SECOND ORDER CONCENTRATION VIA LOGARITHMIC SOBOLEV INEQUALITIES

F. GÖTZE AND H. SAMBALE

Abstract. 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.

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 [T1], [T2]. Further research was done by S. Bobkov, M. Ledoux and others [L1], [B-G1], [B-G2]. 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 [L2], [L3].

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 $(\Omega, \mathcal{A}, \mu) := \otimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ is typically equipped with the Hamming distance $d(x, y) := \operatorname{card}\{k = 1, \ldots, n \colon x_k \neq y_k\}$. A related approach, which is essentially due to M. Ledoux [L1], makes use of certain "difference operators". That is, for any function $f \colon \Omega \to \mathbb{R}$ in $L^2(\mu)$, set

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

and $\mathfrak{d} f := (\mathfrak{d}_1 f, \ldots, \mathfrak{d}_n f)$. See Section 2 for a detailed description of the framework of the difference operators we use in this article. A slight modification of [B-G2, Proposition 2.1] then yields

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

$$\mu\left(|f - \int f d\mu| \geq t\right) \leq 2e^{-t^2/4}.$$

Note that the boundedness of f is in fact a consequence of the condition $|\mathfrak{d} f| \leq 1$ (see Section 2). If we apply Proposition 1.1 to 1-Lipschitz functions with respect to the

Date: May 30, 2016.

1991 *Mathematics Subject Classification.* Primary 60E15, 60F10, 60B20, 62F40.

Key words and phrases. Concentration of measure phenomenon, logarithmic Sobolev inequalities, Hoeffding decomposition, functions on the discrete cube, Bootstrap approximation.

This research was supported by CRC 701.

Hamming distance, we recover the classical concentration inequalities by M. Talagrand (cf. [L1]). Similar results can be derived in the context of "penalties", which can be regarded as generalizations of the Hamming distance [L1]. In [B-G2], a generalized version of Proposition 1.1 is used for deriving concentration inequalities for randomized sums.

In this article, we show a second order analogue of Proposition 1.1. Here, the notion of *second* order concentration has two aspects. Firstly, it refers to the use of difference operators of second order. Secondly, 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 by W. Hoeffding in 1948 [H]. Given a product probability space $(\Omega, \mathcal{A}, \mu) := \otimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ and some function $f \in L^1(\mu)$, the Hoeffding decomposition is the unique decomposition

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

such that $\int h_{i_1 \ldots i_k}(x_{i_1}, \ldots, x_{i_k}) \mu_{i_j}(dx_{i_j}) = 0$ for all $k = 1, \ldots, n$, $1 \le i_1 < \ldots < i_k \le n$ and $j \in \{1, \ldots, k\}$. The sum f_d is called the Hoeffding term of degree d or simply d-th 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. In addition to the "L^2-difference" $\mathfrak{d}$ 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

$$\mathfrak{D}_i f(x) := f(x) - \int_{\Omega_i} f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n) \mu_i(dy_i) \tag{1.3}$$

and $\mathfrak{D}f := (\mathfrak{D}_1 f, \ldots, \mathfrak{D}_n f)$. Higher order differences are defined by iteration, e.g. $\mathfrak{D}_{ij} f := \mathfrak{D}_i(\mathfrak{D}_j f)$ for $1 \le i, j \le n$. In particular, we consider the following modified "Hessian" with respect to $\mathfrak{D}$:

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

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 $(\Omega, \mathcal{A}, \mu) := \otimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ 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*

$$f = \sum_{k=2}^n f_k.$$

Denote by $\mathfrak{d}$ and $\mathfrak{D}$ the difference operators from (1.1) *and* (1.3)*. Assume that the condition*

$$|\mathfrak{d}|\mathfrak{d}f|| \le 1$$

is satisfied and that

$$\int \|\mathfrak{D}^{(2)} f\|_{\mathrm{HS}}^2 d\mu \leq b^2$$

holds for some $b \geq 0$*, where* $\mathfrak{D}^{(2)} f$ *denotes the "de-diagonalized" Hessian of* f *from* (1.4)*, and* $\|\mathfrak{D}^{(2)} f\|_{\mathrm{HS}}$ *denotes its Hilbert-Schmidt norm.*

Then, we have

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

If all the measures μ_i in Theorem 1.2 are Bernoulli measures, we can somewhat sharpen this bound. More precisely:

Corollary 1.3. *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*

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

Using Chebychev's inequality, Theorem 1.2 and Corollary 1.3 for instance imply the estimate

$$\mu(|f| \geq t) \leq 2e^{-ct}$$

for all $t > 0$ and some constant $c = c(b^2)$. The value of the latter constant as given by the bounds in Theorem 1.2 and Corollary 1.3 is not optimal, but optimizing it seems hard. It is possible to obtain a slightly better but still non-optimal constant from the proof of Theorem 1.2 and Corollary 1.3.

In the proof of Theorem 1.2, we shall use yet another difference operator, namely

$$\mathfrak{d}_i^+ f(x) := \Big(\frac{1}{2}\int_{\Omega_i} (f(x) - f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n))_+^2 \mu_i(dy_i)\Big)^{1/2}. \tag{1.5}$$

Here, $f\colon \Omega \to \mathbb{R}$ is any function in $L^2(\mu)$. Moreover, if g is any real-valued function we set $g_+ := \max(g, 0)$ for its positive part. It is straightforward to reformulate Theorem 1.2 using $\mathfrak{d}^+$ instead of $\mathfrak{d}$:

Corollary 1.4. *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* (1.5)*. Then, we have*

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

A generalization of Theorem 1.2 includes functions whose Hoeffding decomposition has a non-vanishing Hoeffding term of first order provided this term 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, for instance

$$\int e^{c|f|} d\mu \leq \int e^{c(|f_1|+|Rf|)} d\mu \leq 2$$

for some constant $c > 0$, we add conditions ensuring $f_1 = \mathcal{O}_P(1)$. The result is the following theorem:

Theorem 1.5. *Let $(\Omega_i, \mathcal{A}_i, \mu_i)$ be probability spaces, and denote by $(\Omega, \mathcal{A}, \mu) := \otimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ their product. Moreover, let $f \colon \Omega \to \mathbb{R}$ be a bounded measurable function such that 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 $\mathbb{E} f = 0$.) Denote by $\mathfrak{d}$ and $\mathfrak{D}$ the difference operators from (1.1) *and* (1.3)*. Suppose that the condition*

$$|\mathfrak{d}|\mathfrak{d}Rf|| \leq 1$$

is satisfied and that we have

$$\int \|\mathfrak{D}^{(2)} f\|_{\mathrm{HS}}^2 d\mu \leq b^2$$

for some $b \geq 0$. Here, $\|\mathfrak{D}^{(2)} f\|_{\mathrm{HS}}$ is the "de-diagonalized" Hessian of f from (1.4)*, and $\|\mathfrak{D}^{(2)} f\|_{\mathrm{HS}}$ denotes its Hilbert Schmidt norm. Furthermore, assume that one of the conditions*

(i) $|\mathfrak{d} f_1|^2 \leq b_0^2$ for some $b_0 \geq 0$ or
(ii) $|\mathfrak{d}|\mathfrak{d} f_1|| \leq 1$ and $\int |\mathfrak{d} f_1|^2 d\mu \leq \alpha^2$ for some $\alpha^2 \geq 0$

is satisfied. Depending on these conditions we have

$$\int \exp\left(\frac{1}{12 + 4b^2 + 7b_0}|f|\right) d\mu \leq 2$$

in case of condition (i) *and*

$$\int \exp\left(\frac{1}{4(3 + b^2 + \alpha^2)}|f|\right) d\mu \leq 2$$

in case of condition (ii).

Similar as in Corollary 1.3, it is possible to improve the constants c_1 and c_2 if all the underlying measures are Bernoulli distributions. We skip details at this point.

1.0.1. *Discussion of Related Inequalities.* Hoeffding decompositions have been studied in particular in the context of U-statistics, that is, statistics of the form

$$U_n(h) = \frac{(n-m)!}{n!} \sum_{i_1 \neq \ldots \neq i_m} h(X_{i_1}, \ldots, X_{i_m}) \tag{1.7}$$

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 $n \geq m$. 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 [D-G]. In particular, there are many inequalities describing their tail behavior starting with Hoeffding's inequalities. That is, for U-statistics like $U_n(h)$ in (1.7), we have

$$P(U_n(h) > t) \le \exp\Big(-\frac{[n/m]t^2}{2M^2}\Big)$$

if the function $h\colon \mathbb{R}^m \to \mathbb{R}$ 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. A. Arcones and E. Giné [A-G] as well as P. Major [M]. These inequalities typically depend on the order m, the second moment σ^2 and some bound M of the kernel h only.

1.0.2. *Applications.* Consider n independent random variables $X_1, \ldots, X_n$ and a statistic T_n of the form

$$\begin{aligned} T_n(X_1, \ldots, X_n) = h_{0,n} &+ \sum_i h_{1,n}(X_i)n^{-1} + \sum_{i<j} h_{2,n}(X_i, X_j)n^{-2} \\ &+ \sum_{i<j<k} h_{3,n}(X_i, X_j, X_k)n^{-3} + \ldots \end{aligned} \tag{1.8}$$

for some "kernel" functions $h_{d,n}$, $d = 0, 1, \ldots, n$, which are degenerate with respect to the X_i. Usually, we then have concentration inequalities of the form

$$P(\sqrt{n}(T_n - h_{0,n}) \ge t) \le e^{-ct},$$

where c is some absolute constant. Using second order concentration, it is possible to sharpen these bounds. Note that here we use a reformulated version of Theorem 1.2 which makes use of a second order version of the difference operator $\mathfrak{d}$ with Hessian $\mathfrak{d}^{(2)}$. See Section 6 for details and notations used below.

Example 1.6. *Let $X_1, \ldots, X_n$ be some independent random variables, and let T_n be a statistic of the form* (1.8)*. Assume we have*

$$\|n\mathfrak{d}^{(2)}T_n\|_{\mathrm{HS}} \le M \qquad \text{and} \qquad |n\mathfrak{d}_i(T_n - \sum_i h_{1,n}(X_i)n^{-1})| \le M \ \ \forall i \tag{1.9}$$

for some universal constant M and with $\mathfrak{d}^{(2)}T_n$ as in (6.3)*. Then, we have*

$$P\Big(n|T_n - h_{0,n} - \sum_i h_{1,n}(X_i)n^{-1}| \ge t\Big) \le 2e^{-ct/M},$$

where c is some numerical constant.

This follows immediately from Theorem 6.1. A simple case where conditions (1.9) are satisfied is the following: Assume that $\|h_{d,n}\|_\infty \equiv \sup_x |h_{d,n}(x)| \le L$ if $d \le m$ and $h_{d,n} \equiv 0$ for all $d \ge m$, where $m \in \mathbb{N}$ is independent of n and where L is some absolute constant.

In particular, we may consider functions on the discrete cube. That is, in Example 1.6, let $X_1, \ldots, X_n$ be i.i.d. random variables with distributions $\mu_i = \frac{1}{2}\delta_{+1} + \frac{1}{2}\delta_{-1}$. In Section 6, we will derive second order concentration results for functions of such random variables. In particular, by Proposition 6.2, it follows that we may replace conditions (1.9) by the single condition

$$\|n\mathfrak{D}^{(2)}T_n\|_{\mathrm{HS}} \le M. \tag{1.10}$$

Here, $\mathfrak{D}^{(2)}T_n$ is the "Hessian" of T_n with respect to $\mathfrak{D}$ defined in (1.4). For instance, if $T_n(X_1,\ldots,X_n) = \alpha_0 + \sum_i n^{-1}\alpha_i X_i + \sum_{i<j} n^{-2}\alpha_{ij}X_iX_j$ for real numbers $\alpha_0, \alpha_i, \alpha_{ij}$, then (1.10) just means $(2\sum_{i<j}\alpha_{ij}^2)^{1/2} \le M$.

We may furthermore apply our results in the context of bootstrap methods. Suppose $X_1,\ldots,X_n,\ldots$ 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 $T_n \equiv T_n(X_1,\ldots,X_n;P) \equiv T_n(\hat{P}_n;P)$ 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 $\theta_n(P) := \mathbb{E}_P h(T_n(X_1,\ldots,X_n;P))$. Given X_1, $\ldots, X_n$, Efron's (nonparametric) bootstrap suggests to estimate $\theta_n(P)$ by $\theta_n(\hat{P}_n)$. That is, if we set

$$B_n(P) = \mathbb{E}^* h(T_n(\hat{P}_n^*, P)) = \frac{1}{n^n}\sum_{i_1,\ldots,i_n=1}^{n} h(T_n(X_{i_1},\ldots,X_{i_n};P)),$$

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 bias problems) it fails asymptotically. D. N. Politis and J. P. Romano [P-R] as well as P. J. Bickel, F. Götze and W. R. van Zwet [G], [B-G-Z] have addressed these problems by introducing m out of n bootstraps, i. e. 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

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

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

Results on the asymptotic consistency of J_m have been obtained by Politis and Romano [P-R] and Götze [G]. For instance, by the result [B-G-Z], if $\frac{m}{n} \to 0$, $m \to \infty$, we have

$$J_m(P) = \theta_m(P) + \mathcal{O}_P((\frac{m}{n})^{1/2}). \tag{1.11}$$

Knowing (or at least estimating) the first order Hoeffding term of J_m, we may sharpen (1.11) by the second order results in this paper:

Theorem 1.7. *Suppose $\frac{m}{n} \to 0$, $m \to \infty$. Let $h = \sum_{i=0}^m h_i$ be the Hoeffding decomposition of $h = h(T_m(X_1,\ldots,X_m))$, and assume that*

$$|\mathfrak{d}_{ij}h| \le \frac{c_1}{m}, \qquad |\mathfrak{d}_i(h - h_0 - h_1)| \le c_2 \tag{1.12}$$

for all $1 \le i < j \le m$ 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

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

Proof. We proceed as in the proof of [B-G-Z, Theorem 1]. That is, suppose T_m does not depend on P. Then, J_m is a U-statistic with kernel $h(T_m(x_1,\ldots,x_m))$ and $\mathbb{E}_P J_m = \theta_m(P)$. Set $RJ_m := J_m - \theta_m(P) - J_{m,1}$ (cf. (1.6)). Using the notations of Theorem 6.1 and conditions (1.12), we have

$$\|\mathfrak{d}^{(2)} RJ_m\|_{\mathrm{HS}} = \mathcal{O}_P(\frac{m}{n}), \qquad |\mathfrak{d}_i RJ_m| = \mathcal{O}_P(\frac{m}{n}) \ \forall i.$$

The proof now follows by applying Theorem 6.1. □

As for the first order Hoeffding term $J_{m,1}(P)$, we have $J_{m,1}(P) = \sum_{i=1}^n h_1(X_i)$ with

$$h_1(X_i) = \frac{m}{n}\big(\mathbb{E}(h(T_m(X_i,X_{j_1},\ldots,X_{j_{m-1}}))|X_i) - \mathbb{E}h(T_m(X_i,X_{j_1},\ldots,X_{j_{m-1}}))\big),$$

where $j_1 < \ldots < j_{m-1}$ is any $m-1$-tuple from $\{1,\ldots,n\}\setminus\{i\}$. Conditions (1.12) imply that $h(T_m(X_1,\ldots,X_m))$ is "normalized", i.e. we have $B_1 = B_2 = \mathcal{O}(1)$ in Theorem 6.1 for $f = h - h_0 - h_1$. Conditions (1.12) may be achieved by requiring h to be sufficiently smooth. Note that without (1.12), we still get $J_m(P) = \theta_m(P) + J_{m,1} + \mathcal{O}_P(\frac{m^2}{n})$ in Theorem 1.7.

1.1. **Differentiable Functions.** In differentiable settings, it seems natural to use the ordinary gradient ∇ instead of difference operators. 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. Let us 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 $f\colon G \to \mathbb{R}$

$$\mathrm{Var}_\mu(f) \le \sigma^2 \int_G |\nabla f|^2 d\mu, \tag{1.13}$$

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 $f\colon G \to \mathbb{R}$

$$\mathrm{Ent}_\mu(f^2) \le 2\sigma^2 \int_G |\nabla f|^2 d\mu, \tag{1.14}$$

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 1.8. *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 $i = 1,\ldots,n$, where $\partial_i f$ denotes the i-th partial derivative of f. Assume that*

$$\int_G f d\mu = 0 \qquad \text{and} \qquad \int_G \partial_i f d\mu = 0 \ \text{for all } i = 1,\ldots,n.$$

Moreover, assume that

$$\|f''(x)\|_{\mathrm{HS}} \le 1 \text{ for all } x \in G \qquad \text{and} \qquad \int_G \|f''(x)\|_{\mathrm{HS}}^2 d\mu \le b^2$$

for some $b \ge 0$, where f'' denotes the Hessian of f and $\|f''\|_{\mathrm{HS}}$ denotes its Hilbert-Schmidt norm.

Then, the following inequality holds:

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

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 $i = 1, \ldots, n$, we may modify f to remove a "linear" term by considering

$$\tilde{f}(x) = f(x) - \mu(f) - \sum_{i=1}^n \mu(\partial_i f)(x_i - \mu(x_i)),$$

where $\mu(h) = \int_G h d\mu$ for any function $h \in L^1(\mu)$. $\tilde{f}$ represents a centered function with centered derivative.

Similarly to Theorem 1.5, we may allow non-vanishing integrals $\mu(\partial_i f)$ in Theorem 1.8 if they are of sufficiently small size. In detail:

Theorem 1.9. *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 $i = 1, \ldots, n$, where $\partial_i f$ denotes the i-th partial derivative of f. Assume that*

$$\int_G f d\mu = 0 \qquad \text{and} \qquad \sum_{i=1}^n \Big(\int_G \partial_i f d\mu\Big)^2 \le b_0^2$$

for some $b_0 \ge 0$. Moreover, assume that

$$\|f''(x)\|_{\mathrm{HS}} \le 1 \text{ for all } x \in G \qquad \text{and} \qquad \int_G \|f''(x)\|_{\mathrm{HS}}^2 d\mu \le b^2$$

for some $b \ge 0$, where f'' denotes the Hessian of f and $\|f''\|_{\mathrm{HS}}$ denotes its Hilbert-Schmidt norm.

Then, we have

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

Here we assume $\sigma > 0$ for the root of the Sobolev constant σ^2.

1.1.1. *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 [B-C-G]. 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 1.8 may then be seen as an intermediate between Theorem 1.2 and the bounds in [B-C-G]. Indeed, if in Theorem 1.8 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 D. Cordero-Erausquin, M. Fradelizi and B. Maurey [CE-F-M].

We would moreover like to mention the results by R. Adamczak and P. Wolff [A-W]. 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 1.8, functions from which we remove a kind of "linear term".

1.1.2. *Applications.* We may apply Theorem 1.8 in the context of random matrix theory. Here we consider two situations. Firstly, let $\{\xi_{jk}, 1 \le j \le k \le N\}$ be a family of independent random variables on some probability space. Assume that the distributions of the ξ_{jk}'s all satisfy a (one-dimensional) logarithmic Sobolev inequality (1.14) with common constant σ^2. Put $\xi_{jk} = \xi_{kj}$ for $1 \le k < j \le N$ and consider a symmetric $N \times N$ random matrix $\Xi = (\xi_{jk})_{1 \le j,k \le N}$ and denote by $\mu^{(N)}$ the joint distribution of its eigenvalues on $\mathbb{R}^N$. By a simple argument using the Hoffman-Wielandt theorem, $\mu^{(N)}$ satisfies a logarithmic Sobolev inequality with constant $\sigma_N^2 = \frac{2\sigma^2}{N}$ (see for instance S. G. Bobkov and F. Götze [B-G3]). Note that similar observations also hold for Hermitean random matrices.

Furthermore, we consider β-ensembles. That is, for $\beta > 0$ fixed, let $\mu_{\beta,V}^{(N)} = \mu^{(N)}$ be the probability distribution on $\mathbb{R}^N$ with density given by

$$\mu^{(N)}(d\lambda) = \frac{1}{Z_N} e^{-\beta N \mathcal{H}(\lambda)} d\lambda, \qquad \mathcal{H}(\lambda) = \frac{1}{2} \sum_{k=1}^N V(\lambda_k) - \frac{1}{N} \sum_{1 \le k < l \le N} \log |\lambda_l - \lambda_k|. \quad (1.15)$$

Here, $V \colon \mathbb{R} \to \mathbb{R}$ 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). Using the convexity of V, we may easily verify that

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

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, $\mu^{(N)}$ satisfies a logarithmic Sobolev inequality (1.14) with constant $\sigma_N^2 = 1/(aN)$. For a detailed discussion see S. G. Bobkov and M. Ledoux [B-L].

Now consider the probability space $(\mathbb{R}^N, \mathbb{B}^N, \mu^{(N)})$, where $\mu^{(N)}$ is either the joint eigenvalue distribution of Ξ or the distribution defined in (1.15). If $f \colon \mathbb{R} \to \mathbb{R}$ is a $\mathcal{C}^1$-smooth function, it is well-known that asymptotic normality

$$S_N = \sum_{j=1}^N (f(\lambda_j) - \mathbb{E} f(\lambda_j)) \Rightarrow \mathcal{N}(0, \sigma_f^2), \qquad (1.16)$$

holds for the self-normalized linear eigenvalue statistics S_N. Here, "$\Rightarrow$" denotes weak convergence, $\mathbb{E}$ means taking the expectation with respect to $\mu^{(N)}$ 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 [J] for the case of β-ensembles and, for general Wigner matrices, A. M. Khorunzhy, B. A. Khoruzhenko and L. A. Pastur [K-K-P] as well as Ya. Sinai and A. Soshnikov [S-S]. Such results have been extensively studied since then. Concentration of measure results have been studied by A. Guionnet and O. Zeitouni [G-Z], 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 $\mathcal{O}_P(1)$. Here we can complement these results by a second order concentration bound:

Proposition 1.10. *Let $\mu^{(N)}$ be the joint eigenvalue distribution of Ξ or the distribution defined in* (1.15)*. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^2$-smooth function with $f'(\lambda_j) \in L^1(\mu^{(N)})$ and second derivatives bounded by some absolute constant, and let*

$$\tilde{S}_N := S_N - \sum_{j=1}^N (\lambda_j - \mathbb{E}(\lambda_j))\mathbb{E}f'(\lambda_j)$$

with S_N as in (1.16)*. Then, we have*

$$\mathbb{E}e^{cN^{1/2}\tilde{S}_N} \leq 2,$$

where $c > 0$ is some absolute constant.

Proposition 1.10 follows from Theorem 1.8 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.

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 then be given in Sections 4 and 5. Here, we will first derive exponential inequalities involving the difference operator $\mathfrak{d}$ by making use of modified logarithmic Sobolev inequalities. After that, we shall relate $\mathfrak{d}$ to the second order difference operators $\mathfrak{D}_{ij}$ in form of the Hessian (1.4). The proof of Theorem 1.2 and its reformulations then follow 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 symmetric Bernoulli variables.

Finally, 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 1.5. Together with a simple application of the Poincaré inequality, this will lead us to the proof of Theorems 1.8 and 1.9.

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

Acknowledgements.

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

2. Difference Operators

Let $(\Omega_1, \mathcal{A}_1), \ldots, (\Omega_n, \mathcal{A}_n)$ be measurable spaces, and denote by $(\Omega, \mathcal{A})$ their product space. Similarly to [B-G1], 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:

Conditions 2.1. *(i) For any bounded measurable function* $f\colon \Omega \to \mathbb{R}$, $\Gamma f = (\Gamma_1 f, \ldots, \Gamma_n f)\colon \Omega \to \mathbb{R}^n$ *is a measurable function with values in* $\mathbb{R}^n$. *We often call* Γ *a* gradient operator *or simply* gradient.

(ii) For all $i = 1, \ldots, n$, *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 particular, we do not suppose Γ to satisfy any sort of "Leibniz rule".

Clearly, the difference operators $\mathfrak{d}$, $\mathfrak{D}$ and $\mathfrak{d}^+$ from (1.1), (1.3) and (1.5) satisfy Conditions 2.1. Note that here we assume the spaces $(\Omega_i, \mathcal{A}_i)$ to be endorsed with probability measures μ_i, whose product measure we denote by μ. We now collect some elementary facts about these three difference operators. In the following assume that $X_1, \ldots, X_n$ is a sequence of independent random variables on some probability space $(\Omega', \mathcal{A}', P)$ with distributions $\mu_1, \ldots, \mu_n$ 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* $i = 1, \ldots, n$, *we have the simple representation*

$$\mathfrak{D}_i f(X) = \frac{1}{2}(f(X) - f(\sigma_i X)),$$

where $X = (X_1, \ldots, X_n)$ *and* $\sigma_i X := (X_1, \ldots, -X_i, \ldots, X_n)$. *Moreover, note that* $\mathfrak{d}_i f = |\mathfrak{D}_i f|$ *and* $\mathfrak{d}_i^+ f = (\mathfrak{D}_i f)_+$.

(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) $\mathfrak{D}_i = Id - \mathbb{E}_i$. *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* $\bar{X}_1, \ldots, \bar{X}_n$ *be a set of independent copies of the random variables* $X_1, \ldots, X_n$. *Set* $T_i f := f(X_1, \ldots, X_{i-1}, \bar{X}_i, X_{i+1}, \ldots, X_n)$ *for any function* $f(X_1, \ldots, X_n)$. *Then, we have*

$$\mathfrak{d}_i f(X) = \Big(\frac{1}{2}\bar{\mathbb{E}}_i (f(X) - T_i f(X))^2\Big)^{1/2}. \tag{2.1}$$

Here, $\bar{\mathbb{E}}_i$ *denotes the expectation with respect to* $\bar{X}_i$. *By independence, we can rewrite* (2.1) *as*

$$\begin{aligned}\mathfrak{d}_i f(X) &= \Big(\frac{1}{2}\big((f(X) - \mathbb{E}_i f(X))^2 + \mathbb{E}_i (f(X) - \mathbb{E}_i f(X))^2\big)\Big)^{1/2} \\ &= \Big(\frac{1}{2}\big((\mathfrak{D}_i f(X))^2 + \mathbb{E}_i (\mathfrak{D}_i f(X))^2\big)\Big)^{1/2},\end{aligned} \tag{2.2}$$

where $\mathbb{E}_i$ *denotes the expectation with respect to* X_i.

(4) *Similarly, we have*

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

for any $f(X) \in L^2(P)$ *as well as*

$$\mathfrak{D}_i f(X) = \bar{\mathbb{E}}_i(f(X) - T_i f(X))$$

for any $f(X) \in L^1(P)$.

By induction over n, f is bounded if and only if $|\mathfrak{D}f|$ is bounded. Using (2.2), the same holds for $|\mathfrak{d}f|$ instead of $|\mathfrak{D}f|$.

Note that the difference operator $\mathfrak{D}$ is closely related to the Hoeffding decomposition (1.2). In essence, proving (1.2) is based on the identity $\mathbb{E}_i + \mathfrak{D}_i = Id$ with $\mathfrak{D}_i$ as in (1.3). We finally get

$$h_{i_1\ldots i_k}(X_{i_1}, \ldots, X_{i_k}) = \Big(\prod_{j \notin \{i_1,\ldots i_k\}} \mathbb{E}_j \prod_{l \in \{i_1,\ldots i_k\}} \mathfrak{D}_l \Big) f(X_1, \ldots, X_n).$$

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

$$\mathfrak{L} := \sum_{i \neq j} \mathfrak{D}_{ij}. \tag{2.3}$$

Calling (2.3) a Laplacian is justified for several reasons. First of all, (2.3) 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 (2.3) is invariant under permutations, i. e.

$$\mathfrak{L}f(x) = \mathfrak{L}f(\pi(x))$$

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* $(\Omega, \mathcal{A}, \mu) := \otimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ *their product. Moreover,* 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* (2.3)*, and we write* $(d)_2 = d(d-1)$. *Thus, the d-th 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 $f_d(x_1, \ldots, x_n) = \sum_{i_1 < \ldots < i_d} h_{i_1\ldots i_d}(x_{i_1}, \ldots, x_{i_d})$ as in (1.2). Fix $i_1 < \ldots < i_d$. Then, we get

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

Therefore, we have

$$\mathfrak{D}_i f_d(x_1,\dots,x_n) = \sum_{\substack{i_1<\dots<i_d \\ i\in\{i_1,\dots,i_d\}}} h_{i_1\dots i_d}(x_{i_1},\dots,x_{i_d}) \tag{2.4}$$

and consequently

$$\mathfrak{D}_{ij} f_d(x_1,\dots,x_n) = \sum_{\substack{i_1<\dots<i_d \\ i,j\in\{i_1,\dots,i_d\}}} h_{i_1\dots i_d}(x_{i_1},\dots,x_{i_d}). \tag{2.5}$$

Hence it remains to check how often each term $h_{i_1\dots i_d}(x_{i_1},\dots,x_{i_d})$ 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 $i,j\in\{i_1,\dots,i_d\}$ replicates the summand $h_{i_1\dots i_d}(x_{i_1},\dots,x_{i_d})$ precisely once. As there are $d(d-1)=(d)_2$ such pairs, we arrive at the result. □

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

$$\mathfrak{L}_1 := \sum_i \mathfrak{D}_i, \qquad \mathfrak{L}_2 := \mathfrak{L}_1^2 \qquad \text{and more generally} \qquad \mathfrak{L}_k := \mathfrak{L}_1^k$$

for any $k\in\{1,2,\dots,n\}$. Another one is given by

$$\mathfrak{L}_k^* := \sum_{i_1\neq i_2\neq\dots\neq i_k} \mathfrak{D}_{i_1}\dots\mathfrak{D}_{i_k}$$

for any $k\in\{1,2,\dots,n\}$. It is possible to relate these two families to each other by representing the $\mathfrak{L}_k^*$ as polynomials in $\mathfrak{L}_1$, e.g. 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 (2.3), 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.

It would be interesting to study concentration of higher order for functions of independent random variables with the help of the operators $\mathfrak{L}_k^*$, but it seems that this will get more involved than the second order case and needs a different set of technical tools for proving concentration. We intend to return to this question in the future.

3. Modified Logarithmic Sobolev 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

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 measurable function $f \colon \Omega \to \mathbb{R}$ such that the following integrals are finite we have*

$$\operatorname{Ent}(e^f) \le \frac{\sigma^2}{2} \int |\Gamma f|^2 e^f d\mu. \tag{3.1}$$

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

This definition goes back to [B-G1], where it is called LSI_{σ^2}. The term "modified logarithmic Sobolev inequality" is due to Ledoux [L3, Chapter 5.3], where other modifications of logarithmic Sobolev inequalities are discussed as well. The difference between the usual form of the LSI and modified one in (3.1) is motivated by the fact that the difference operators from Example 2.2 do not 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 $\Gamma = \mathfrak{d}$ and $\Gamma = \mathfrak{d}^+$ from Example 2.2. 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 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 [B-G2]. For the reader's convenience we include a sketch of its proof here. Noting that we only need (1.1) in dimension one in the present situation, we apply Jensen's inequality to get

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

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:

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 [B-G2], 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:

Lemma 3.4. *For all $i = 1, \ldots, n$, 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 $\mu_1 \otimes \ldots \otimes \mu_n$ on $(\Omega_1 \times \ldots \times \Omega_n, \mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n)$ also satisfies the modified LSI* (3.1) *with Sobolev constant $\sigma^2 = \max_{i=1,\ldots,n} \sigma_i^2$ with respect to $\mathfrak{d}$.*

As in the case the usual logarithmic Sobolev inequality, this is a consequence of the subadditivity 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.

For technical reasons, 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 $2\sigma^2$. As a result, we can transport Propositions 3.2 and 3.3 and Lemma 3.4 to the $\mathfrak{d}^+$ difference operators. We summarize these results in the following proposition:

Proposition 3.5. *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 $\mu_1 \otimes \ldots \otimes \mu_n$ on $(\Omega_1 \times \ldots \times \Omega_n, \mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n)$ satisfies the modified LSI* (3.1) *with Sobolev constant $\sigma^2 = 4$ with respect to $\mathfrak{d}^+$ as in* (1.5)*. If all the Ω_i are two-point spaces, we can take $\sigma^2 = 2$.*

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 the gradient operator $\mathfrak{d}$. In Bobkov and Götze [B-G1], it was proved that for all bounded measurable functions $f \colon \Omega \to \mathbb{R}$ such that $\int f d\mu = 0$, we have

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

The proof of (4.1) is similar to the proof of inequality (4.4) which will be sketched in the proof of Lemma 4.1.

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

$$\int e^{tu^2} d\mu \leq \exp\left(c(t) \int u^2 d\mu \right)$$

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 [A-M-S] 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 $|\nabla f| \leq 1$, we immediately get

$$|\nabla f^2| = 2|f||\nabla f| \leq 2|f|.$$

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, namely $\mathfrak{d}^+$ as introduced in (1.5). Indeed, 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

$$\begin{aligned}(f(x)^2 - f(y)^2)_+^2 &= (|f(x)| + |f(y)|)^2(|f(x)| - |f(y)|)_+^2 \\ &\leq 4|f(x)|^2(|f(x)| - |f(y)|)_+^2.\end{aligned}$$

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

$$|\mathfrak{d}^+ f^2| \leq 2|f|. \tag{4.2}$$

The same holds for product measures, i. e. the multivariate case. Now (4.2) leads us back to the basic inequality needed to estimate large deviations in [B-G1]. We therefore arrive at the following lemma:

Lemma 4.1. *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 (1.5). Moreover, let $f\colon \Omega \to \mathbb{R}$ be a bounded measurable function such that $|\mathfrak{d}|f|| \leq 1$. Then we have*

$$\int e^{tf^2} d\mu \leq \exp\left(\frac{t}{1 - 2\tilde{\sigma}^2 t} \int f^2 d\mu\right) \tag{4.3}$$

for all $t \in [0, \frac{1}{2\tilde{\sigma}^2})$.

Proof. We adapt the arguments from [B-G1], p. 6 f. First, consider the inequality

$$\int e^f d\mu \leq \left(\int e^{\lambda f + (1-\lambda)\tilde{\sigma}^2|\mathfrak{d}^+ f|^2/2} d\mu\right)^{1/\lambda} \tag{4.4}$$

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 (4.4), we use the well-known “variational formula”

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

which can be shown by Young’s inequality in the form

$$uv \leq u \log u - u + e^v$$

for all $u \geq 0$ and $v \in \mathbb{R}$, for instance. See [L3, 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

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

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

$$\lambda \int f e^f d\mu + (1-\lambda)\mathrm{Ent}(e^f) - \beta \int e^f d\mu \leq \mathrm{Ent}(e^f).$$

This is equivalent to

$$\lambda \int e^f d\mu \log \int e^f d\mu - \beta \int e^f d\mu \leq 0,$$

from which we directly get (4.4).

We now apply (4.4) 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 (4.2) (note that $|\mathfrak{d}|f|| \le 1$ implies $|\mathfrak{d}^+|f|| \le 1$), this gives us

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

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

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

Now we set

$$u(s) := \int e^{sf^2/(2\tilde{\sigma}^2)} d\mu,$$

$s \in (0, 1]$. Then we get

$$0 \ge \frac{1}{1-s}\left[s(1-s)u'(s) - u(s) \log u(s) \right],$$

or equivalently

$$0 \ge \frac{1-s}{s}\frac{u'(s)}{u(s)} - \frac{1}{s^2} \log u(s).$$

Hence, the function

$$v(s) := \exp\left(\frac{1-s}{s} \log u(s) \right)$$

is non-increasing in s, and therefore we have $v(s) \le \lim_{s \downarrow 0} v(s) =: v(0^+)$ for all $s \in (0, 1]$.

Note that

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

Thus, we have

$$\exp\left(\frac{1-s}{s} \log u(s) \right) \le \exp\left(\frac{1}{2\tilde{\sigma}^2} \int f^2 d\mu \right)$$

for all $s \in (0, 1]$, or equivalently

$$\int e^{sf^2/(2\tilde{\sigma}^2)} d\mu \le \exp\left(\frac{1}{2\tilde{\sigma}^2} \frac{s}{1-s} \int f^2 d\mu \right).$$

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

Combining inequalities (4.1) and (4.3), we now get the following result.

Proposition 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 $\sigma^2 > 0$ with respect to the gradient operator $\mathfrak{d}$ and which moreover satisfies the modified LSI (3.1) with Sobolev constant*

$\tilde{\sigma}^2$ *with respect to the gradient operator* $\mathfrak{d}^+$. *Furthermore, let* $f\colon \Omega \to \mathbb{R}$ *be a bounded measurable function such that* $\int f d\mu = 0$ *and* $|\mathfrak{d}|\mathfrak{d}f|| \le 1$. *Then, we have*

$$\int \exp\left(\frac{1}{2\sigma\tilde{\sigma}} f\right) d\mu \le \exp\left(\frac{1}{2\tilde{\sigma}^2} \int |\mathfrak{d}f|^2 d\mu\right). \tag{4.5}$$

Proof. First, applying (4.1) to λf leads to

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

Moreover, (4.3) with $t = \lambda^2\sigma^2$ for any $\lambda \in [0, \frac{1}{\sqrt{2}\sigma\tilde{\sigma}})$ and with f replaced by $|\mathfrak{d}f|$ gives us

$$\int e^{\lambda^2\sigma^2|\mathfrak{d}f|^2} d\mu \le \exp\left(\frac{\lambda^2\sigma^2}{1 - 2\sigma^2\tilde{\sigma}^2\lambda^2} \int |\mathfrak{d}f|^2 d\mu\right).$$

Combining these two inequalities then yields

$$\int e^{\lambda f} d\mu \le \exp\left(\frac{\lambda^2\sigma^2}{1 - 2\sigma^2\tilde{\sigma}^2\lambda^2} \int |\mathfrak{d}f|^2 d\mu\right).$$

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 (4.5), we may now study relations of the form

$$\gamma \int |\mathfrak{d}f|^2 d\mu \le \int \|\mathfrak{D}^{(2)} f\|_{\mathrm{HS}}^2 d\mu$$

for some constant $\gamma > 0$, where $\mathfrak{D}^{(2)}$ is the "de-diagonalized" Hessian of f with respect to the "Hoeffding" type difference operator $\mathfrak{D}$ introduced in (1.4). 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* $(\Omega, \mathcal{A}, \mu) := \otimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ *their product. Consider* $f, g \in L^2(\mu)$. *Then, we have:*

(1)

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

where $\mathfrak{D}_i$ *is the difference operator from* (1.3).

(2)

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

where $\mathfrak{D}$ *is the gradient operator from* (1.3) *and the integral has to be understood componentwise.*

(3)

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

where $\mathfrak{L}$ *is the Laplacian as in* (2.3).

Hence, the difference operators $\mathfrak{D}_i$, the gradient operator $\mathfrak{D}$ and the Laplacian $\mathfrak{L}$ are in some sense selfadjoint operators on $L^2(\mu)$.

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 (1.3) 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 $(\Omega, \mathcal{A}, \mu) := \otimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ 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 $f = f_d$, i. e. 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 $\mathfrak{L} f_k = (k)_2 f_k$. Consequently, we have

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

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

$$f_k(X_1, \ldots, X_n) = \sum_{i_1 < \ldots < i_k} h_{i_1 \ldots i_k}(X_{i_1}, \ldots, X_{i_k}),$$

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 $T_{i_j} h_{i_1 \ldots i_k}(X_{i_1}, \ldots, X_{i_d}) = h_{i_1 \ldots i_k}(X_{i_1}, \ldots, \bar{X}_{i_j}, \ldots, X_{i_k})$ (cf. Example 2.2(3)). Then,

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

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 (2.4), we therefore get

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

Using orthogonality, it follows that

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

As in the proof of Theorem 2.3, it remains to check how often each term $\mathbb{E}h^2_{i_1\ldots i_k}(X_{i_1},\ldots,X_{i_k})$ 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 $\mathbb{E}h_{i_1\ldots i_k}(X_{i_1},\ldots,X_{i_k})$ exactly once. Consequently, it follows that $\mathbb{E}|\mathfrak{d}f_k|^2 = k\mathbb{E}f_k^2$, or

$$\int |\mathfrak{d}f_k|^2 d\mu = k\int f_k^2 d\mu. \qquad (**)$$

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

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

This finally completes the proof. □

We are now ready to prove our main theorems:

Proof of Theorem 1.2 and Corollary 1.3. First, combining Proposition 4.2, Proposition 5.2 with $d = 2$ and the assumptions from Theorem 1.2 leads to

$$\int \exp\left(\frac{1}{2\sigma\tilde{\sigma}}f\right)d\mu \le \exp\left(\frac{1}{2\tilde{\sigma}^2}\int \|\mathfrak{D}^{(2)}f\|^2_{\mathrm{HS}}d\mu\right) \le \exp\left(\frac{b^2}{2\tilde{\sigma}^2}\right) \qquad (5.1)$$

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

$$\begin{aligned}\int \exp\left(\frac{1}{2\sigma\tilde{\sigma}}|f|\right)d\mu &\le \int\left(\exp\left(\frac{1}{2\sigma\tilde{\sigma}}f\right) + \exp\left(\frac{1}{2\sigma\tilde{\sigma}}(-f)\right)\right)d\mu\\ &\le 2\exp\left(\frac{b^2}{2\tilde{\sigma}^2}\right).\end{aligned}$$

Thus, by applying Hölder's inequality we obtain

$$\int \exp\left(\frac{1}{2\sigma\tilde{\sigma}\kappa}|f|\right) d\mu \leq \left(\int \exp\left(\frac{1}{2\sigma\tilde{\sigma}}|f|\right) d\mu\right)^{1/\kappa} \leq \left(2\exp\left(\frac{b^2}{2\tilde{\sigma}^2}\right)\right)^{1/\kappa}$$

for all $\kappa \geq 1$. The last term is bounded by 2 if

$$\kappa \geq \left(\log 2 + \frac{1}{2\tilde{\sigma}^2}b^2\right) / \log 2,$$

or equivalently

$$\frac{1}{2\sigma\tilde{\sigma}\kappa} \leq \frac{\log 2}{2\sigma\tilde{\sigma}\log 2 + \frac{\sigma}{\tilde{\sigma}}b^2}.$$

By Proposition 3.2, Proposition 3.3, Lemma 3.4 and Proposition 3.5, 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

$$\int \exp\left(\frac{\log 2}{\sqrt{32}\log 2 + \frac{1}{\sqrt{2}}b^2}|f|\right) d\mu \leq 2 \tag{5.2}$$

if $\sigma^2 = 2$ and $\tilde{\sigma}^2 = 4$ and

$$\int \exp\left(\frac{\log 2}{\sqrt{8}\log 2 + \frac{1}{\sqrt{2}}b^2}|f|\right) d\mu \leq 2 \tag{5.3}$$

if $\sigma^2 = 1$ and $\tilde{\sigma}^2 = 2$. Noting that

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

for all $x \geq 0$ completes the proof. □

Simple modifications of the above arguments now lead to the proof of Corollary 1.4:

Proof of Corollary 1.4. 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

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

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

$$\frac{1}{2\tilde{\sigma}^2\kappa} \leq \frac{\log 2}{2\tilde{\sigma}^2\log 2 + b^2}.$$

Plugging in $\tilde{\sigma}^2 = 4$ and noting that

$$\frac{\log 2}{8\log 2 + x} \geq \frac{1}{8 + 2x}$$

for all $x \geq 0$ completes the proof. □

Finally, we prove Theorem 1.5.

Proof of Theorem 1.5. The basic argument is as follows: if we have two functions φ_1 and φ_2 on $\mathbb{R}^n$ both satisfying

$$\int e^{c_i|\varphi_i|} d\mu \leq 2 \qquad (*)$$

for some constants $c_i > 0$, $i = 1, 2$, it follows that

$$\begin{aligned} \int e^{\min(c_1,c_2)|\varphi_1+\varphi_2|/2} d\mu &\leq \int e^{c_1|\varphi_1|/2} e^{c_2|\varphi_2|/2} d\mu \\ &\leq \Big(\int e^{c_1|\varphi_1|} d\mu\Big)^{1/2} \Big(\int e^{c_2|\varphi_2|} d\mu\Big)^{1/2} \leq 2 \end{aligned} \tag{5.4}$$

due to the Cauchy-Schwarz inequality. In our situation, we set $\varphi_1 = f_1$ and $\varphi_2 = Rf$. Hence, we only have to check $(*)$.

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 (2.5). To prove the bound for f_1, assuming condition (ii), we apply Proposition 4.2 and then proceed as in the proof of Theorem 1.2. Using the notation from $(*)$, this leads to

$$c_1 = \frac{1}{6 + 2\alpha^2} \qquad \text{and} \qquad c_2 = \frac{1}{6 + 2b^2},$$

and hence we can estimate $\min(c_1, c_2)/2$ as stated in Theorem 1.5

It remains to check $(*)$, bounding f_1 using condition (i). Here, inequality (4.1) yields

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

for any $\lambda > 0$, thus

$$\int e^{\lambda |f_1|} d\mu \leq 2 e^{\sigma^2 \lambda^2 b_0^2}.$$

As in the proof of Theorem 1.2, it follows that

$$\int e^{\lambda |f_1|/\kappa} d\mu \leq \Big(2 e^{\sigma^2 \lambda^2 b_0^2}\Big)^{1/\kappa}.$$

for all $\kappa \geq 1$. The right hand side is bounded by 2 if

$$\frac{\lambda}{\kappa} \leq \frac{\lambda \log 2}{\log 2 + \lambda^2 \sigma^2 b_0^2}.$$

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

$$\frac{1}{2} c_1 = \frac{(\log 2)^{1/2}}{4\sqrt{2} b_0} \geq \frac{1}{7 b_0}.$$

With c_2 as in the first part of the proof we arrive at the bound given in the theorem. □

Compared to Theorem 1.2, Theorem 1.5 needs conditions which are more involved and hence are not always easy to check.

6. Evaluating Second Order Difference Operators

In Theorem 1.2, checking the condition $\int \|\mathfrak{D}^{(2)} f\|^2_{\mathrm{HS}} d\mu \le b^2$ is straightforward, once we know the Hoeffding decomposition of f. 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.

To this end, in addition to the first order L^2-differences $\mathfrak{d}$, let us define second order L^2-differences. Recall the operators T_i from Remark 2.2(3) given by

$$T_i f(X_1, \ldots, X_n) = f(X_1, \ldots, \bar{X}_i, \ldots, X_n),$$

where $\bar{X}_1, \ldots, \bar{X}_n$ is a set of independent copies of the random variables $X_1, \ldots, X_n$. Setting $T_{ij} = T_i \circ T_j$, let

$$\mathfrak{d}_{ij} f(X) = \Big(\frac{1}{4} \bar{\mathbb{E}}_{ij} (f(X) - T_i f(X) - T_j f(X) + T_{ij} f(X))^2 \Big)^{1/2} \tag{6.1}$$

for any $i \neq j$. Here, $\bar{\mathbb{E}}_{ij}$ means taking the expectation with respect to $\bar{X}_i$ and $\bar{X}_j$. Note that similarly to (2.2), (6.1) can be rewritten as

$$\mathfrak{d}_{ij} f = \Big(\frac{1}{4} \big((\mathfrak{D}_{ij} f)^2 + \mathbb{E}_i (\mathfrak{D}_{ij} f)^2 + \mathbb{E}_j (\mathfrak{D}_{ij} f)^2 + \mathbb{E}_{ij} (\mathfrak{D}_{ij} f)^2 \big) \Big)^{1/2}, \tag{6.2}$$

where $\mathbb{E}_i$ and $\mathbb{E}_j$ denote taking the expectation with respect to X_i and X_j, respectively, and $\mathbb{E}_{ij}$ means taking the expectation with respect to X_i and X_j. Similarly to (1.4), we now define another sort of "Hessian" now based L^2-differences by

$$(\mathfrak{d}^{(2)} f(X))_{ij} := \begin{cases} \mathfrak{d}_{ij} f(X), & i \neq j, \\ 0, & i = j. \end{cases} \tag{6.3}$$

Theorem 6.1. *Let $(\Omega_i, \mathcal{A}_i, \mu_i)$ be probability spaces, and denote by $(\Omega, \mathcal{A}, \mu) := \otimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$ 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*

$$f = \sum_{k=2}^n f_k.$$

Denote by $\mathfrak{d}$ the difference operator introduced in (1.1). *Assume that the condition*

$$\|\mathfrak{d}^{(2)} f\|_{\mathrm{HS}} \le B_1 \tag{6.4}$$

is satisfied for some $B_1 \ge 0$, where $\mathfrak{d}^{(2)} f$ denotes the "de-diagonalized" Hessian of f from (6.3) *and $\|\mathfrak{d}^{(2)} f\|_{\mathrm{HS}}$ denotes its Hilbert-Schmidt norm, and assume that*

$$|\mathfrak{d}_i f| \le B_2 \qquad \text{for all } i = 1, \ldots, n \tag{6.5}$$

holds for some $B_2 \ge 0$.

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

$$|\mathfrak{d}|\mathfrak{d}f(X)|| = \Big(\sum_{i=1}^n \frac{1}{2}\bar{\mathbb{E}}_i(|\mathfrak{d}f(X)| - |T_i\mathfrak{d}f(X)|)^2\Big)^{1/2} \tag{6.6}$$

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 $a := \sum_{j=1}^n{}^{(i)}(\mathfrak{d}_jf)^2$, $b := (\mathfrak{d}_if)^2$, $c := \sum_{j=1}^n{}^{(i)}(T_i\mathfrak{d}_jf)^2$ and $d := (T_i\mathfrak{d}_if)^2$ for any $1 \le i \le n$, we arrive at

$$\begin{aligned}(|\mathfrak{d}f| - |T_i\mathfrak{d}f|)^2 = (\sqrt{a+b} - \sqrt{c+d})^2 &\le \Big(|\sqrt{a} - \sqrt{c}| + \frac{|b-d|}{\sqrt{a+b}}\Big)^2 \\ &\le 2\Big((\sqrt{a} - \sqrt{c})^2 + \frac{(b-d)^2}{a+b}\Big).\end{aligned} \tag{6.7}$$

(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, we have

$$\begin{aligned}&(\sqrt{a} - \sqrt{c})^2 = \Big(\big(\sum_{j=1}^n{}^{(i)}(\mathfrak{d}_jf)^2\big)^{1/2} - \big(\sum_{j=1}^n{}^{(i)}(T_i\mathfrak{d}_jf)^2\big)^{1/2}\Big)^2 \\ &\le \sum_{j=1}^n{}^{(i)}(\mathfrak{d}_jf - T_i\mathfrak{d}_jf)^2 = \frac{1}{2}\sum_{j=1}^n{}^{(i)}\big((\bar{\mathbb{E}}_j(f - T_jf)^2)^{1/2} - (\bar{\mathbb{E}}_j(T_if - T_{ij}f)^2)^{1/2}\big)^2 \\ &\le \frac{1}{2}\sum_{j=1}^n{}^{(i)}\bar{\mathbb{E}}_j(f - T_jf - T_if + T_{ij}f)^2\end{aligned} \tag{6.8}$$

Combining (6.6), (6.7) and (6.8) together with the trivial estimate $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$ for all $x, y \ge 0$ then yields

$$|\mathfrak{d}|\mathfrak{d}f(X)|| \le \sqrt{2}\Bigg(\|\mathfrak{d}^{(2)}f(X)\|_{\mathrm{HS}} + \Bigg(\frac{1}{2}\sum_{i=1}^n \bar{\mathbb{E}}_i\frac{((\mathfrak{d}_if(X))^2 - (T_i\mathfrak{d}_if(X))^2)^2}{|\mathfrak{d}f(X)|^2}\Bigg)^{1/2}\Bigg). \tag{6.9}$$

We may further estimate the last term by

$$\Bigg(\sum_{i=1}^n \bar{\mathbb{E}}_i\frac{|(\mathfrak{d}_if(X))^2 - (T_i\mathfrak{d}_if(X))^2|}{|\mathfrak{d}f(X)|^2}\Bigg)^{1/2} \sup_{x\in\mathrm{supp}(\mu)}\max_{i=1,\ldots,n}|\mathfrak{d}_if(x)|. \tag{6.10}$$

We now claim that

$$\Bigg(\sum_{i=1}^n \bar{\mathbb{E}}_i\frac{|(\mathfrak{d}_if(X))^2 - (T_i\mathfrak{d}_if(X))^2|}{|\mathfrak{d}f(X)|^2}\Bigg)^{1/2} \le 1. \tag{6.11}$$

To see this, we first assume that the Hoeffding decomposition of f consists of the second order term only, i. e. $f(X_1, \ldots, X_n) = \sum_{i<j} h_{ij}(X_i, X_j)$. The general case is then proved similarly using the orthogonality of the Hoeffding terms. Note that by Remark 2.2(3)

and the elementary properties of the Hoeffding decomposition (1.2), we have

$$(\mathfrak{d}_i f)^2 = \frac{1}{2}\bar{\mathbb{E}}_i(\sum_{j=1}^n {}^{(i)}h_{ij} - \sum_{j=1}^n {}^{(i)}T_i h_{ij})^2 = \frac{1}{2}\Big((\sum_{j=1}^n {}^{(i)}h_{ij})^2 + \mathbb{E}_i(\sum_{j=1}^n {}^{(i)}h_{ij})^2\Big).$$

Here we set $h_{ij} = h_{ji}$ for $i > j$. Consequently, we get

$$|(\mathfrak{d}_i f)^2 - (T_i\mathfrak{d}_i f(X))^2| \le \frac{1}{2}\Big((\sum_{j=1}^n {}^{(i)}h_{ij})^2 + (\sum_{j=1}^n {}^{(i)}T_i h_{ij})^2\Big).$$

Taking expectations, we obtain

$$\bar{\mathbb{E}}_i|(\mathfrak{d}_i f)^2 - (T_i\mathfrak{d}_i f(X))^2| \le (\mathfrak{d}_i f)^2,$$

from which the claim follows in the case $f = f_2$. In the general case, we may replace $\sum_{j=1}^n {}^{(i)}h_{ij}$ by

$$\sum_{j=1}^n {}^{(i)}h_{ij} + \sum_{j<k} {}^{(i)}h_{ijk} + \ldots,$$

where the summation extends over all terms up to the order n. From here on the proof is similar to the case $f = f_2$.

Combining (6.9), (6.10) and (6.11) with the assumptions from the theorem, we therefore arrive at

$$|\mathfrak{d}|\mathfrak{d}f|| \le \sqrt{2}B_1 + B_2.$$

Moreover, by Remark 2.2(4) and Jensen's inequality (or (6.2), alternatively), we have $(\mathfrak{D}_{ij}f(x))^2 \le 4(\mathfrak{d}_{ij}f(x))^2$ and hence

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

Finally, consider the "normalized" function $f/(\sqrt{2}B_1 + B_2)$ and use $(*)$ in (5.2) and (5.3) from the proof of Theorem 1.2, respectively. The proof of Theorem 6.1 then follows by elementary computations. □

As for conditions (6.4) and (6.5), 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 symmetric Bernoulli 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

$$f(X_1, \ldots, X_n) = \alpha_0 + \sum_{i=1}^n \alpha_i X_i + \sum_{i<j} \alpha_{ij} X_i X_j + \ldots, \qquad (6.12)$$

where the coefficients α_I (with a suitable multi-index I) are real numbers and the summation extends over all terms till order n. More precisely, we have

$$\alpha_{i_1 \ldots i_d} = \mathbb{E}f(X_1, \ldots, X_n)X_{i_1} \cdots X_{i_d}$$

for any $i_1 < \ldots < i_d$, $d = 0, 1, \ldots, n$. This representation is called the *Fourier-Walsh expansion* of the function f, and the expression on the right-hand side of (6.12) is also known as a *Rademacher chaos*. It is immediately clear that (6.12) is at the same time the Hoeffding decomposition of f. Applying Corollary 1.3 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*

$$f(x_1, \ldots, x_n) := \sum_{i<j} \alpha_{ij} x_i x_j + \sum_{i<j<k} \alpha_{ijk} x_i x_j x_k + \ldots,$$

where the sum goes up to order n and the $\alpha_{i_1 \ldots i_d}$ 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 (1.4). Then, we have

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

Proof. First consider functions of the form

$$f\colon \mathbb{R}^n \to \mathbb{R}; \qquad f(x_1, \ldots, x_n) := \sum_{i<j} \alpha_{ij} x_i x_j \qquad (*)$$

with $\alpha_{ij} \in \mathbb{R}$. We also set $\alpha_{ij} = \alpha_{ji}$ for all $i > j$. Using (2.5), we easily get

$$\int \|\mathfrak{D}^{(2)} f\|_{\mathrm{HS}}^2 d\mu = \|\mathfrak{D}^{(2)} f\|_{\mathrm{HS}}^2 = \sum_{i \ne j} (\alpha_{ij} x_i x_j)^2 = \sum_{i \ne j} \alpha_{ij}^2. \tag{6.13}$$

Hence, we need to check the condition $|\mathfrak{d}|\mathfrak{d}f|| \le 1$ for all $x \in \{\pm 1\}^n$ only. By Remark 2.2(1), we can write

$$|\mathfrak{d}|\mathfrak{d}f(x)|| = \frac{1}{2}\Big(\sum_{k=1}^n (|\mathfrak{d}f(x)| - |\mathfrak{d}f(\sigma_k x)|)^2\Big)^{1/2}. \tag{6.14}$$

Now use Remark 2.2(1), (2.4) and the notations from the proof of Theorem 6.1 to show

$$(\mathfrak{d}_i f(x))^2 = (\mathfrak{D}_i f(x))^2 = \Big(\sum_{j=1}^n {}^{(i)} \alpha_{ij} x_i x_j\Big)^2.$$

If $i \ne k$, we therefore get

$$(\mathfrak{d}_i f(x))^2 = \Big(\sum_{j=1}^n {}^{(i,k)} \alpha_{ij} x_i x_j + \alpha_{ik} x_i x_k\Big)^2,$$

$$(\mathfrak{d}_i f(\sigma_k x))^2 = \Big(\sum_{j=1}^n {}^{(i,k)} \alpha_{ij} x_i x_j - \alpha_{ik} x_i x_k\Big)^2.$$

If $i = k$, we have $(\mathfrak{d}_k f(x))^2 = (\mathfrak{d}_k f(\sigma_k x))^2$. Now define two vectors v and w by

$$v_i := \begin{cases} \sum_{j=1}^n {}^{(i,k)} \alpha_{ij} x_i x_j, & i \ne k, \\ \sum_{j=1}^n {}^{(k)} \alpha_{kj} x_k x_j, & i = k, \end{cases} \qquad w_i := \begin{cases} \alpha_{ik} x_i x_k, & i \ne k, \\ 0, & i = k. \end{cases}$$

Then, we have

$$\begin{aligned} ||\mathfrak{d}f(x)| - |\mathfrak{d}f(\sigma_k(x))|| &= ||v + w| - |v - w|| \\ &\le 2|w| = 2\Big(\sum_{i=1}^n {}^{(k)} (\alpha_{ik} x_i x_k)^2\Big)^{1/2}. \end{aligned} \tag{6.15}$$

Combining (6.13), (6.14) and 6.15, we get

$$|\mathfrak{d}|\mathfrak{d}f(x)|| \leq (\sum_{i\neq k}(\alpha_{ik}x_ix_k)^2)^{1/2} = \|\mathfrak{D}^{(2)}f(x)\|_{\mathrm{HS}}$$

Now apply Corollary 1.3 to the "normalized" function f/B. Simple estimates then complete the proof for functions of type (∗).

For general functions f, note that as above, we have

$$(\mathfrak{d}_i f(x))^2 = (\mathfrak{D}_i f(x))^2 = \big(\sum_{j=1}^{n}{}^{(i)}\alpha_{ij}x_ix_j + \sum_{j<l}{}^{(i)}\alpha_{ijl}x_ix_jx_l + \ldots\big)^2,$$

where the summation extends over all terms up to order n. Consequently, if $i \neq l$, we have

$$\begin{aligned}(\mathfrak{d}_i f(x))^2 =& \big((\sum_{j=1}^{n}{}^{(i,k)}\alpha_{ij}x_ix_j + \sum_{j<l}{}^{(i,k)}\alpha_{ijl}x_ix_jx_l + \ldots)\\ &+ (\alpha_{ik}x_ix_k + \sum_{l=1}^{n}{}^{(i,k)}\alpha_{ikl}x_ix_kx_l + \ldots)\big)^2,\\ (\mathfrak{d}_i f(\sigma_k x))^2 =& \big((\sum_{j=1}^{n}{}^{(i,k)}\alpha_{ij}x_ix_j + \sum_{j<l}{}^{(i,k)}\alpha_{ijl}x_ix_jx_l + \ldots)\\ &- (\alpha_{ik}x_ix_k + \sum_{l=1}^{n}{}^{(i,k)}\alpha_{ikl}x_ix_kx_l + \ldots)\big)^2,\end{aligned}$$

while for $i = k$ we have $(\mathfrak{d}_i f(x))^2 = (\mathfrak{d}_i f(\sigma_k x))^2$. Noting that

$$\alpha_{ik}x_ix_l + \sum_{l=1}^{n}{}^{(i,k)}\alpha_{ikl}x_ix_kx_l + \ldots = \mathfrak{D}_{ik}f(x),$$

we can proceed as above, which finishes the proof. □

7. Differentiable Functions: Proofs

In order to prove Theorem 1.8, 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 $f\colon M \to \mathbb{R}$ is a continuous function, we may define the generalized modulus of the gradient by

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

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

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

for any $x \in M$.

Using the generalized modulus of the gradient, we have the following analogues of inequalities (4.1) and (4.3) 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

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

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

$$\int e^{tu^2} d\mu \leq \exp\left(\frac{t}{1 - 2\sigma^2 t} \int u^2 d\mu\right) \tag{7.4}$$

for any $0 \leq t < \frac{1}{2\sigma^2}$. Inequalities (7.3) and (7.4) are both due to S. Bobkov and F. Götze [B-G1]. As mentioned in Section 4, (7.4) actually goes back to S. Aida, T. Masuda and I. Shikegawa [A-M-S].

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 4.2, 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* (1.14) *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 $|\nabla^* |\nabla^* f|| \leq 1$. Here, $|\nabla^* f|$ is the generalized modulus of the gradient from* (7.1)*. Then, we have*

$$\int_G \exp\left(\frac{1}{2\sigma^2} f\right) d\mu \leq \exp\left(\frac{1}{2\sigma^2} \int_G |\nabla^* f|^2 d\mu\right).$$

Proposition 7.1 is a special case of [B-C-G, Proposition 2.1]. If f is a $\mathcal{C}^2$-function, the condition $|\nabla^* |\nabla^* f|| \leq 1$ 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*

$$|\nabla^* |\nabla^* f(x)|| \leq \|f''(x)\|_{\mathrm{HS}}$$

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

Proof. For the function $v(x) = |\nabla^* f(x)|^2 = |\nabla f(x)|^2 = \sum_{j=1}^n (\partial_j f(x))^2$, we have

$$\partial_i v(x) = 2 \sum_{j=1}^n \partial_{ij} f(x) \partial_j f(x), \qquad i = 1, \ldots, n.$$

Hence, by Cauchy's inequality,

$$(\partial_i v(x))^2 \leq 4 \sum_{j=1}^n (\partial_{ij} f(x))^2 \sum_{j=1}^n (\partial_j f(x))^2 = 4 |\nabla f(x)| (\partial_{ij} f(x))^2,$$

and thus

$$|\nabla v(x)| \leq 2 |\nabla f(x)| \|f''(x)\|_{\mathrm{HS}}.$$

Now consider the function $w(x) = |\nabla f(x)| = \sqrt{v(x)}$. In the region $|\nabla f(x)| \neq 0$, w is differentiable, and we get

$$|\nabla w(x)| = \frac{|\nabla v(x)|}{2\sqrt{v(x)}} \leq \|f''(x)\|_{\mathrm{HS}}$$

as required.

If $|\nabla f(x)| = 0$, we have $\partial_i f(x) = 0$ for all $i = 1, \ldots, n$. By Taylor expansion at the point x, we have

$$\partial_i f(x+h) = \sum_{j=1}^n \partial_{ij} f(x) h_j + o(|h|) \qquad \text{as } h = (h_1, \ldots, h_n) \to 0,$$

where $|h| = \sqrt{h_1^2 + \ldots + h_n^2}$. Using Cauchy's inequality, this yields

$$(\partial_i f(x+h))^2 \leq |h|^2 \sum_{j=1}^n (\partial_{ij} f(x))^2 + o(|h|^2),$$

and therefore

$$|\nabla f(x+h)| \leq |h| \|f''(x)\|_{\mathrm{HS}} + o(|h|).$$

Hence, by the definition (7.1),

$$|\nabla^* w(x)| = \limsup_{h \to 0} \frac{|w(x+h) - w(x)|}{|h|} = \limsup_{h \to 0} \frac{|\nabla f(x+h)|}{|h|} \leq \|f''(x)\|_{\mathrm{HS}}.$$

Lemma 7.2 is proved. □

We can now prove Theorem 1.8:

Proof of Theorem 1.8. Given a function f as in Theorem 1.8, applying Proposition 7.1 together with Lemma 7.2 yields

$$\int_G \exp\left(\frac{1}{2\sigma^2} f\right) d\mu \leq \exp\left(\frac{1}{2\sigma^2} \int_G |\nabla f|^2 d\mu\right). \tag{7.5}$$

Since μ satisfies a logarithmic Sobolev inequality with constant σ^2, it also satisfies a Poincaré inequality (1.13) with constant σ^2. Therefore, since $\int_G \partial_i f d\mu = 0$ for all i, we have

$$\int_G (\partial_i f)^2 d\mu \leq \sigma^2 \sum_{j=1}^n \int_G (\partial_{ij} f)^2 d\mu$$

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

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

Combining (7.5), (7.6) and the assumptions from Theorem 1.8, we arrive at

$$\int_G \exp\left(\frac{1}{2\sigma^2} f\right) d\mu \leq \exp\left(\frac{1}{2} \int_G \|f''\|_{\mathrm{HS}}^2 d\mu\right) \leq \exp\left(\frac{b^2}{2}\right).$$

The rest of the proof is similar to the proof of Theorem 1.2. We finally arrive at the inequality

$$\frac{1}{2\sigma^2 \kappa} \leq \frac{\log 2}{2\sigma^2 \log 2 + b^2 \sigma^2}.$$

Noting that

$$\frac{\log 2}{2\sigma^2 \log 2 + x\sigma^2} \geq \frac{1}{2\sigma^2(1+x)}$$

for all $x \geq 0$ finishes the proof. □

Finally, we prove Theorem 1.9:

Proof of Theorem 1.9. The proof is similar to the proof of Theorem 1.5 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

$$\varphi_1(x) = \sum_{i=1}^n \mu(\partial_i f)(x_i - \mu(x_i)), \qquad \varphi_2(x) = f(x) - \varphi_1(x).$$

We now apply the basic argument (5.4) from the proof of Theorem 1.5. Here we need to check that

$$\int e^{c_i|\varphi_i|} d\mu \leq 2$$

for $i = 1, 2$ and some constants $c_1, c_2 > 0$. By Theorem 1.8 applied to φ_2, we may choose

$$\frac{1}{2} c_2 = \frac{1}{4\sigma^2(1+b^2)}.$$

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

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

for all $\lambda > 0$. Proceeding as in the proof of Theorem 1.5, we obtain

$$\frac{1}{2} c_1 = \frac{(\log 2)^{1/2}}{4\sigma b_0} \geq \frac{1}{5\sigma b_0},$$

which easily yields the desired result. □

References

[A-W] Adamczak, R., Wolff, P.: *Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order.* Probability Theory Related Fields 162(3) (2015), 531–586.

[A-M-S] Aida, S., Masuda, T., Shikegawa, I.: *Logarithmic Sobolev inequalities and exponential integrability.* J. Func. Anal. 126 (1994), 83–101.

[A-G] Arcones, M. A., Giné, E.: *Limit theorems for U-processes.* Ann. Prob. 21 (1993), 1494–1542.

[B-G-Z] Bickel, P. J., Götze, F., van Zwet, W. R.: *Resampling fewer than n observations: gains, losses, and remedies for losses.* Statistica Sinica 7 (1997), 1–32.

[B-C-G] Bobkov, S. G., Chistyakov, G. P., Götze, F.: *Second Order Concentration on the Sphere.* Preprint, arXiv:1502.04178.

[B-G1] Bobkov, S. G., Götze, F.: *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities.* J. Func. Anal. 163(1) (1999), 1–28.

[B-G2] Bobkov, S. G., Götze, F.: *Concentration inequalities and limit theorems for randomized sums.* Probability Theory Related Fields 137 (2007), 49–81.

[B-G3] Bobkov, S. G., Götze, F.: *Concentration of empirical distribution functions with applications to non-i.i.d. models.* Bernoulli 16(4) (2010), 1385–1414.

[B-G-H] Bobkov, S. G., Götze, F., Houdré, C.: *On Gaussian and Bernoulli covariance representations.* Bernoulli 7(3) (2001), 439–451.

[B-G-T] Bobkov, S. G., Götze, F., Tikhomirov, A. N.: *On Concentration of Empirical Measures and Convergence to the Semi-circle Law*. J. Theor. Probab. 23 (2010), 792–823.

[B-L] Bobkov, S. G., Ledoux, M.: *From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities*. Geom. Funct. Anal. 10(5) (2000), 1028–1052.

[CE-F-M] Cordero-Erausquin, D., Fradelizi, M., Maurey, B.: The (B) conjecture for the Gaussian measure of dilates of symmetric convex sets and related problems. J. Funct. Anal. 214(2) (2004), 410–427.

[D-G] de la Peña, V., Giné, E.: *Decoupling. From Dependence to Independence*. Springer, 1999.

[G-G-M] Giné, E., Götze, F., Mason, D. M.: *When is the Student t-statistic asymptotically standard normal?*. Ann. Prob. 25(3) (1997), 1514–1531.

[G] Götze, F. IMS Bulletin (1993).

[Gr] Gross, L.: *Logarithmic Sobolev inequalities*. Amer. J. Math 97(4) (1975), 1061–1083.

[G-Z] Guionnet, A., Zeitouni, O.: *Concentration of the spectral measure for large matrices*. Electron. Commun. Prob. 5 (2000), 119–136.

[H] Hoeffding, W.: *A class of statistics with asymptotically normal distribution*. Ann. Math. Statist. 19 (1948), 293–325.

[J] Johansson, K.: *On fluctuations of eigenvalues of random Hermitean matrices*. Duke Math. J. 91(1) (1998), 151–204.

[K-K-P] Khorunzhy, A. M., Khoruzhenko, B. A., Pastur, L. A.: *Asymptotic properties of large random matrices with independent entries*. J. Math. Phys. 37 (1996), 5033-5060.

[L1] Ledoux, M.: *On Talagrand's deviation inequalities for product measures*. ESAIM Prob. & Stat. 1 (1996), 63–87.

[L2] Ledoux, M.: *Concentration of measure and logarithmic Sobolev inequalities. Séminaire de Probabilités XXXIII*. Lecture Notes in Math. 1709, 120–216. Springer, 1999.

[L3] Ledoux, M.: *The Concentration of Measure Phenomenon*. American Mathematical Society, 2001.

[M] Major, P.: *On the estimation of multiple random integrals and U-statistics*. Lecture Notes in Math. 2079. Springer, 2013.

[M-O-O] Mossel, E., O'Donnell, R., Oleszkiewicz, K.: *Noise stability of functions with low influences: Invariance and optimality*. Ann. Math. 171 (2010), 295–341.

[P-R] Politis, D. N., Romano, J. P.: *Large sample confidence regions based on subsamples under minimal assumptions*. Ann. Statist. 22 (1994), 2031–2050.

[R-L-T-V] Robins, J., Li, L., Tchetgen Tchetgen, E., van der Vaart, A.: *Asymptotic Normality of Quadratic Estimators*. Preprint, arXiv:1512.02280.

[R-V] Rudelson, M., Vershynin, R.: *Hanson-Wright inequality and sub-gaussian concentration*. Electronic Communications in Probability 18 (2013), 1–9.

[S] Sambale, H.: *Second Order Concentration for Functions of Independent Random Variables*. Ph.D. thesis, Bielefeld Univ., 2015.

[S-S] Sinai, Ya., Soshnikov, A.: *A central limit theorem for traces of large random matrices with independent matrix elements*. Bol. Soc. Brasil. Mat (N.S.) 29 (1998), 1–24.

[T1] Talagrand, M.: *Concentration of measure and isoperimetric inequalities in product spaces*. Publications Mathématiques de l'I.H.E.S. 81(9) (1995), 73–205.

[T2] Talagrand, M.: *New concentration inequalities in product spaces*. Invent. Math. 126 (1996), 505–563.

Friedrich Götze, Faculty of Mathematics, Bielefeld University, Bielefeld, Germany

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

Holger Sambale, Faculty of Mathematics, Bielefeld University, Bielefeld, Germany

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