x) := \Big(\frac{1}{2} \int_{\Omega_i} \big(f(x) - f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n) \big)_+^2 \mu_i(dy_i) \Big)^{1/2}. \tag{4.2}$$

Here, $f\colon \Omega \to \mathbb{R}$ is any function in $L^2(\mu)$, and $g_+ := \max(g, 0)$ denotes the positive part of any real-valued function g. As always, $\mathfrak{d}^+ f = (\mathfrak{d}_1^+ f, \ldots, \mathfrak{d}_n^+ f)$.

Let $f\colon \Omega \to \mathbb{R}$ be any measurable function on some probability space $(\Omega, \mathcal{A}, \mu)$. Then, for any $x, y \in \Omega$ we have

$$\big(f(x)^2 - f(y)^2\big)_+^2 = \big(|f(x)| + |f(y)|\big)^2 \big(|f(x)| - |f(y)|\big)_+^2 \le 4 |f(x)|^2 \big(|f(x)| - |f(y)|\big)_+^2.$$

Taking integrals and roots, we thus get that for any function $f\colon \Omega \to \mathbb{R}$ in $L^2(\mu)$ such that $|\mathfrak{d}^+ |f|| \le 1$, we have

$$\big| \mathfrak{d}^+ f^2 \big| \le 2|f|. \tag{4.3}$$

The same holds for product measures, that is, the multivariate case.

In the sequel, we also need modified LSI results for $\mathfrak{d}^+$. It is easily seen that if some measurable space $(\Omega, \mathcal{A})$ equipped with a probability measure μ satisfies the modified LSI (3.1) with Sobolev constant $\sigma^2 > 0$ with respect to $\mathfrak{d}$, it also satisfies the modified LSI (3.1) with respect to $\mathfrak{d}^+$, and the Sobolev constant can be chosen $2\sigma^2$. Hence, we can transport Propositions 3.2 and 3.3 and Lemma 3.4 to the $\mathfrak{d}^+$ difference operators. In fact, results of this type can already be found in [24] (Proposition 5.8) or [12] (e.g., Proposition 10).

Proposition 4.1. *For all $i = 1, \ldots, n$, let $(\Omega_i, \mathcal{A}_i)$ be measurable spaces equipped with probability measures μ_i. Then, the product measure $\mu_1 \otimes \cdots \otimes \mu_n$ on $(\Omega_1 \times \cdots \times \Omega_n, \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n)$*

satisfies the modified LSI (3.1) with Sobolev constant $\sigma^2 = 4$ with respect to $\mathfrak{d}^+$ as in (4.2). If all the Ω_i are two-point spaces, we can take $\sigma^2 = 2$.

Now (4.3) leads us back to the basic inequality needed to estimate large deviations in [6]. We therefore arrive at the following lemma.

Lemma 4.2. *Let μ be a probability measure on some measurable space $(\Omega, \mathcal{A})$ which satisfies the modified LSI (3.1) with Sobolev constant $\tilde{\sigma}^2 > 0$ with respect to the gradient operator $\mathfrak{d}^+$ from (4.2). Moreover, let $f\colon \Omega \to \mathbb{R}$ be a bounded measurable function such that $|\mathfrak{d}|f|| \le 1$. Then, for all $t \in [0, \frac{1}{2\tilde{\sigma}^2})$ we have*

$$\int e^{tf^2}\, d\mu \le \exp\Big(\frac{t}{1-2\tilde{\sigma}^2 t}\int f^2\, d\mu\Big). \tag{4.4}$$

Proof. We adapt the arguments from [6], p. 6 f. First, consider the inequality

$$\int e^f\, d\mu \le \Big(\int e^{\lambda f + (1-\lambda)\tilde{\sigma}^2|\mathfrak{d}^+ f|^2/2}\, d\mu\Big)^{1/\lambda} \tag{4.5}$$

for all bounded measurable functions $f\colon \Omega \to \mathbb{R}$ and all $\lambda \in (0,1]$. Here, we have already plugged in $\mathfrak{d}^+$ as our choice of the difference operator. To deduce (4.5), we use the well-known "variational formula"

$$\mathrm{Ent}(g) = \sup\Big\{\int gh\, d\mu\colon h\colon \Omega \to \mathbb{R} \text{ measurable s. th. } \int e^h\, d\mu \le 1\Big\},$$

which can be shown by Young's inequality in the form $uv \le u\log u - u + e^v$ for all $u \ge 0$ and $v \in \mathbb{R}$, for instance. See [24], Proposition 5.6, for details. If we set $g := e^f$ and $h := \lambda f + (1-\lambda)\tilde{\sigma}^2|\mathfrak{d}^+ f|^2/2 - \beta$ with $\beta = \log\int e^{\lambda f + (1-\lambda)\tilde{\sigma}^2|\mathfrak{d}^+ f|^2/2}\, d\mu$, we have $\int e^h\, d\mu = 1$ and thus

$$\int \big(\lambda f + (1-\lambda)\tilde{\sigma}^2\big|\mathfrak{d}^+ f\big|^2/2 - \beta\big)e^f\, d\mu \le \mathrm{Ent}\big(e^f\big).$$

Since f satisfies the modified LSI (3.1) with constant $\tilde{\sigma}^2$, it follows that

$$\begin{aligned} &\lambda\int fe^f\, d\mu + (1-\lambda)\,\mathrm{Ent}\big(e^f\big) - \beta\int e^f\, d\mu \le \mathrm{Ent}\big(e^f\big) \\ \Leftrightarrow\quad &\lambda\int e^f\, d\mu \log\int e^f\, d\mu - \beta\int e^f\, d\mu \le 0, \end{aligned}$$

from which we directly get (4.5).

We now apply (4.5) to the function $sf^2/(2\tilde{\sigma}^2)$ with $0 < s < 1$ and $\lambda = (p-s)/(1-s)$ for any $p \in (s,1]$. Together with (4.3) (note that $|\mathfrak{d}|f|| \le 1$ implies $|\mathfrak{d}^+|f|| \le 1$), this gives

$$\int e^{sf^2/(2\tilde{\sigma}^2)}\, d\mu \le \Big(\int \exp\Big(\frac{psf^2}{2\tilde{\sigma}^2}\Big)\, d\mu\Big)^{(1-s)/(p-s)}.$$

For $p = 1$ both sides are equal, and as for $p < 1$ the upper inequality holds, we get that the logarithm of the left-hand side (considered as a function of p) must increase more rapidly at $p = 1$ than that of the right-hand side. We thus consider the derivatives of the logarithms of both sides at $p = 1$ and arrive at the inequality

$$0 \geq \frac{1}{1-s}\Big[(1-s)\int \frac{sf^2}{2\tilde{\sigma}^2} e^{sf^2/(2\tilde{\sigma}^2)}\,d\mu - \int e^{sf^2/(2\tilde{\sigma}^2)}\,d\mu \log \int e^{sf^2/(2\tilde{\sigma}^2)}\,d\mu\Big].$$

Now we set $u(s) := \int e^{sf^2/(2\tilde{\sigma}^2)}\,d\mu$, $s \in (0, 1]$. Then we get

$$0 \geq \frac{1}{1-s}\big[s(1-s)u'(s) - u(s)\log u(s)\big] \quad \Leftrightarrow \quad 0 \geq \frac{1-s}{s}\frac{u'(s)}{u(s)} - \frac{1}{s^2}\log u(s).$$

Hence, the function $v(s) := \exp(\frac{1-s}{s}\log u(s))$ is non-increasing in s, and therefore we have $v(s) \leq \lim_{s\downarrow 0} v(s) =: v(0^+)$ for all $s \in (0, 1]$.

Note that

$$v(0^+) = \lim_{s\downarrow 0}\big(u(s)^{(1-s)/s}\big) = \lim_{s\downarrow 0}\Big(\int e^{sf^2/(2\tilde{\sigma}^2)}\,d\mu\Big)^{(1-s)/s} = \exp\Big(\frac{1}{2\tilde{\sigma}^2}\int f^2\,d\mu\Big).$$

Thus, for all $s \in (0, 1]$ we have

$$\begin{aligned} \exp\Big(\frac{1-s}{s}\log u(s)\Big) &\leq \exp\Big(\frac{1}{2\tilde{\sigma}^2}\int f^2\,d\mu\Big) \\ \Leftrightarrow \quad \int e^{sf^2/(2\tilde{\sigma}^2)}\,d\mu &\leq \exp\Big(\frac{1}{2\tilde{\sigma}^2}\frac{s}{1-s}\int f^2\,d\mu\Big). \end{aligned}$$

Setting $t = s/(2\tilde{\sigma}^2)$ completes the proof. □

Combining inequalities (4.1) and (4.4), we now get the following result.

Proposition 4.3. *Let μ be a probability measure on some measurable space $(\Omega, \mathcal{A})$ which satisfies the modified LSI (3.1) with Sobolev constant $\sigma^2 > 0$ with respect to $\mathfrak{d}$ and which moreover satisfies the modified LSI (3.1) with Sobolev constant $\tilde{\sigma}^2$ with respect to $\mathfrak{d}^+$. Furthermore, let $f : \Omega \to \mathbb{R}$ be a bounded measurable function such that $\int f\,d\mu = 0$ and $|\mathfrak{d}|\mathfrak{d}f|| \leq 1$. Then, we have*

$$\int \exp\Big(\frac{1}{2\sigma\tilde{\sigma}}f\Big)\,d\mu \leq \exp\Big(\frac{1}{2\tilde{\sigma}^2}\int |\mathfrak{d}f|^2\,d\mu\Big). \tag{4.6}$$

Proof. First, applying (4.1) to λf and then (4.4) with $t = \lambda^2\sigma^2$ for any $\lambda \in [0, \frac{1}{\sqrt{2}\sigma\tilde{\sigma}})$ and with f replaced by $|\mathfrak{d}f|$ leads to

$$\int e^{\lambda f}\,d\mu \leq \int e^{\lambda^2\sigma^2|\mathfrak{d}f|^2}\,d\mu \leq \exp\Big(\frac{\lambda^2\sigma^2}{1 - 2\sigma^2\tilde{\sigma}^2\lambda^2}\int |\mathfrak{d}f|^2\,d\mu\Big).$$

Setting $\lambda = \frac{1}{2\sigma\tilde{\sigma}}$ completes the proof. □

5. Relating first and second order difference operators

In order to remove the first order difference operator on the right-hand side of (4.6), we may now study relations of the form $\gamma \int |\mathfrak{d} f|^2 \, d\mu \leq \int \|\mathfrak{d}^{(2)} f\|_{\mathrm{HS}}^2 \, d\mu$ for some constant $\gamma > 0$. Note that due to (2.6), we may replace $\mathfrak{d}^{(2)} f$ by "Hoeffding" differences $\mathfrak{D}^{(2)} f$ on the right-hand side, which enables us to make use of the "harmonic analysis" arguments established in Section 2. Indeed, one of our main tools is the following lemma about partial integration and self-adjointness for difference operators and the discrete Laplacian $\mathfrak{L}$ defined on functions of independent random variables.

Lemma 5.1. *Let $(\Omega_i, \mathcal{A}_i, \mu_i)$ be probability spaces, and denote by $(\Omega, \mathcal{A}, \mu) := \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ their product. Let $\mathfrak{D} = (\mathfrak{D}_i)_i$ be the difference operator from (2.1), and let $\mathfrak{L}$ be the Laplacian as in (2.7). Then, for any $f, g \in L^2(\mu)$ we have*:

1.
$$\int (\mathfrak{D}_i f) g \, d\mu = \int f (\mathfrak{D}_i g) \, d\mu = \int (\mathfrak{D}_i f)(\mathfrak{D}_i g) \, d\mu.$$

2.
$$\int (\mathfrak{D} f) g \, d\mu = \int f (\mathfrak{D} g) \, d\mu,$$
where $\mathfrak{D}$ the integral has to be understood componentwise.

3.
$$\int (\mathfrak{L} f) g \, d\mu = \int f (\mathfrak{L} g) \, d\mu = \sum_{i \neq j} \int (\mathfrak{D}_{ij} f)(\mathfrak{D}_{ij} g) \, d\mu.$$

Proof. The proof is elementary. Note that in order to prove (2) and (3), we only need to check (1). Part 1 in turn follows from the fact that by Fubini's theorem, we have

$$\int g \Big(\int f \, d\mu_i \Big) d\mu = \int \Big(\int f \, d\mu_i \Big) \Big(\int g \, d\mu_i \Big) d\mu = \int f \Big(\int g \, d\mu_i \Big) d\mu.$$

For (3), note that we always have $\mathfrak{D}_{ij} f = \mathfrak{D}_{ji} f$ for any i, j by (2.1) and Fubini's theorem. □

Using this result, we can prove an inequality of the desired type.

Proposition 5.2. *Let $(\Omega_i, \mathcal{A}_i, \mu_i)$ be probability spaces, and denote by $(\Omega, \mathcal{A}, \mu) := \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ their product. Let $f \in L^2(\mu)$ be a function such that its Hoeffding decomposition with respect to μ is given by $f = \sum_{k=d}^n f_k$ for some $d \geq 2$. Then, we have*

$$\int |\mathfrak{d} f|^2 \, d\mu \leq \frac{1}{d-1} \int \|\mathfrak{d}^{(2)} f\|_{\mathrm{HS}}^2 \, d\mu.$$

Equality holds if $f = f_d$, that is, the Hoeffding decomposition of f consists of a single term only. Here, $\|\cdot\|_{\mathrm{HS}}$ denotes the Hilbert Schmidt norm of a matrix.

Proof. First, let $f = f_k$. Then, applying Lemma 5.1(3) leads to

$$\int \|\mathfrak{D}^{(2)} f_k\|_{\mathrm{HS}}^2 \, d\mu = \sum_{i \neq j} \int (\mathfrak{D}_{ij} f_k)(\mathfrak{D}_{ij} f_k) \, d\mu = \int f_k \mathfrak{L} f_k \, d\mu.$$

Moreover, Theorem 2.3 yields $\mathfrak{L} f_k = (k)_2 f_k$. Together with (2.6), this yields

$$\int \|\mathfrak{D}^{(2)} f_k\|_{\mathrm{HS}}^2 \, d\mu = (k)_2 \int f_k^2 \, d\mu. \tag{$*$}$$

On the other hand, if $X_1, \ldots, X_n$ is a sequence of independent random variables with distributions μ_i, $i = 1, \ldots, n$, we have $f_k(X_1, \ldots, X_n) = \sum_{i_1 < \cdots < i_k} h_{i_1 \cdots i_k}(X_{i_1}, \ldots, X_{i_k})$, where the summands on the right-hand side are pairwise orthogonal in L^2. Here, we used the notation of the proof of Theorem 2.3.

Now let $\bar{X}_1, \ldots, \bar{X}_n$ be a sequence of independent copies of the random variables $X_1, \ldots, X_n$, and additionally consider the functions $T_{i_j} h_{i_1 \cdots i_k}(X_{i_1}, \ldots, X_{i_d}) = h_{i_1 \cdots i_k}(X_{i_1}, \ldots, \bar{X}_{i_j}, \ldots, X_{i_k})$ (cf. Example 2.2(3)). Then,

$$\bigcup_{i_1 < \cdots < i_k} \big\{ h_{i_1 \cdots i_k}(X_{i_1}, \ldots, X_{i_k}) \big\} \cup \big\{ T_{i_j} h_{i_1 \cdots i_k}(X_{i_1}, \ldots, X_{i_k}), \, j = 1, \ldots, k \big\}$$

is still a (larger) family of pairwise orthogonal functions in L^2, now integrating with respect to the X_i and the $\bar{X}_i$.

Similarly to the deduction of (2.8), we therefore get

$$\begin{aligned} \big(\mathfrak{d}_i f_k(X_1, \ldots, X_n)\big)^2 &= \frac{1}{2} \bar{\mathbb{E}}_i (f_k - T_i f_k)^2 \\ &= \frac{1}{2} \bar{\mathbb{E}}_i \Bigg(\sum_{\substack{i_1 < \cdots < i_k \\ i \in \{i_1, \ldots, i_k\}}} \big(h_{i_1 \cdots i_k}(X_{i_1}, \ldots, X_{i_k}) - T_i h_{i_1 \cdots i_k}(X_{i_1}, \ldots, X_{i_k}) \big) \Bigg)^2. \end{aligned}$$

Using orthogonality, it follows that

$$\begin{aligned} &\mathbb{E}\big(\mathfrak{d}_i f_k(X_1, \ldots, X_n)\big)^2 \\ &\quad = \sum_{\substack{i_1 < \cdots < i_k \\ i \in \{i_1, \ldots, i_k\}}} \frac{1}{2} \big(\mathbb{E} \bar{\mathbb{E}}_i \big(h_{i_1 \cdots i_k}^2 (X_{i_1}, \ldots, X_{i_k}) + T_i h_{i_1 \cdots i_k}^2 (X_{i_1}, \ldots, X_{i_k}) \big) \big) \\ &\quad = \sum_{\substack{i_1 < \cdots < i_k \\ i \in \{i_1, \ldots, i_k\}}} \mathbb{E} h_{i_1 \cdots i_k}^2 (X_{i_1}, \ldots, X_{i_k}). \end{aligned}$$

As in the proof of Theorem 2.3, it remains to check how often each term $\mathbb{E} h_{i_1 \cdots i_k}^2 (X_{i_1}, \ldots, X_{i_k})$ appears in $\mathbb{E}|\mathfrak{d} f_k|^2 = \sum_i \mathbb{E}(\mathfrak{d}_i f_k)^2$. However, it is clear that each $i \in \{i_1, \ldots, i_k\}$ replicates the summand $\mathbb{E}h_{i_1\cdots i_k}(X_{i_1},\dots,X_{i_k})$ exactly once. Consequently, it follows that $\mathbb{E}|\mathfrak{d}f_k|^2 = k\mathbb{E}f_k^2$, or

$$\int |\mathfrak{d}f_k|^2\,d\mu = k\int f_k^2\,d\mu. \tag{$**$}$$

Comparing ($*$) and ($**$) completes the proof in case of $f = f_k$.

For functions with arbitrary Hoeffding expansion we shall use the orthogonality of the terms of the Hoeffding decomposition to get

$$\int |\mathfrak{d}f|^2\,d\mu = \sum_{k=d}^{n} \frac{1}{k-1}\int \|\mathfrak{D}^{(2)} f_k\|_{\mathrm{HS}}^2\,d\mu \le \frac{1}{d-1}\int \|\mathfrak{D}^{(2)} f\|_{\mathrm{HS}}^2\,d\mu.$$

In view of (2.6), this finally completes the proof. □

We are now ready to prove Theorem 1.2. In fact, using the results established in Sections 2–4, we may easily obtain some complementary results which can be shown along the lines of the proof of Theorem 1.2. For instance, we have the following slight sharpening of Theorem 1.2 if all the measures μ_i are Bernoulli measures.

Proposition 5.3. *Using the notations of Theorem 1.2, let all the μ_i be of the form $\mu_i = p_i\delta_{a_i} + (1-p_i)\delta_{b_i}$, where $a_i, b_i \in \mathbb{R}$, $p_i \in (0,1)$ for all i, and δ_x denotes the Dirac measure at $x \in \mathbb{R}$. Then, assuming the conditions of Theorem 1.2, we have*

$$\int \exp\Big(\frac{1}{3+2b^2}|f|\Big)\,d\mu \le 2.$$

We now prove give a joint proof of Theorem 1.2 and Proposition 5.3.

Proof of Theorem 1.2 and Proposition 5.3. First, combining Proposition 4.3, Proposition 5.2 with $d = 2$ and the assumptions from Theorem 1.2 leads to

$$\int \exp\Big(\frac{1}{2\sigma\tilde{\sigma}} f\Big)\,d\mu \le \exp\Big(\frac{1}{2\tilde{\sigma}^2}\int \|\mathfrak{d}^{(2)} f\|_{\mathrm{HS}}^2\,d\mu\Big) \le \exp\Big(\frac{b^2}{2\tilde{\sigma}^2}\Big) \tag{5.1}$$

if μ satisfies the modified LSI (3.1) with constant $\sigma^2 > 0$ with respect to $\mathfrak{d}$ and furthermore with constant $\tilde{\sigma}^2 > 0$ with respect to $\mathfrak{d}^+$. Now, from (5.1) we get

$$\int \exp\Big(\frac{1}{2\sigma\tilde{\sigma}}|f|\Big)\,d\mu \le \int \Big(\exp\Big(\frac{1}{2\sigma\tilde{\sigma}} f\Big) + \exp\Big(\frac{1}{2\sigma\tilde{\sigma}}(-f)\Big)\Big)\,d\mu \le 2\exp\Big(\frac{b^2}{2\tilde{\sigma}^2}\Big).$$

Thus, by applying Hölder's inequality we obtain

$$\int \exp\Big(\frac{1}{2\sigma\tilde{\sigma}\kappa}|f|\Big)\,d\mu \le \Big(\int \exp\Big(\frac{1}{2\sigma\tilde{\sigma}}|f|\Big)\,d\mu\Big)^{1/\kappa} \le \Big(2\exp\Big(\frac{b^2}{2\tilde{\sigma}^2}\Big)\Big)^{1/\kappa}$$

for all $\kappa \geq 1$. The last term is bounded by 2 if $\kappa \geq (\log 2 + b^2/(2\tilde{\sigma}^2))/\log 2$, or equivalently $1/(2\sigma\tilde{\sigma}\kappa) \leq \log 2/(2\sigma\tilde{\sigma}\log 2 + \sigma\tilde{\sigma}^{-1}b^2)$.

By Proposition 3.2, Proposition 3.3, Lemma 3.4 and Proposition 4.1, we can set $\sigma^2 = 2$ and $\tilde{\sigma}^2 = 4$ or, in the Bernoulli case, $\sigma^2 = 1$ and $\tilde{\sigma}^2 = 2$. We thus choose

$$\begin{aligned} &\int \exp\Big(\frac{\log 2}{\sqrt{32}\log 2 + \frac{1}{\sqrt{2}}b^2}|f|\Big)\,d\mu \leq 2, \\ &\int \exp\Big(\frac{\log 2}{\sqrt{8}\log 2 + \frac{1}{\sqrt{2}}b^2}|f|\Big)\,d\mu \leq 2 \end{aligned} \tag{5.2}$$

for $\sigma^2 = 2$ and $\tilde{\sigma}^2 = 4$ or $\sigma^2 = 1$ and $\tilde{\sigma}^2 = 2$, respectively. The proof of completed by noting that for all $x \geq 0$,

$$\frac{\log 2}{\sqrt{32}\log 2 + \frac{1}{\sqrt{2}}x} \geq \frac{1}{6+2x} \quad \text{and} \quad \frac{\log 2}{\sqrt{8}\log 2 + \frac{1}{\sqrt{2}}x} \geq \frac{1}{3+2x}.$$

□

Moreover, it is straightforward to reformulate Theorem 1.2 using $\mathfrak{d}^+$ instead of $\mathfrak{d}$. The proof is easily obtained by simple modifications of the above arguments:

Proposition 5.4. *Using the notations of Theorem 1.2, we require that $|\mathfrak{d}^+|\mathfrak{d}^+ f|| \leq 1$, where $\mathfrak{d}^+$ is the difference operator from (4.2). Then, we have*

$$\int \exp\Big(\frac{1}{2(4+b^2)}|f|\Big)\,d\mu \leq 2.$$

Proof. Note that we can use (4.1) with $\mathfrak{d}$ replaced by $\mathfrak{d}^+$, that is $\int e^f\,d\mu \leq \int e^{\tilde{\sigma}^2|\mathfrak{d}^+ f|^2}\,d\mu$ for any bounded measurable function $f\colon \Omega \to \mathbb{R}$ with $\int f\,d\mu = 0$. Proceeding as in Section 4 then leads to the inequality

$$\int \exp\Big(\frac{1}{2\tilde{\sigma}^2} f\Big)\,d\mu \leq \exp\Big(\frac{1}{2\tilde{\sigma}^2}\int |\mathfrak{d}^+ f|^2\,d\mu\Big)$$

if $f\colon \Omega \to \mathbb{R}$ is any bounded measurable function such that $\int f\,d\mu = 0$ and $|\mathfrak{d}^+|\mathfrak{d}^+ f|| \leq 1$. Since $\int |\mathfrak{d}^+ f|^2\,d\mu \leq \int |\mathfrak{d} f|^2\,d\mu$, we can now use Proposition 5.2 as well. The remaining part of the proof is similar to the proof of Theorem 1.2. Thus, we finally arrive at the inequality $1/(2\tilde{\sigma}^2\kappa) \leq \log 2/(2\tilde{\sigma}^2\log 2 + b^2)$. Plugging in $\tilde{\sigma}^2 = 4$ and noting that $\log 2/(8\log 2 + x) \geq 1/(8+2x)$ for all $x \geq 0$ completes the proof. □

Finally, we prove Theorem 1.3.

Proof of Theorem 1.3. The basic argument is as follows: if we have two functions φ_1 and φ_2 on $\mathbb{R}^n$ both satisfying $\int e^{c_i|\varphi_i|}\,d\mu \le 2$ for some constants $c_i > 0$, $i = 1, 2$, it follows that

$$\begin{aligned}\int e^{\min(c_1,c_2)|\varphi_1+\varphi_2|/2}\,d\mu &\le \int e^{c_1|\varphi_1|/2}e^{c_2|\varphi_2|/2}\,d\mu \\ &\le \Big(\int e^{c_1|\varphi_1|}\,d\mu\Big)^{1/2}\Big(\int e^{c_2|\varphi_2|}\,d\mu\Big)^{1/2} \le 2 \end{aligned}\tag{5.3}$$

due to the Cauchy–Schwarz inequality. In our situation, we set $\varphi_1 = f_1$ and $\varphi_2 = Rf$.

The bound for Rf is obvious by Theorem 1.2 and the fact that $\mathfrak{D}_{ij}Rf = \mathfrak{D}_{ij}f$ for all $i \neq j$ in view of (2.9). This leads to $c_2 = 1/(6 + 2b^2)$. It remains to bound f_1. Here, inequality (4.1) yields

$$\int e^{\lambda f_1}\,d\mu \le \int e^{\sigma^2\lambda^2|\mathfrak{d}f_1|^2}\,d\mu \le e^{\sigma^2\lambda^2 b_0^2}$$

for any $\lambda > 0$, thus $\int e^{\lambda|f_1|}\,d\mu \le 2e^{\sigma^2\lambda^2 b_0^2}$. As in the proof of Theorem 1.2, it follows that

$$\int e^{\lambda|f_1|/\kappa}\,d\mu \le \big(2e^{\sigma^2\lambda^2 b_0^2}\big)^{1/\kappa}$$

for all $\kappa \ge 1$. The right-hand side is bounded by 2 if $\lambda/\kappa \le \lambda\log 2/(\log 2 + \lambda^2\sigma^2 b_0^2)$. Here, the expression on the right-hand side attains a maximum at $\lambda = (\log 2)^{1/2}/(\sigma b_0)$ whose value is $(\log 2)^{1/2}/(2\sigma b_0)$. Plugging in $\sigma^2 = 2$, we get $c_1/2 = (\log 2)^{1/2}/(4\sqrt{2}b_0) \ge 1/(7b_0)$, and hence we can estimate $\min(c_1, c_2)/2$ as stated in Theorem 1.3. □

6. Evaluating second order difference operators

In Theorem 1.2, checking the condition $\int \|\mathfrak{d}^{(2)}f\|_{\mathrm{HS}}^2\,d\mu \le b^2$ is typically straightforward, once we know the Hoeffding decomposition of f (cf. (2.6), enabling us to use the "Hoeffding" differences $\mathfrak{D}$). In contrast, evaluating the condition $|\mathfrak{d}|\mathfrak{d}f|| \le 1$ tends to be more involved. Therefore, we shall provide a reformulated version of Theorem 1.2 with conditions which are easier to apply.

Theorem 6.1. *Let $(\Omega_i, \mathcal{A}_i, \mu_i)$ be probability spaces, and denote by $(\Omega, \mathcal{A}, \mu) := \bigotimes_{i=1}^n(\Omega_i, \mathcal{A}_i, \mu_i)$ their product. Moreover, let $f\colon \Omega \to \mathbb{R}$ be a bounded measurable function so that its Hoeffding decomposition with respect to μ is given by $f = \sum_{k=2}^n f_k$. Assume that the conditions*

$$\big\|\mathfrak{d}^{(2)}f\big\|_{\mathrm{HS}} \le B_1 \quad \text{and} \quad \max_{i=1,\dots,n}|\mathfrak{d}_i f| \le B_2 \tag{6.1}$$

are satisfied for some $B_1, B_2 \ge 0$, where $\|\mathfrak{d}^{(2)}f\|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt norm of $\mathfrak{d}^{(2)}f$. Then, we have

$$\int \exp\Big(\frac{c}{B_1 + B_2}|f|\Big)\,d\mu \le 2$$

for some numerical constant $c > 0$. A possible choice is $c = 1/11$. If all the underlying measures μ_i are two-point measures, we can take $c = 1/7$.

Proof. For a set of independent random variables $X_1, \ldots, X_n$ with distributions μ_i, write

$$\big|\mathfrak{d}\big|\mathfrak{d} f(X)\big|\big| = \left(\sum_{i=1}^n \frac{1}{2}\bar{\mathbb{E}}_i\big(\big|\mathfrak{d} f(X)\big| - \big|T_i \mathfrak{d} f(X)\big|\big)^2\right)^{1/2} \tag{6.2}$$

with $X = (X_1, \ldots, X_n)$ and T_k as in Remark 2.2(3). Without loss of generality, we may assume that $|\mathfrak{d} f| \neq 0$. To simplify notation, we introduce the convention that $\sum^{(j)}$ means summation extending over all indexes but j. Similarly, $\sum^{(j,k)}$ denotes summation over all indexes but j and k. Now, setting $a := \sum_{j=1}^{n}{}^{(i)} (\mathfrak{d}_j f)^2$, $b := (\mathfrak{d}_i f)^2$, $c := \sum_{j=1}^{n}{}^{(i)} (T_i \mathfrak{d}_j f)^2$ and $d := (T_i \mathfrak{d}_i f)^2$ for any $1 \le i \le n$, we arrive at

$$\begin{aligned}
\big(|\mathfrak{d} f| - |T_i \mathfrak{d} f|\big)^2 &= (\sqrt{a+b} - \sqrt{c+d})^2 = \left(\frac{a+b-c-d}{\sqrt{a+b} + \sqrt{c+d}}\right)^2 \\
&\le \left(|\sqrt{a} - \sqrt{c}| + \frac{|b-d|}{\sqrt{a+b}}\right)^2 \le 2\left((\sqrt{a} - \sqrt{c})^2 + \frac{(b-d)^2}{a+b}\right).
\end{aligned} \tag{6.3}$$

(Using the simpler estimate $|\sqrt{a+b} - \sqrt{c+d}| \le |\sqrt{a} - \sqrt{c}| + |\sqrt{b} - \sqrt{d}|$ instead would essentially lead to a condition on first order differences only.) Moreover,

$$\begin{aligned}
(\sqrt{a} - \sqrt{c})^2 &= \left(\left(\sum_{j=1}^n{}^{(i)} (\mathfrak{d}_j f)^2\right)^{1/2} - \left(\sum_{j=1}^n{}^{(i)} (T_i \mathfrak{d}_j f)^2\right)^{1/2}\right)^2 \\
&\le \sum_{j=1}^n{}^{(i)} (\mathfrak{d}_j f - T_i \mathfrak{d}_j f)^2 \\
&= \frac{1}{2}\sum_{j=1}^n{}^{(i)} \big(\big(\bar{\mathbb{E}}_j (f - T_j f)^2\big)^{1/2} - \big(\bar{\mathbb{E}}_j (T_i f - T_{ij} f)^2\big)^{1/2}\big)^2 \\
&\le \frac{1}{2}\sum_{j=1}^n{}^{(i)} \bar{\mathbb{E}}_j (f - T_j f - T_i f + T_{ij} f)^2.
\end{aligned} \tag{6.4}$$

Combining (6.2), (6.3) and (6.4) together with the trivial estimate $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$ for all $x, y \ge 0$ then yields

$$\big|\mathfrak{d}\big|\mathfrak{d} f(X)\big|\big| \le \sqrt{2}\left(\big\|\mathfrak{d}^{(2)} f(X)\big\|_{\mathrm{HS}} + \left(\frac{1}{2}\sum_{i=1}^n \bar{\mathbb{E}}_i \frac{((\mathfrak{d}_i f(X))^2 - (T_i \mathfrak{d}_i f(X))^2)^2}{|\mathfrak{d} f(X)|^2}\right)^{1/2}\right). \tag{6.5}$$

We may further estimate the last term by

$$\Big(\sum_{i=1}^n \bar{\mathbb{E}}_i \frac{|(\mathfrak{d}_i f(X))^2 - (T_i \mathfrak{d}_i f(X))^2|}{|\mathfrak{d} f(X)|^2}\Big)^{1/2} \sup_{x\in \mathrm{supp}(\mu)} \max_{i=1,\ldots,n} \big|\mathfrak{d}_i f(x)\big|. \tag{6.6}$$

We now claim that

$$\Big(\sum_{i=1}^n \bar{\mathbb{E}}_i \frac{|(\mathfrak{d}_i f(X))^2 - (T_i \mathfrak{d}_i f(X))^2|}{|\mathfrak{d} f(X)|^2}\Big)^{1/2} \le 1. \tag{6.7}$$

To see this, recall that by (2.3), $(\mathfrak{d}_i f(X))^2 = ((\mathfrak{D}_i f(X))^2 + \mathbb{E}_i(\mathfrak{D}_i f(X))^2)/2$, and therefore

$$\big|\big(\mathfrak{d}_i f(X)\big)^2 - \big(T_i \mathfrak{d}_i f(X)\big)^2\big| \le \big(\big(\mathfrak{D}_i f(X)\big)^2 + \big(T_i \mathfrak{D}_i f(X)\big)^2\big)/2.$$

Taking expectations yields $\bar{\mathbb{E}}_i |(\mathfrak{d}_i f(X))^2 - (T_i \mathfrak{d}_i f(X))^2| \le (\mathfrak{d}_i f(X))^2$, which proves (6.7).

Combining (6.5), (6.6) and (6.7) with the assumptions from the theorem, we therefore arrive at $|\mathfrak{d}|\mathfrak{d} f|| \le \sqrt{2} B_1 + B_2$. Moreover, by (2.5), we have $(\mathfrak{D}_{ij} f(x))^2 \le 4(\mathfrak{d}_{ij} f(x))^2$ and hence

$$\int \|\mathfrak{D}^{(2)} f\|_{\mathrm{HS}}^2 \, d\mu \le 4B_1^2. \tag{$*$}$$

Finally, consider the "normalized" function $f/(\sqrt{2}B_1 + B_2)$ and use ($*$) in (5.2) from the proof of Theorem 1.2, respectively. The proof of Theorem 6.1 then follows by elementary computations. □

As for conditions (6.1), note that in typical cases (for instance, if the function f is symmetric) we have $B_1 = \Theta(B_2)$ as $n \to \infty$.

For functions of independent Rademacher variables taking values in $\{\pm 1\}$, we don't seem to need first order differences. It is well-known that such functions can be represented in the form

$$f(X_1, \ldots, X_n) = \alpha_0 + \sum_{i=1}^n \alpha_i X_i + \sum_{i<j} \alpha_{ij} X_i X_j + \cdots, \tag{6.8}$$

where the coefficients α_I (with a suitable multi-index I) are real numbers and the summation extends over all terms up to the order n. More precisely, we have $\alpha_{i_1\cdots i_d} = \mathbb{E} f(X_1, \ldots, X_n) X_{i_1} \cdots X_{i_d}$ for any $i_1 < \cdots < i_d$, $d = 0, 1, \ldots, n$. This representation is called the *Fourier–Walsh expansion* of the function f, and the expression on the right-hand side of (6.8) is also known as a *Rademacher chaos*. It is immediately clear that (6.8) is at the same time the Hoeffding decomposition of f. Applying Corollary 5.3 to functions of this type leads to the following result.

Proposition 6.2. *Let μ be the product measure of n symmetric Bernoulli distributions $\mu_i = \frac{1}{2}\delta_{+1} + \frac{1}{2}\delta_{-1}$ on $\{\pm 1\}$, and define $f\colon \mathbb{R}^n \to \mathbb{R}$ by $f(x_1, \ldots, x_n) := \sum_{i<j} \alpha_{ij} x_i x_j +$*

$\sum_{i<j<k} \alpha_{ijk} x_i x_j x_k + \cdots$, *where the sum goes up to order* n *and the* $\alpha_{i_1 \cdots i_d}$ *are any real numbers. Set* $B := \sup_{x \in \{\pm 1\}^n} \|\mathfrak{D}^{(2)} f(x)\|_{\mathrm{HS}}$ *with* $\mathfrak{D}^{(2)} f(x)$ *as in* (2.2). *Then, we have*

$$\int \exp\Big(\frac{1}{5B}|f|\Big)\, d\mu \le 2.$$

Proof. First, note that similarly to Remark 2.2(1), for products of symmetric Bernoulli distributions we have $\mathfrak{d}_{ij} f = |\mathfrak{D}_{ij} f|$ for any $i \neq j$ and consequently $\|\mathfrak{d}^{(2)} f\|_{\mathrm{HS}} = \|\mathfrak{D}^{(2)} f\|_{\mathrm{HS}}$. Therefore, in view of Corollary 5.3, it suffices to prove that $|\mathfrak{d}|\mathfrak{d} f|| \le \|\mathfrak{d}^{(2)} f\|_{\mathrm{HS}}$ on $\mathrm{supp}(\mu)$.

To this end, note that for any $i = 1, \ldots, n$, by the fact that $T_i|\mathfrak{d} f| = |T_i \mathfrak{d} f|$ and the reverse triangular inequality,

$$\big(\mathfrak{d}_i |\mathfrak{d} f|\big)^2 = \frac{1}{2}\bar{\mathbb{E}}_i\big(|\mathfrak{d} f| - |T_i \mathfrak{d} f|\big)^2 \le \frac{1}{2}\bar{\mathbb{E}}_i |\mathfrak{d} f - T_i \mathfrak{d} f|^2. \tag{6.9}$$

Here, the difference $\mathfrak{d} f - T_i \mathfrak{d} f$ is defined componentwise. Using the Fourier–Walsh expansion (6.8) and the fact that $x_i^2 = 1$ on $\mathrm{supp}(\mu)$, it is easy to see that $T_i \mathfrak{d}_i f = \mathfrak{d}_i f$. Therefore, using the notations from the proof of Theorem 6.1,

$$\begin{aligned} |\mathfrak{d} f - T_i \mathfrak{d} f|^2 &= \frac{1}{2}\sum_{j=1}^{n}{}^{(i)} \Big(\big(\bar{\mathbb{E}}_j (f - T_j f)^2\big)^{1/2} - \big(\bar{\mathbb{E}}_j (T_i f - T_{ij} f)^2\big)^{1/2}\Big)^2 \\ &\le \frac{1}{2}\sum_{j=1}^{n}{}^{(i)} \bar{\mathbb{E}}_j (f - T_j f - T_i f + T_{ij} f)^2. \end{aligned} \tag{6.10}$$

Here, the last step follows from the reverse triangular inequality again (for the norm $(\bar{\mathbb{E}}_j(\cdot)^2)^{1/2}$). Combining (6.9) and (6.10) and summing over $i = 1, \ldots, n$ finishes the proof. □

7. Differentiable functions: Proofs

In order to prove Theorem 1.4, we need to adapt some of the elements of the proof of Theorem 1.2 from the previous sections. For that, if (M, d) is a metric space and $f : M \to \mathbb{R}$ is a continuous function, we may define the generalized modulus of the gradient by

$$\big|\nabla^* f(x)\big| = \limsup_{y \to x} \frac{|f(x) - f(y)|}{d(x, y)} \tag{7.1}$$

for any $x \in M$, where the limsup is assigned to be zero at isolated points. By the continuity of f, $x \mapsto |\nabla^* f(x)|$ is a Borel-measurable function. If f is a differentiable function on some open subset $G \subset \mathbb{R}^n$, the generalized modulus of the gradient agrees with the Euclidean norm of the usual gradient. We may iterate the generalized modulus of the gradient by setting for any $x \in M$

$$\big|\nabla^* \big|\nabla^* f(x)\big|\big| := \limsup_{y \to x} \frac{||\nabla^* f(x)| - |\nabla^* f(y)||}{d(x, y)}. \tag{7.2}$$

Using the generalized modulus of the gradient, we have the following analogues of inequalities (4.1) and (4.4) from Section 4. Let (M, d) be a metric space, equipped with some Borel probability measure μ which satisfies a logarithmic Sobolev inequality with constant σ^2. Moreover, let $u\colon M \to \mathbb{R}$ be a μ-integrable locally Lipschitz function. Then, we have

$$\int e^{u-\int u\,d\mu}\,d\mu \le \int e^{\sigma^2|\nabla^* u|^2}\,d\mu. \tag{7.3}$$

Moreover, if we additionally require $|\nabla^* u| \le 1$, we have

$$\int e^{tu^2}\,d\mu \le \exp\Big(\frac{t}{1-2\sigma^2 t}\int u^2\,d\mu\Big) \tag{7.4}$$

for any $0 \le t < 1/(2\sigma^2)$. As mentioned in Section 4, (7.3) and (7.4) are due to [6] and [2].

Now consider $M = G$, where $G \subset \mathbb{R}^n$ is some open subset equipped with the Euclidean metric. By proceeding as in the proof of Proposition 4.3, we arrive at the following exponential moment inequality.

Proposition 7.1. *Let $G \subset \mathbb{R}^n$ be some open set, and let μ be a probability measure on $(G, \mathcal{B}(G))$ which satisfies the logarithmic Sobolev inequality* (1.8) *with Sobolev constant $\sigma^2 > 0$. Furthermore, let $f\colon G \to \mathbb{R}$ be a locally Lipschitz μ-integrable function with μ-mean zero such that $|\nabla^* f|$ is locally Lipschitz and $|\nabla^*|\nabla^* f|| \le 1$. Here, $|\nabla^* f|$ is the generalized modulus of the gradient from* (7.1). *Then, we have*

$$\int_G \exp\Big(\frac{1}{2\sigma^2} f\Big)\,d\mu \le \exp\Big(\frac{1}{2\sigma^2}\int_G |\nabla^* f|^2\,d\mu\Big).$$

Proposition 7.1 is a special case of [5], Proposition 2.1. If f is a $\mathcal{C}^2$-function, the condition $|\nabla^*|\nabla^* f|| \le 1$ can be simplified by the following lemma.

Lemma 7.2. *Let $G \subset \mathbb{R}^n$ be some open set. Then, for any $\mathcal{C}^2$-smooth function $f\colon G \to \mathbb{R}$, the function $|\nabla^* f|$ is locally Lipschitz and satisfies*

$$\big|\nabla^*\big|\nabla^* f(x)\big|\big| \le \big\|f''(x)\big\|_{\mathrm{Op}}$$

at all points $x \in G$, where $f''(x)$ denotes the Hessian of f at $x \in G$.

Proof. By chain rule, $|\nabla f(x)|$ is differentiable on $\{|\nabla f(x)| \neq 0\}$ with $\nabla|\nabla f(x)| = \frac{1}{|\nabla f(x)|} \times f''(x)\nabla f(x)$, which immediately yields the desired result if $|\nabla f(x)| \neq 0$.

It remains to consider the case $|\nabla f(x)| = 0$. Here, for any $v \in \mathbb{R}^n$ such that $|v| = 1$, by Taylor expansion we obtain $\langle \nabla f(x+h), v\rangle = \langle f''(x)v, h\rangle + o(|h|)$ as $h \to 0$. Here, the o-term can be bounded by a quantity which does not depend on the choice of v. Therefore, dividing by $|h|$ and taking limits according to (7.1), the proof is finished by noting that

$$\big|\nabla^*|\nabla f(x)|\big| = \limsup_{h\to 0} \frac{|\nabla f(x+h)|}{|h|} \le \sup\Big\{\Big\langle f''(x)v, \frac{h}{|h|}\Big\rangle \colon |v|=1, h\neq 0\Big\}$$
$$= \|f''(x)\|_{\mathrm{Op}}. \qquad \square$$

We can now prove Theorems 1.4 and 1.5:

Proof of Theorem 1.4. Given a function f as in Theorem 1.4, applying Proposition 7.1 together with Lemma 7.2 yields

$$\int_G \exp\Big(\frac{1}{2\sigma^2} f\Big)\, d\mu \le \exp\Big(\frac{1}{2\sigma^2}\int_G |\nabla f|^2\, d\mu\Big). \tag{7.5}$$

Since μ satisfies a logarithmic Sobolev inequality with constant σ^2, it also satisfies a Poincaré inequality (1.7) with constant σ^2. Therefore, since $\int_G \partial_i f\, d\mu = 0$ for all i, we have $\int_G (\partial_i f)^2\, d\mu \le \sigma^2 \sum_{j=1}^n \int_G (\partial_{ij} f)^2\, d\mu$ for all $i = 1, \ldots, n$, where $\partial_{ij} f(x) = \frac{d^2 f(x)}{dx_i dx_j}$. Summing up over all i, we get

$$\int_G |\nabla f|^2\, d\mu \le \sigma^2 \int_G \|f''\|_{\mathrm{HS}}^2\, d\mu. \tag{7.6}$$

Combining (7.5), (7.6) and the assumptions from Theorem 1.4, we arrive at

$$\int_G \exp\Big(\frac{1}{2\sigma^2} f\Big)\, d\mu \le \exp\Big(\frac{1}{2}\int_G \|f''\|_{\mathrm{HS}}^2\, d\mu\Big) \le \exp\Big(\frac{b^2}{2}\Big).$$

The rest of the proof is similar to the proof of Theorem 1.2. We finally arrive at the inequality $1/(2\sigma^2\kappa) \le \log 2/(2\sigma^2 \log 2 + b^2\sigma^2)$. Noting that $\log 2/(2\sigma^2 \log 2 + x\sigma^2) \ge 1/(2\sigma^2(1+x))$ for all $x \ge 0$ finishes the proof. $\square$

Proof of Theorem 1.5. The proof is similar to the proof of Theorem 1.3 assuming condition (i) from the latter theorem. Setting $\mu[h] = \int_G h\, d\mu$ for any $h \in L^1(\mu)$, write $f = \varphi_1 + \varphi_2$ with $\varphi_1(x) = \sum_{i=1}^n \mu[\partial_i f](x_i - \mu[x_i])$, $\varphi_2(x) = f(x) - \varphi_1(x)$. We now apply the basic argument (5.3) from the proof of Theorem 1.3. Here we need to check that $\int e^{c_i|\varphi_i|}\, d\mu \le 2$ for $i = 1, 2$ and some constants $c_1, c_2 > 0$. By Theorem 1.4 applied to φ_2, we may choose $c_2/2 = 1/(4\sigma^2(1+b^2))$.

For estimating the function φ_1, note that $|\nabla\varphi_1|^2 = \sum_{i=1}^n (\mu[\partial_i f])^2 \le \sigma^2 b_0^2$ by assumption. Therefore, applying (7.3) yields

$$\int e^{\lambda\varphi_1}\, d\mu \le \int e^{\sigma^2\lambda^2|\nabla\varphi_1|^2}\, d\mu \le e^{\sigma^4\lambda^2 b_0^2}$$

for all $\lambda > 0$. Proceeding as in the proof of Theorem 1.3, we obtain $c_1/2 = (\log 2)^{1/2}/(4\sigma^2 b_0) \ge 1/(5\sigma^2 b_0)$, which easily yields the desired result. $\square$

8. Applications

8.1. Functions of independent random variables

As a first example, we shall consider a certain type of statistics related to Hoeffding-type expansions. In detail, for any $n \in \mathbb{N}$, let $X_1, \dots, X_n$ be independent random variables and T_n a statistic of the form

$$T_n(X_1, \dots, X_n) = h_{0,n} + \sum_i h_{1,n}(X_i)n^{-1} + \sum_{i<j} h_{2,n}(X_i, X_j)n^{-2} \\ + \sum_{i<j<k} h_{3,n}(X_i, X_j, X_k)n^{-3} + \cdots . \tag{8.1}$$

Here, $h_{d,n}$, $d = 0, 1, \dots, n$, are some "kernel" functions which are completely degenerate with respect to the X_i. Usually, we then have concentration inequalities of the form $P(\sqrt{n}(T_n - h_{0,n}) \geq t) \leq e^{-ct^2}$, where c is some absolute constant. Using second order concentration, it is possible to sharpen these bounds. Here we mainly use the results from Section 6.

Example 8.1. Let $X_1, \dots, X_n$ be some independent random variables, and let T_n be a statistic of the form (8.1). Assume we have

$$\|n\mathfrak{d}^{(2)} T_n\|_{\mathrm{HS}} \leq M \quad \text{and} \quad \left| n\mathfrak{d}_i\left(T_n - \sum_i h_{1,n}(X_i)n^{-1}\right)\right| \leq M \quad \forall i \tag{8.2}$$

for some universal constant M and with $\mathfrak{d}^{(2)}T_n$ as in (1.4). Then, there exists some numerical constant $c > 0$ such that

$$P\left(n\left|T_n - h_{0,n} - \sum_i h_{1,n}(X_i)n^{-1}\right| \geq t\right) \leq 2e^{-ct/M}.$$

This follows immediately from Theorem 6.1. In particular, conditions (8.2) are satisfied if $\|h_{d,n}\|_\infty \equiv \sup_x |h_{d,n}(x)| \leq L$ for $d \leq m$ and $h_{d,n} \equiv 0$ for all $d \geq m$, where $m \in \mathbb{N}$ is independent of n and where L is some absolute constant.

A special case is given by functions on the discrete cube, that is, we assume $X_1, \dots, X_n$ to be i.i.d. random variables with distributions $\mu_i = \frac{1}{2}\delta_{+1} + \frac{1}{2}\delta_{-1}$. In this situation, by Proposition 6.2, we may replace conditions (8.2) by the single condition $\|n\mathfrak{D}^{(2)}T_n\|_{\mathrm{HS}} \leq M$. Here, $\mathfrak{D}^{(2)}T_n$ is the "Hessian" of T_n with respect to $\mathfrak{D}$ defined in (2.2). For instance, if $T_n(X_1, \dots, X_n) = \alpha_0 + \sum_i n^{-1}\alpha_i X_i + \sum_{i<j} n^{-2}\alpha_{ij}X_iX_j$ for real numbers α_0, α_i, α_{ij}, then $\|n\mathfrak{D}^{(2)}T_n\|_{\mathrm{HS}} \leq M$ just means $n^{-1}(2\sum_{i<j}\alpha_{ij}^2)^{1/2} \leq M$.

As a second example, we shall consider additive functionals of partial sums, that is, functionals of the form

$$S_f := S_f(X) := \sum_{i=1}^n f\left(\sum_{j=1}^i X_j\right). \tag{8.3}$$

Random variables of this kind appear for example, as additive functionals of random walks (cf. [11]). Here we obtain the following result.

Example 8.2. Let $X_1, \ldots, X_n$ be a set of independent random variables, and let $f\colon \mathbb{R} \to \mathbb{R}$ be a bounded measurable function. Consider $S_f = S_f(X)$ as defined in (8.3). Then, there exists some numerical constant $c > 0$ such that for any $t \geq 0$,

$$P\big(|S_f - \mathbb{E}S_f| \geq t\big) \leq 2\exp\Big(-c\frac{t}{n^2\|f\|_\infty}\Big).$$

Proof. The proof is obtained by combining Theorem 1.3 and Theorem 6.1. For that, we simply have to calculate the respective differences of first and second order.

To start, note that the first order Hoeffding term of $S_f(X)$ is given by

$$S_f^1(X) = \sum_{\nu=1}^n \Big(\sum_{i \geq \nu} \Big(\mathbb{E}^{(\nu)} f\Big(\sum_{j=1}^i X_j\Big) - \mathbb{E} f\Big(\sum_{j=1}^i X_j\Big)\Big)\Big),$$

where $\mathbb{E}^{(\nu)}$ denotes taking the expectation with respect to all the random variables $X_1, \ldots, X_n$ but X_ν. It follows that for any $\nu = 1, \ldots, n$,

$$\big(\mathfrak{d}_\nu S_f^1(X)\big)^2 = \frac{1}{2}\bar{\mathbb{E}}_\nu\Big(\sum_{i\geq\nu}\Big(\mathbb{E}^{(\nu)} f\Big(\sum_{j=1}^i X_j\Big) - \mathbb{E}^{(\nu)} f\Big(\sum_{j=1}^i T_\nu X_j\Big)\Big)\Big)^2 \leq 2\|f\|_\infty^2 (n-\nu+1)^2,$$

and consequently $|\mathfrak{d}S_f^1(X)|^2 \leq 2\|f\|_\infty^2 \sum_{\nu=1}^n (n-\nu+1)^2 = \frac{1}{3}n(n+1)(2n+1)\|f\|_\infty^2$.

Next, we need to check the second order conditions from Theorem 1.3, that is, (1.5) for $RS_f^1(X) := S_f(X) - S_f^1(X) - \mathbb{E}S_f(X)$ (noting that $\mathfrak{d}^{(2)}RS_f^1(X) = \mathfrak{d}^{(2)}S_f^1(X)$). As shown in (the proof of) Theorem 6.1, these conditions can be replaced by (6.1). To see the first condition in (6.1), for any $\nu \neq \mu$,

$$\begin{aligned}
&\big(\mathfrak{d}_{\nu\mu} RS_f(X)\big)^2 \\
&\quad = \big(\mathfrak{d}_{\nu\mu} S_f(X)\big)^2 \\
&\quad = \frac{1}{4}\bar{\mathbb{E}}_{\nu\mu}\Big(\sum_{i\geq\nu\vee\mu}\Big(f\Big(\sum_{j=1}^i X_j\Big) - T_\nu f\Big(\sum_{j=1}^i X_j\Big) - T_\mu f\Big(\sum_{j=1}^i X_j\Big) + T_{\nu\mu} f\Big(\sum_{j=1}^i X_j\Big)\Big)\Big)^2 \\
&\quad \leq 4\|f\|_\infty^2\big(n - (\nu\vee\mu) + 1\big)^2
\end{aligned}$$

using similar arguments as above, and therefore $\|\mathfrak{d}^{(2)}S_f(X)\|_{\mathrm{HS}}^2 = \sum_{\nu\neq\mu}(\mathfrak{d}_{\nu\mu}S_f(X))^2 \leq Cn^4\|f\|_\infty^2$ for some numerical constant $C > 0$. Moreover, to see the second condition in (6.1), for any $\nu = 1, \dots, n$,

$$\begin{aligned}
&\big(\mathfrak{d}_\nu\big(S_f(X) - S_f^1(X) - \mathbb{E}S_f(X)\big)\big)^2 \\
&\quad = \frac{1}{2}\bar{\mathbb{E}}_\nu\left(\sum_{i\ge\nu}\left(f\left(\sum_{j=1}^i X_j\right) - T_\nu f\left(\sum_{j=1}^i X_j\right) - \mathbb{E}^{(\nu)} f\left(\sum_{j=1}^i X_j\right) + \mathbb{E}^{(\nu)} T_\nu f\left(\sum_{j=1}^i X_j\right)\right)\right)^2 \\
&\quad \le 8\|f\|_\infty^2 (n-\nu+1)^2.
\end{aligned}$$

Combining these estimates we easily arrive at the result. □

We may furthermore apply our results in the context of bootstrap methods. Suppose $X_1, \dots, X_n, \dots$ are random elements taking values in $\mathbb{R}^p$ (or some other separable metric space) which are independent and identically distributed from some distribution $P \in \mathcal{P}_0$. Here, $\mathcal{P}_0$ is a set of probability measures on $\mathbb{R}^p$ which contains all discrete measures. By $\hat{P}_n$ we denote the empirical measure of the first n observations. Let $T_n \equiv T_n(X_1, \dots, X_n; P) \equiv T_n(\hat{P}_n; P)$ be a sequence of symmetric statistics which may depend on the distribution P, and let h be a bounded real function defined on the range of T_n.

Here we are interested in estimating $\theta_n(P) := \mathbb{E}_P h(T_n(X_1, \dots, X_n; P))$. Given $X_1, \dots, X_n$, Efron's (nonparametric) bootstrap suggests to estimate $\theta_n(P)$ by $\theta_n(\hat{P}_n)$. That is, if we set

$$B_n(P) = \frac{1}{n^n} \sum_{i_1,\dots,i_n=1}^{n} h\big(T_n(X_{i_1}, \dots, X_{i_n}; P)\big),$$

Efron's bootstrap is given by $B_n(\hat{P}_n)$. In many situations, this bootstrap can be successfully applied, but in a number of examples (in particular due to bias problems) it fails asymptotically. These problems have been addressed by D.N. Politis and J.P. Romano [27], F. Götze [16] and P.J. Bickel, F. Götze and W.R. van Zwet [4] by introducing the m out of n bootstraps, that is, sampling from an i.i.d. sample of size n m-times independently with or without replacement. For instance, in the case of sampling without replacement (also called the $\binom{n}{m}$ bootstrap), we consider

$$J_{m,n}(P) = J_m(P) = \frac{1}{\binom{n}{m}} \sum_{i_1<\dots<i_m} h\big(T_m(X_{i_1}, \dots, X_{i_m}; P)\big)$$

as an estimator of $\theta_n(P)$. Then, the $\binom{n}{m}$ bootstrap estimator of $J_{m,n}(P)$ is given by $J_{m,n}(\hat{P}_n)$.

In order to access the accuracy of this estimate, one would have to estimate the error involved in replacing P by $\hat{P}_n$ in the functional $J_{m,n}(P)$. Under sufficient smoothness conditions for the dependence on P, this would lead to first or second order Hoeffding expansions involving a kernel of $m+1$ or $m+2$ variables, respectively. This would be necessary for evaluating the bias term of this bootstrap estimator. For the sake of brevity, we shall consider the estimate of the variance term for the original P only at this point.

At first order, the variance part of the error has been estimated in [27] and [16]. For instance, by [4], Theorem 1, if $\frac{m}{n} \to 0$, $m \to \infty$, we have $J_m(P) = \theta_m(P) + \mathcal{O}_P((m/n)^{1/2})$. Knowing (or at least estimating) the first order Hoeffding term of $J_m(P)$, we may sharpen this result by the second order results in this paper:

Proposition 8.3. *Suppose $\frac{m}{n} \to 0, m \to \infty$. Let $h = \sum_{i=0}^m h_i$ be the Hoeffding decomposition of $h = h(T_m(X_1, \ldots, X_m; P))$, and assume that*

$$|\mathfrak{d}_{ij} h| \leq \frac{c_1}{m}, \qquad \big|\mathfrak{d}_i(h - h_0 - h_1)\big| \leq c_2 \tag{8.4}$$

for all $1 \leq i < j \leq m$ and all $i = 1, \ldots, m$, respectively, where c_1 and c_2 are some absolute constants. Let $J_{m,1}(P)$ denote the first order Hoeffding term of $J_m(P)$. Then, we have

$$J_m(P) = \theta_m(P) + J_{m,1}(P) + \mathcal{O}_P\Big(\frac{m}{n}\Big).$$

Proof. Noting that $\mathbb{E}_P J_m(P) = \theta_m(P)$, let us check the conditions from Theorem 6.1 for $RJ_m(P) := J_m(P) - \theta_m(P) - J_{m,1}(P)$ (cf. (1.6)). Using $\mathfrak{d}^{(2)} RJ_m(P) = \mathfrak{d}^{(2)} J_m(P)$ and (8.4), by elementary counting arguments we obtain

$$\begin{aligned}
\big\|\mathfrak{d}^{(2)} J_m(P)\big\|_{\mathrm{HS}} &= \Bigg(\sum_{j \neq k \leq n} \Bigg(\mathfrak{d}_{jk} \frac{1}{\binom{n}{m}} \sum_{\substack{i_1 < \cdots < i_m \\ j,k \in \{i_1, \ldots, i_m\}}} h(X_{i_1}, \ldots, X_{i_m}; P)\Bigg)^2\Bigg)^{1/2} \\
&\leq \Bigg(\sum_{j \neq k \leq n} \frac{\binom{n-2}{m-2}}{\binom{n}{m}^2} \sum_{\substack{i_1 < \cdots < i_m \\ j,k \in \{i_1, \ldots, i_m\}}} \big(\mathfrak{d}_{jk} h(X_{i_1}, \ldots, X_{i_m}; P)\big)^2\Bigg)^{1/2} \leq c_1 \frac{m}{n}.
\end{aligned}$$

Here, to see the first inequality, we may rewrite $\mathfrak{d}_{jk}$ by (2.4) and use the inequality $\mathbb{E}(\sum_{i=1}^{\nu} f_i)^2 \leq \nu \sum_{i=1}^{\nu} \mathbb{E} f_i^2$ with $\mathbb{E}$ replaced by $\bar{\mathbb{E}}_{jk}$ and $\nu = \binom{n-2}{m-2}$. Similarly, we have $|\mathfrak{d}_i RJ_m(P)| \leq c_2 \frac{m}{n}$ for all i. The proof now follows by applying Theorem 6.1. □

As for the first order Hoeffding term $J_{m,1}(P)$, we have $J_{m,1}(P) = \sum_{i=1}^n g_1(X_i)$ with

$$g_1(X_i) = \frac{m}{n}\Big(\mathbb{E}_P\big(h\big(T_m(X_i, X_{j_1}, \ldots, X_{j_{m-1}})\big)|X_i\big) - \mathbb{E}_P h\big(T_m(X_i, X_{j_1}, \ldots, X_{j_{m-1}})\big)\Big),$$

where $j_1 < \cdots < j_{m-1}$ is any $(m-1)$-tuple from $\{1, \ldots, n\} \setminus \{i\}$. Conditions (8.4) imply that $h(T_m(X_1, \ldots, X_m); P)$ is "normalized", i.e. we have $B_1 = B_2 = \mathcal{O}(1)$ in Theorem 6.1 for $f = h - h_0 - h_1$. This may be achieved by requiring h to be sufficiently smooth.

In fact, in many applications, we can only assume $|\mathfrak{d}_{ij} h| \leq c_1$. In this case, we still get $J_m(P) = \theta_m(P) + J_{m,1}(P) + \mathcal{O}_P(m^2/n)$ in Proposition 8.3. A typical situation is $h = 1_A$ for some measurable set $A \subset \mathbb{R}$, i.e. we estimate the probability of $\{h(T_n) \in A\}$. Here, we clearly have $|\mathfrak{d}_{ij} h| \leq c_1$ and $|\mathfrak{d}_i(h - h_0 - h_1)| \leq c_2$. Consequently, while we cannot achieve the error of Proposition 8.3 in this situation, we still get an improved consistency result especially for small m.

8.2. Differentiable functions

We may apply Theorem 1.4 in the context of random matrix theory. Here we consider two cases.

Case 1 (Wigner matrices). Let $\{\xi_{jk}, 1 \le j \le k \le N\}$ be a family of independent real-valued random variables whose distributions all satisfy a logarithmic Sobolev inequality (1.8) with common constant σ^2. Putting $\xi_{jk} = \xi_{kj}$ for $1 \le k < j \le N$, we consider a symmetric $N \times N$ random matrix $\Xi = (\xi_{jk}/\sqrt{N})_{1 \le j,k \le N}$. Denote by $\mu^{(N)} = \mu$ the joint distribution of its ordered eigenvalues $\lambda_1 \le \cdots \le \lambda_N$ on $\mathbb{R}^N$ (in fact, $\lambda_1 < \cdots < \lambda_N$ a.s.). By a simple argument using the Hoffman–Wielandt theorem, μ satisfies a logarithmic Sobolev inequality with constant $\sigma_N^2 = 2\sigma^2/N$ (see for instance [8]). Note that similar observations also hold for Hermitian random matrices.

Case 2 (β-ensembles). For $\beta > 0$ fixed, let $\mu_{\beta,V}^{(N)} = \mu^{(N)} = \mu$ be the probability distribution on $\mathbb{R}^N$ with density given by

$$\mu(d\lambda) = \frac{1}{Z_N} e^{-\beta N \mathcal{H}(\lambda)}\, d\lambda, \qquad \mathcal{H}(\lambda) = \frac{1}{2}\sum_{k=1}^{N} V(\lambda_k) - \frac{1}{N} \sum_{1 \le k < l \le N} \log(\lambda_l - \lambda_k) \tag{8.5}$$

for $\lambda = (\lambda_1, \ldots, \lambda_N)$ such that $\lambda_1 < \cdots < \lambda_N$. Here, $V : \mathbb{R} \to \mathbb{R}$ is a strictly convex $\mathcal{C}^2$-smooth function and Z_N is a normalization constant. It is well-known that for $\beta = 1, 2, 4$, these probability measures correspond to the distributions of the classical invariant random matrix ensembles (orthogonal, unitary and symplectic, respectively). For other β, one can interpret (8.5) as particle systems on the real line with Coulomb interactions. Using the convexity of V, we may easily verify that

$$\mathcal{H}''(\lambda) \ge a\,\mathrm{Id} \tag{8.6}$$

uniformly in λ, where $\mathcal{H}''(\lambda)$ denotes the Hessian of $\mathcal{H}$, Id denotes the $N \times N$ identity matrix and $a > 0$ is some constant. As a consequence, by the classical Bakry-Emery criterion, μ satisfies a logarithmic Sobolev inequality (1.8) with constant $\sigma_N^2 = 1/(aN)$. For a detailed discussion see S.G. Bobkov and M. Ledoux [10].

Now consider the probability space $(\mathbb{R}^N, \mathbb{B}^N, \mu)$, where μ is either the joint eigenvalue distribution of Ξ or the distribution defined in (8.5). If $f : \mathbb{R} \to \mathbb{R}$ is a $\mathcal{C}^1$-smooth function, it is well-known that asymptotic normality

$$S_N = \sum_{j=1}^{N} \big(f(\lambda_j) - \mu\big[f(\lambda_j)\big]\big) \Rightarrow \mathcal{N}\big(0, \sigma_f^2\big) \tag{8.7}$$

holds for the self-normalized linear eigenvalue statistics S_N. Here, "$\Rightarrow$" denotes weak convergence, $\mu[\cdot]$ means integration with respect to μ and $\mathcal{N}(0, \sigma_f^2)$ denotes a normal distribution with mean zero and variance σ_f^2 depending on f.

This result goes back to K. Johansson [20] for the case of β-ensembles and, for general Wigner matrices, A.M. Khorunzhy, B.A. Khoruzhenko and L.A. Pastur [21] as well as Ya. Sinai and

A. Soshnikov [30]. Such results have been extensively studied since then. Concentration of measure results have been obtained by A. Guionnet and O. Zeitouni [18], proving concentration inequalities centered at the mean using techniques by Talagrand and Ledoux discussed in the introduction. In particular, they proved that S_N has fluctuations of order $\mathcal{O}_P(1)$ if f' is absolutely bounded. Here we can complement these results by a second order concentration bound which only requires f'' to be absolutely bounded.

Proposition 8.4. *Let μ be the joint distribution of the ordered eigenvalues of Ξ or the β-ensemble distribution defined in (8.5). Let $f\colon \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^2$-smooth function with $f'(\lambda_j) \in L^1(\mu)$ and second derivatives bounded by some constant $\gamma > 0$, and let $\tilde{S}_N := S_N - \sum_{j=1}^N (\lambda_j - \mu[\lambda_j])\mu[f'(\lambda_j)]$ with S_N as in (8.7). Then, we have*

$$\int \exp\bigl(cN^{1/2}|\tilde{S}_N|\bigr)\, d\mu \leq 2,$$

where $c = c(\gamma) > 0$ is some constant. If μ is the eigenvalue distribution of Ξ, c moreover depends on the Sobolev constant σ^2, and if μ is the β-ensemble distribution (8.5), c also depends on the quantity a from (8.6).

Proposition 8.4 follows from Theorem 1.4 and the fact that the Sobolev constant σ_N^2 is of order $1/N$. In view of the self-normalized property of S_N, the fluctuation result for $\tilde{S}_N$ is of the next order, although the scaling is of order $\sqrt{N}$ only.

Results of this type are useful in situations where f' is not bounded (i.e. [18] cannot be applied), in particular if f grows at most quadratically. In this case, concentration results for S_N may be obtained by considering $\tilde{S}_N$ and the "linear" part separately. Controlling the linear part is usually an easy task, while $\tilde{S}_N$ can be handled by Proposition 8.4. The idea of splitting eigenvalue statistics into a "linear" term and a remainder also appears in the analysis of interacting particle systems. That is, in (8.5), another quadratic "interaction energy" term of the form $\frac{1}{N}\sum_{i<j} h(\lambda_j - \lambda_i)$ is added to $\mathcal{H}(\lambda)$, where h is a "kernel" function with suitable properties. These particle system have been studied by F. Götze and M. Venker [17], including a concentration of measure result similar to our bounds for the recentered interaction energy h, removing both the expected value and a linear term (cf. e.g., [28] and Remark 4.8 there).

Indeed, second order results may also be used for establishing concentration bounds for quadratic eigenvalue statistics. The idea of studying higher order statistics of eigenvalues can already be found in A. Lytova and L. Pastur [25], where the authors proved a law of large numbers and a central limit theorem for U-statistics of eigenvalues. Following [17], an interesting question is whether the self-normalization phenomenon extends to what might be informally called a "double self-normalization" (at the level of the fluctuations, in our framework). That is, we shall examine whether quadratic statistics which are "recentered" in a suitable way may have fluctuations of a better order than $\mathcal{O}_P(N)$, i.e. possibly even $\mathcal{O}_P(1)$.

To this aim, we consider a sufficiently smooth "kernel" function $g\colon \mathbb{R}^2 \to \mathbb{R}$ and set $T_N := \sum_{j\neq k} g(\lambda_j, \lambda_k)$. Rescaling $T_N - \mu[T_N]$, we arrive at asymptotic normality. To obtain second order concentration of measure results, we shall also center around a linear correction term according to (1.9). That is, we define $Q_N = Q_N(\lambda)$ by

$$
\begin{aligned}
Q_N &:= \sum_{j\neq k} g(\lambda_j,\lambda_k) - \sum_{j\neq k}\mu\big[g(\lambda_j,\lambda_k)\big] \\
&\quad - \sum_{i=1}^{N}\Bigg(\sum_{k:k\neq i}\big(\mu\big[g_x(\lambda_i,\lambda_k)\big]+\mu\big[g_y(\lambda_k,\lambda_i)\big]\big)\Bigg)\big(\lambda_i-\mu[\lambda_i]\big). \qquad (8.8)
\end{aligned}
$$

Here, g_x, g_y etc. denote partial derivatives. For instance, if $g(x,y)=xy$, Q_N has the form

$$
Q_N(\lambda) = \sum_{j\neq k}\Big(\big(\lambda_j-\mu[\lambda_j]\big)\big(\lambda_k-\mu[\lambda_k]\big) - \mu\big[\big(\lambda_j-\mu[\lambda_j]\big)\big(\lambda_k-\mu[\lambda_k]\big)\big]\Big).
$$

In particular, this demonstrates that it is natural to remove a "linear" term in this context (also recall the discussion at the beginning of Section 1).

Proposition 8.5. *Let μ be the joint distribution of the ordered eigenvalues of Ξ or the β-ensemble distribution defined in (8.5). Let $g\colon \mathbb{R}^2\to\mathbb{R}$ be a $\mathcal{C}^2$-smooth function with first order derivatives in $L^1(\mu)$ and second order derivatives bounded by some constant $\gamma>0$. Consider Q_N as defined in (8.8). Then, for some constant $c=c(\gamma)>0$,*

$$
\int \exp\Big(\frac{c}{N^{1/2}}|Q_N|\Big)\,d\mu \le 2. \qquad (8.9)
$$

In the special case of $g(x,y):=xy$, we have

$$
\int \exp\big(c|Q_N|\big)\,d\mu \le 2. \qquad (8.10)
$$

If μ is the eigenvalue distribution of Ξ, c moreover depends on the Sobolev constant σ^2, and if μ is the β-ensemble distribution (8.5), c also depends on the quantity a from (8.6).

Proof. To check the conditions of Theorem 1.4, note that the Hessian of Q_N has entries

$$
\begin{aligned}
\big(Q_N''\big)_{ij} &= g_{xy}(\lambda_i,\lambda_j)+g_{yx}(\lambda_j,\lambda_i), \quad i\neq j, \\
\big(Q_N''\big)_{ii} &= \sum_{k:k\neq i}\big(g_{xx}(\lambda_i,\lambda_k)+g_{yy}(\lambda_k,\lambda_i)\big).
\end{aligned}
$$

Using the boundedness of the second derivatives of g it follows easily that $\|Q_N''\|_{\mathrm{Op}}\le cN$. Here, $c=c(\gamma)$ denotes a numerical constant which will vary from line to line throughout the proof.

On the other hand, we have

$$\int \|Q_N''\|_{\mathrm{HS}}^2 \, d\mu = \sum_{i \neq j} \int \big(g_{xy}(\lambda_i, \lambda_j) + g_{yx}(\lambda_j, \lambda_i)\big)^2 \, d\mu \\ + \sum_{i=1}^N \int \Big(\sum_{k : k \neq i} \big(g_{xx}(\lambda_i, \lambda_k) + g_{yy}(\lambda_k, \lambda_i)\big)\Big)^2 \, d\mu.$$

Here, the sum corresponding to the off-diagonal terms can clearly be bounded by cN^2, while the sum corresponding to the diagonal terms can only be bounded by cN^3 in general. Therefore, $\int \|Q_N''\|_{\mathrm{HS}}^2 \, d\mu \le cN^3$.

Finally, if $g(x, y) := xy$, we have $g_{xx} \equiv g_{yy} \equiv 0$, and consequently $\int \|Q_N''\|_{\mathrm{HS}}^2 \, d\mu \le cN^2$. Applying Theorem 1.4 finishes the proof. □

In case of $g(x, y) := xy$, by (8.10), Q_N has fluctuations of order $\mathcal{O}_P(1)$, which can be regarded as an extension of the self-normalizing property to a second order situation at least on the level of the fluctuations of Q_N.

Unfortunately, this property does not seem to hold in full strength for general kernels g. To explain this, note that Q_N may be decomposed into a "pure" quadratic part D_N and a remainder term $Q_N - D_N$, where $D_N = D_N(\lambda)$ is given by

$$D_N = \frac{1}{2} \sum_{i=1}^N \mu[\partial_{ii} Q_N]\big(\lambda_i^2 - 2\mu[\lambda_i]\lambda_i + 2\mu[\lambda_i]^2 - \mu\big[\lambda_i^2\big]\big). \tag{8.11}$$

Arguing similarly as in the proof of Proposition 8.5, for an arbitrary kernel g, $Q_N - D_N$ has fluctuations of order $\mathcal{O}_P(1)$, while the fluctuations of D_N are of a larger order $\mathcal{O}_P(N^{1/2})$. Therefore, we obtain a factor of $1/N^{1/2}$ in (8.9). If $g(x, y) = xy$, we have $D_N \equiv 0$, and hence we arrive at a different result (8.10). On the other hand, considering the case of $g(x, y) = f(x)$ for some function f with bounded second order derivatives, we arrive at Proposition 8.4 again. In particular, this shows that in general, the additional factor $1/N^{1/2}$ in (8.9) cannot be removed.

Acknowledgements

We wish to thank Sergey Bobkov for important suggestions concerning modified log-Sobolev inequalities which helped to improve this paper and Holger Kösters and Arthur Sinulis for many fruitful discussions.

This research was supported by the German Research Council (DFG) via the CRC 701 *Spectral Structures and Topological Methods in Mathematics* and the CRC 1283 *Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications*.

References

[1] Adamczak, R. and Wolff, P. (2015). Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probab. Theory Related Fields* **162** 531–586. MR3383337 https://doi.org/10.1007/s00440-014-0579-3

[2] Aida, S., Masuda, T. and Shigekawa, I. (1994). Logarithmic Sobolev inequalities and exponential integrability. *J. Funct. Anal.* **126** 83–101. MR1305064 https://doi.org/10.1006/jfan.1994.1142

[3] Arcones, M.A. and Giné, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21** 1494–1542. MR1235426

[4] Bickel, P.J., Götze, F. and van Zwet, W.R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statist. Sinica* **7** 1–31. MR1441142

[5] Bobkov, S.G., Chistyakov, G.P. and Götze, F. (2017). Second-order concentration on the sphere. *Commun. Contemp. Math.* **19** 1650058, 20. MR3670792 https://doi.org/10.1142/S0219199716500589

[6] Bobkov, S.G. and Götze, F. (1999). Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** 1–28. MR1682772 https://doi.org/10.1006/jfan.1998.3326

[7] Bobkov, S.G. and Götze, F. (2007). Concentration inequalities and limit theorems for randomized sums. *Probab. Theory Related Fields* **137** 49–81. MR2278452 https://doi.org/10.1007/s00440-006-0500-9

[8] Bobkov, S.G. and Götze, F. (2010). Concentration of empirical distribution functions with applications to non-i.i.d. models. *Bernoulli* **16** 1385–1414. MR2759184 https://doi.org/10.3150/10-BEJ254

[9] Bobkov, S.G., Götze, F. and Sambale, H. (2018). Higher order concentration of measure. *Commun. Contemp. Math.*, online first. https://doi.org/10.1142/S0219199718500438.

[10] Bobkov, S.G. and Ledoux, M. (2000). From Brunn–Minkowski to Brascamp–Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.* **10** 1028–1052. MR1800062 https://doi.org/10.1007/PL00001645

[11] Borodin, A.N. and Ibragimov, I.A. (1994). Limit theorems for functionals of random walks. *Tr. Mat. Inst. Steklova* **195** 286. MR1368394

[12] Boucheron, S., Lugosi, G. and Massart, P. (2003). Concentration inequalities using the entropy method. *Ann. Probab.* **31** 1583–1614. MR1989444 https://doi.org/10.1214/aop/1055425791

[13] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities*. Oxford: Oxford Univ. Press. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[14] Cordero-Erausquin, D., Fradelizi, M. and Maurey, B. (2004). The (B) conjecture for the Gaussian measure of dilates of symmetric convex sets and related problems. *J. Funct. Anal.* **214** 410–427. MR2083308 https://doi.org/10.1016/j.jfa.2003.12.001

[15] de la Peña, V.H. and Giné, E. (1999). *Decoupling: From Dependence to Independence. Probability and Its Applications (New York)*. New York: Springer. MR1666908 https://doi.org/10.1007/978-1-4612-0537-1

[16] Götze, F. (1993). IMS Bulletin.

[17] Götze, F. and Venker, M. (2014). Local universality of repulsive particle systems and random matrices. *Ann. Probab.* **42** 2207–2242. MR3265167 https://doi.org/10.1214/13-AOP844

[18] Guionnet, A. and Zeitouni, O. (2000). Concentration of the spectral measure for large matrices. *Electron. Commun. Probab.* **5** 119–136. MR1781846 https://doi.org/10.1214/ECP.v5-1026

[19] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. MR0026294 https://doi.org/10.1214/aoms/1177730196

[20] Johansson, K. (1998). On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.* **91** 151–204. MR1487983 https://doi.org/10.1215/S0012-7094-98-09108-6

[21] Khorunzhy, A.M., Khoruzhenko, B.A. and Pastur, L.A. (1996). Asymptotic properties of large random matrices with independent entries. *J. Math. Phys.* **37** 5033–5060. MR1411619 https://doi.org/10.1063/1.531589

[22] Ledoux, M. (1995/97). On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Stat.* **1** 63–87. MR1399224 https://doi.org/10.1051/ps:1997103

[23] Ledoux, M. (1999). Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXIII. Lecture Notes in Math.* **1709** 120–216. Berlin: Springer. MR1767995 https://doi.org/10.1007/BFb0096511

[24] Ledoux, M. (2001). *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. Providence, RI: Amer. Math. Soc. MR1849347

[25] Lytova, A. and Pastur, L. (2008). On asymptotic behavior of multilinear eigenvalue statistics of random matrices. *J. Stat. Phys.* **133** 871–882. MR2461187 https://doi.org/10.1007/s10955-008-9644-6

[26] Major, P. (2013). *On the Estimation of Multiple Random Integrals and U-Statistics. Lecture Notes in Math.* **2079**. Heidelberg: Springer. MR3087566 https://doi.org/10.1007/978-3-642-37617-7

[27] Politis, D.N. and Romano, J.P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050. MR1329181 https://doi.org/10.1214/aos/1176325770

[28] Robins, J.M., Li, L., Tchetgen, E.T. and van der Vaart, A. (2016). Asymptotic normality of quadratic estimators. *Stochastic Process. Appl.* **126** 3733–3759. MR3565475 https://doi.org/10.1016/j.spa.2016.04.005

[29] Sambale, H. (2015). Second order concentration for functions of independent random variables. Ph.D. thesis, Bielefeld Univ.

[30] Sinai, Y. and Soshnikov, A. (1998). Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Bol. Soc. Bras. Mat.* **29** 1–24. MR1620151 https://doi.org/10.1007/BF01245866

[31] Talagrand, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. Inst. Hautes Études Sci.* **81** 73–205. MR1361756

[32] Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563. MR1419006 https://doi.org/10.1007/s002220050108

Received May 2018 and revised November 2018