

HIGHER ORDER CONCENTRATION OF MEASURE

S. G. BOBKOV, F. GÖTZE, AND H. SAMBALE

ABSTRACT. We study sharpened forms of the concentration of measure phenomenon typically centered at stochastic expansions of order $d - 1$ for any $d \in \mathbb{N}$. The bounds are based on d -th order derivatives or difference operators. In particular, we consider deviations of functions of independent random variables and differentiable functions over probability measures satisfying a logarithmic Sobolev inequality, and functions on the unit sphere. Applications include concentration inequalities for U -statistics as well as for classes of symmetric functions via polynomial approximations on the sphere (Edgeworth-type expansions).

1. INTRODUCTION

In this article, we study higher order versions of the concentration of measure phenomenon. Referring to the use of derivatives or difference operators of higher order, say d , the notion of *higher* order concentration has several aspects. In particular, instead of the classical problem about deviations of f around the mean $\mathbb{E}f$, one may consider potentially smaller fluctuations of $f - \mathbb{E}f - f_1 - \dots - f_d$, where f_1, \dots, f_d are “lower order terms” of f with respect to a suitable decomposition, such as a Taylor-type decomposition or the Hoeffding decomposition of f .

Starting with the works of Milman in local theory of Banach spaces, and of Borell, Sudakov, and Tsirelson within the framework of Gaussian processes, the concentration of measure phenomenon has been intensely studied during the past decades. This study includes important contributions due to Talagrand and other researchers in the 1990s, cf. Milman and Schechtman [M-S], Talagrand [T], Ledoux [L1], [L2], [L3]; a more recent survey is authored by Boucheron, Lugosi and Massart [B-L-M].

As another previous work, let us mention Adamczak and Wolff [A-W], who exploited certain Sobolev-type inequalities or subgaussian tail conditions to derive exponential tail inequalities for functions with bounded higher-order derivatives (evaluated in terms of tensor-product matrix norms). While in [A-W], concentration around the mean is studied, the idea of sharpening concentration inequalities for Gaussian measures by requiring orthogonality to linear functions also appears in Wolff [W] as well as in Cordero-Erausquin, Fradelizi and Maurey [CE-F-M].

Our research started with second order results for functions on the n -sphere orthogonal to linear functions [B-C-G], with an approach which was continued in [G-S] in presence of logarithmic Sobolev inequalities. This includes discrete models as well as differentiable functions on open subsets of \mathbb{R}^n . Here, we adapt in particular Sobolev type inequalities introduced by Boucheron, Bousquet, Lugosi and Massart

Date: May 15, 2018.

1991 Mathematics Subject Classification. Primary 60E15, 60F10; secondary 41A10, 41A80.

Key words and phrases. Concentration of measure phenomenon, logarithmic Sobolev inequalities, Hoeffding decomposition, functions on the discrete cube, Efron–Stein inequality.

This research was supported by CRC 1283. The work of S. G. Bobkov was supported by NSF grant DMS-1612961 and by the Russian Academic Excellence Project ‘5-100’.

[B-B-L-M], and thus extend some of the results from [G-S] to arbitrary higher orders. Developing the algebra of higher order difference operators, we moreover came across a higher order extension of the well-known Efron–Stein inequality.

1.1. Functions of independent random variables. Let $X = (X_1, \dots, X_n)$ be a random vector in \mathbb{R}^n with independent components, defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. First, we state higher order exponential inequalities in terms of the difference operator which is frequently used in the method of bounded differences.

Let $(\bar{X}_1, \dots, \bar{X}_n)$ be an independent copy of X . Given $f(X) \in L^\infty(\mathbb{P})$, define

$$T_i f(X) = T_i f = f(X_1, \dots, X_{i-1}, \bar{X}_i, X_{i+1}, \dots, X_n), \quad i = 1, \dots, n,$$

$$(1.1) \quad \mathfrak{h}_i f(X) = \frac{1}{2} \|f(X) - T_i f(X)\|_{i, \infty}, \quad \mathfrak{h}f = (\mathfrak{h}_1 f, \dots, \mathfrak{h}_n f),$$

where $\|\cdot\|_{i, \infty}$ denotes the L^∞ -norm with respect to (X_i, \bar{X}_i) . Depending on the random variables X_j , $j \neq i$, $\mathfrak{h}_i f$ thus provides a uniform upper bound on the differences with respect to the i -th coordinate (up to constant). Based on \mathfrak{h} , it is possible to define higher order difference operators $\mathfrak{h}_{i_1 \dots i_d}$ ($d \in \mathbb{N}$) by setting

$$(1.2) \quad \begin{aligned} \mathfrak{h}_{i_1 \dots i_d} f(X) &= \frac{1}{2^d} \left\| \prod_{s=1}^d (Id - T_{i_s}) f(X) \right\|_{i_1, \dots, i_d, \infty} \\ &= \frac{1}{2^d} \left\| f(X) + \sum_{k=1}^d (-1)^k \sum_{1 \leq s_1 < \dots < s_k \leq d} T_{i_{s_1} \dots i_{s_k}} f(X) \right\|_{i_1, \dots, i_d, \infty}, \end{aligned}$$

where $T_{i_1 \dots i_d} = T_{i_1} \circ \dots \circ T_{i_d}$, and $\|\cdot\|_{i_1, \dots, i_d, \infty}$ denotes the L^∞ -norm with respect to X_{i_1}, \dots, X_{i_d} and $\bar{X}_{i_1}, \dots, \bar{X}_{i_d}$. For instance,

$$\mathfrak{h}_{ij} f = \frac{1}{4} \|f - T_i f - T_j f + T_{ij} f\|_{i, j, \infty} \quad \text{for } i \neq j.$$

Based on (1.2), we define hypermatrices of d -th order differences as follows:

$$(1.3) \quad (\mathfrak{h}^{(d)} f(X))_{i_1 \dots i_d} = \begin{cases} \mathfrak{h}_{i_1 \dots i_d} f(X), & \text{if } i_1, \dots, i_d \text{ are distinct,} \\ 0, & \text{else.} \end{cases}$$

For short, we freely write $\mathfrak{h}^{(d)} f$ instead of $\mathfrak{h}^{(d)} f(X)$. Since $T_{ii} \equiv T_i$, we necessarily have $\mathfrak{h}_{ii} f = \frac{1}{2} \mathfrak{h}_i f$. Therefore, removing the d -th order differences in which some indexes appear more than once can be interpreted as removing lower order differences.

Moreover, define $|\mathfrak{h}^{(d)} f|_{\text{HS}}$ to be the Euclidean norm of $\mathfrak{h}^{(d)} f$ regarded as an element of \mathbb{R}^{n^d} . For instance, $|\mathfrak{h}^{(1)} f|_{\text{HS}}$ is the Euclidean norm of $\mathfrak{h}f$, and $|\mathfrak{h}^{(2)} f|_{\text{HS}}$ is the Hilbert–Schmidt norm of the ‘‘Hessian’’ $\mathfrak{h}^{(2)} f$. Also, put

$$(1.4) \quad \|\mathfrak{h}^{(d)} f\|_{\text{HS}, p} = (\mathbb{E} |\mathfrak{h}^{(d)} f|_{\text{HS}}^p)^{1/p}, \quad p \in (0, \infty].$$

Using these notations, the following result holds for any fixed integer $d = 1, \dots, n$.

Theorem 1.1. *Let $f = f(X)$ be in $L^\infty(\mathbb{P})$ with $\mathbb{E}f = 0$. If the conditions*

$$(1.5) \quad \|\mathfrak{h}^{(k)} f\|_{\text{HS}, 2} \leq 1 \quad (k = 1, \dots, d-1)$$

and

$$(1.6) \quad \|\mathfrak{h}^{(d)} f\|_{\text{HS}, \infty} \leq 1$$

are satisfied, then, for some universal constant $c > 0$,

$$\mathbb{E} \exp (c|f|^{2/d}) \leq 2.$$

Here, a possible choice is $c = 1/(208 e)$.

In the case $d = 1$, (1.5) does not contain any constraint, while (1.6) means the boundedness of fluctuations of f along every coordinate. Here we arrive at the well-known assertion on Gaussian deviations of $f(X)$. For growing d , the conclusion is getting somewhat weaker, however it holds as well under the potentially much weaker assumptions (1.5)–(1.6). To interpret them in case $d \geq 2$, let us recall the notion of a Hoeffding decomposition, introduced by Hoeffding in 1948 [Hoe]. Given a function $f(X) \in L^1(\mathbb{P})$, it is the unique decomposition

$$\begin{aligned} (1.7) \quad f(X_1, \dots, X_n) &= \mathbb{E}f(X) + \sum_{1 \leq i \leq n} h_i(X_i) + \sum_{1 \leq i < j \leq n} h_{ij}(X_i, X_j) + \dots \\ &= f_0 + f_1 + f_2 + \dots + f_n \end{aligned}$$

such that $\mathbb{E}_{i_s} h_{i_1 \dots i_k}(X_{i_1}, \dots, X_{i_k}) = 0$ whenever $1 \leq i_1 < \dots < i_k \leq n$, $s = 1, \dots, k$, where \mathbb{E}_i denotes the expectation with respect to X_i . The sum f_d is called Hoeffding term of degree d or simply d -th Hoeffding term of f . Provided that $f(X) \in L^2(\mathbb{P})$, the system $\{f_i(X)\}_{i=0}^n$ forms an orthogonal decomposition of $f(X)$ in $L^2(\mathbb{P})$.

It is not hard to see that $\mathfrak{h}_{i_1 \dots i_k} f = \mathfrak{h}_{i_1 \dots i_k}(\sum_{i=k}^n f_i)$ whenever $i_1 < \dots < i_k$, $k \leq d$. In this sense, (1.5) controls the lower order Hoeffding terms f_1, \dots, f_{d-1} , while the behaviour of $\sum_{i=d}^n f_i$ is mainly controlled by (1.6). The relationship between these two conditions may be illustrated by considering a special class of functions f like *multilinear polynomials*, that is

$$\begin{aligned} (1.8) \quad f(X_1, \dots, X_n) &= \alpha_0 + \sum_{i=1}^n \alpha_i X_i + \sum_{i < j} \alpha_{ij} X_i X_j + \dots \\ &= f_0 + f_1 + f_2 + \dots + f_n \quad (\alpha_I \in \mathbb{R}). \end{aligned}$$

Proposition 1.2. *Let X_1, \dots, X_n be bounded and such that $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$ for $i = 1, \dots, d$. If $f(X)$ is a multilinear polynomial (1.8) of the form $f = \sum_{k=d}^n f_k$, then (1.6) implies (1.5).*

Note that under the conditions of Proposition 1.2, the Hoeffding decomposition of f can be read off the polynomial structure: $h_{i_1 \dots i_k}(X_{i_1}, \dots, X_{i_k}) = \alpha_{i_1 \dots i_k} X_{i_1} \cdots X_{i_k}$.

In particular, let X_1, \dots, X_n be independent Rademacher variables, i. e. X_i 's have distribution $\frac{1}{2}\delta_{+1} + \frac{1}{2}\delta_{-1}$, where δ_x denotes the Dirac measure in x . In this case, any function $f(X)$ can be written as a multilinear polynomial with coefficients

$$\alpha_{i_1 \dots i_k} = \mathbb{E}f(X_1, \dots, X_n) X_{i_1} \cdots X_{i_k}, \quad 1 \leq i_1 < \dots < i_k \leq n.$$

This representation is known as *Fourier–Walsh expansion* of f . Consequently, Theorem 1.1 yields a d -th order concentration result on the discrete hypercube, and if f has Fourier–Walsh expansion of type $f = \sum_{k=d}^n f_k$ satisfying (1.6), the concentration bound given in Theorem 1.1 holds.

Using Rademacher variables in Theorem 1.1 gives rise to a concentration inequality for U -statistics with completely degenerate kernel functions. There are many results on the distributional properties of U -statistics (cf. de la Peña and Giné

[D-G] for an overview). Starting with Hoeffding's inequalities (e. g. [D-G], Theorem 4.1.8), we especially refer to the results by Arcones and Giné [A-G] and Major [M]. By combining elements of the proof of Theorem 1.1 and a classical result on randomized U -statistics by de la Peña and Giné, we arrive at the following:

Corollary 1.3. *Let X_1, \dots, X_n be i.i.d. random elements in a measurable space (S, \mathcal{S}) , and h be a measurable function on S^d ($1 \leq d \leq n$) such that $|h| \leq M$ for some constant M . If h is completely degenerate, i. e. $\mathbb{E}_i h(X_1, \dots, X_d) = 0$ for all $i = 1, \dots, d$, then, for some positive constant $c = c(d, M)$, the U -statistic*

$$f(X_1, \dots, X_n) = \frac{(n-d)!}{n!} \sum_{i_1 \neq \dots \neq i_d} h(X_{i_1}, \dots, X_{i_d})$$

satisfies

$$\mathbb{E} \exp(cn |f|^{2/d}) \leq 2.$$

By Chebychev's inequality, Theorem 1.1 immediately yields a deviation bound

$$\mathbb{P}\{|f(X)| \geq t\} \leq 2e^{-ct^{2/d}}, \quad t \geq 0.$$

More precisely, we get refined tail estimates similar to Adamczak [A], Theorem 7, or Adamczak and Wolff [A-W], Theorem 3.3.

Corollary 1.4. *Let $f = f(X)$ be in $L^\infty(\mathbb{P})$ with $\mathbb{E}f = 0$. For all $t \geq 0$, putting*

$$\eta_f(t) = \min \left(\frac{t^{2/d}}{\|\mathfrak{h}^{(d)}f\|_{\text{HS},\infty}^{2/d}}, \min_{k=1,\dots,d-1} \frac{t^{2/k}}{\|\mathfrak{h}^{(k)}f\|_{\text{HS},2}^{2/k}} \right),$$

we have

$$\mathbb{P}\{|f| \geq t\} \leq e^2 \exp\{-\eta_f(t)/41(de)^2\}.$$

Moreover, it is possible to give a version of Theorem 1.1 for suprema of suitable classes of functions. To this end, we need some more notation. Let \mathfrak{F} be a class of functions $f = f(X)$ in $L^\infty(\mathbb{P})$, where as before $X = (X_1, \dots, X_n)$ is a vector of independent random variables. Then, for $i_1 \neq \dots \neq i_d$, $d = 1, \dots, n$, we define

$$\mathfrak{h}_{i_1 \dots i_d}^*(\mathfrak{F}) = \sup_{f \in \mathfrak{F}} \mathfrak{h}_{i_1 \dots i_d} f(X)$$

as a structural supremum (eventually taken over a countable subset of \mathfrak{F}), and put

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

The notations used in (1.4) are similarly adapted. This leads to the following result.

Theorem 1.5. *If $\|\mathfrak{h}^{*(k)}(\mathfrak{F})\|_{\text{HS},2} \leq 1$ for all $k = 1, \dots, d-1$ and $\|\mathfrak{h}^{*(d)}(\mathfrak{F})\|_{\text{HS},\infty} \leq 1$, then*

$$\mathbb{E} \exp \left\{ c \left| \sup_{f \in \mathfrak{F}} |f| - \mathbb{E} \sup_{f \in \mathfrak{F}} |f| \right|^{2/d} \right\} \leq 2$$

with some universal constant $c > 0$.

Finally, to provide another application, recall the example of additive functionals of partial sums (e. g. random walks)

$$(1.9) \quad S_f = S_f(X) = \sum_{i=1}^n f\left(\sum_{j=1}^i X_j\right).$$

In [G-S] we proved a second order concentration result for functionals of this type, which may be reproved and somewhat sharpened by applying Corollary 1.4:

Proposition 1.6. *Given a bounded, Borel measurable function $f: \mathbb{R} \rightarrow \mathbb{R}$, for any $t \geq 0$,*

$$\mathbb{P}(|S_f - \mathbb{E}S_f| \geq t) \leq e^2 \exp \left\{ -c \min \left(\frac{t^2}{n^3 \|f\|_\infty^2}, \frac{t}{n^2 \|f\|_\infty} \right) \right\},$$

where $c > 0$ is some numerical constant.

1.2. Higher order Efron–Stein inequality. Given independent random variables X_1, \dots, X_n , we denote by $\mathbb{E}_i f(X) = \mathbb{E}_i f$ and $\text{Var}_i f(X) = \mathbb{E}_i (f(X) - \mathbb{E}_i f(X))^2$ the expected value and variance with respect to X_i . By a well-known result of Efron and Stein [E-S], the variance functional is subadditive in the sense that

$$(1.10) \quad \text{Var} f(X) \leq \mathbb{E} \sum_{i=1}^n \text{Var}_i f(X).$$

It is possible to restate (1.10) in terms of difference operators which we introduce below. As before, let $\bar{X}_1, \dots, \bar{X}_n$ be a set of independent copies of X_1, \dots, X_n and $T_i f = f(X_1, \dots, X_{i-1}, \bar{X}_i, X_{i+1}, \dots, X_n)$. We use $\bar{\mathbb{E}}_i$ to denote the expectation with respect to \bar{X}_i , and put $x_+ = \max(x, 0)$ and $x_- = \max(-x, 0)$ for a number x .

Definition 1.7. *Let $f(X) = f(X_1, \dots, X_n)$ be a measurable function. For $i = 1, \dots, n$, under proper integrability assumptions, put:*

(i)

$$\mathbf{v}_i f(X) = (\text{Var}_i f(X))^{1/2}, \quad \mathbf{v} f = (\mathbf{v}_1 f, \dots, \mathbf{v}_n f);$$

(ii)

$$\mathfrak{D}_i f(X) = f(X) - \mathbb{E}_i f(X), \quad \mathfrak{D} f = (\mathfrak{D}_1 f, \dots, \mathfrak{D}_n f);$$

(iii)

$$\mathfrak{d}_i f(X) = \left(\frac{1}{2} \bar{\mathbb{E}}_i (f(X) - T_i f(X))^2 \right)^{1/2}, \quad \mathfrak{d} f = (\mathfrak{d}_1 f, \dots, \mathfrak{d}_n f);$$

(iv)

$$\mathfrak{d}_i^+ f(X) = \left(\frac{1}{2} \bar{\mathbb{E}}_i (f(X) - T_i f(X))_+^2 \right)^{1/2}, \quad \mathfrak{d}^+ f = (\mathfrak{d}_1^+ f, \dots, \mathfrak{d}_n^+ f);$$

(v)

$$\mathfrak{d}_i^- f(X) = \left(\frac{1}{2} \bar{\mathbb{E}}_i (f(X) - T_i f(X))_-^2 \right)^{1/2}, \quad \mathfrak{d}^- f = (\mathfrak{d}_1^- f, \dots, \mathfrak{d}_n^- f).$$

Various relations between these difference operators are discussed in Section 3. In particular, it is easy to see that, for $f(X) \in L^2(\mathbb{P})$,

$$(1.11) \quad \mathbb{E} |\mathbf{v}f|^2 = \mathbb{E} |\mathfrak{D}f|^2 = \mathbb{E} |\mathfrak{d}f|^2 = 2 \mathbb{E} |\mathfrak{d}^+ f|^2 = 2 \mathbb{E} |\mathfrak{d}^- f|^2,$$

where $|\cdot|$ denotes the Euclidean norm in \mathbb{R}^n . Therefore, we may equivalently state the Efron–Stein inequality as

$$(1.12) \quad \begin{aligned} \text{Var}f &\leq \mathbb{E} |\mathbf{v}f|^2, & \text{Var}f &\leq \mathbb{E} |\mathfrak{D}f|^2, & \text{Var}f &\leq \mathbb{E} |\mathfrak{d}f|^2, \\ \text{Var}f &\leq 2 \mathbb{E} |\mathfrak{d}^+ f|^2 & \text{or} & & \text{Var}f &\leq 2 \mathbb{E} |\mathfrak{d}^- f|^2. \end{aligned}$$

Equality in (1.12) holds iff the Hoeffding decomposition of f consists of the expected value and the first order term only, namely for $f(X) = \mathbb{E}f(X) + \sum_{i=1}^n h_i(X_i)$. Thus, the Efron–Stein inequality may be restated as the fact that any product probability measure satisfies a Poincaré-type inequality with respect to any of the difference operators \mathbf{v} , \mathfrak{D} and \mathfrak{d} with constant $\sigma^2 = 1$ (like (1.17) below). The same statement applies as well to the difference operators \mathfrak{d}^+ and \mathfrak{d}^- but with constant $\sigma^2 = 2$.

To introduce higher order versions of the Efron–Stein inequality, we need to define higher order differences based on the difference operators from Definition 1.7. For \mathfrak{D} , this is achieved by iteration, i.e. $\mathfrak{D}_{ij}f = \mathfrak{D}_i(\mathfrak{D}_j f)$ or, in general, $\mathfrak{D}_{i_1 \dots i_d} f = \mathfrak{D}_{i_1}(\dots(\mathfrak{D}_{i_d} f))$ for $1 \leq i_1, \dots, i_d \leq n$. To generalize \mathbf{v} , we set similarly to (1.2)

$$(1.13) \quad \begin{aligned} \mathbf{v}_{i_1 \dots i_d} f(X) &= \left(\mathbb{E}_{i_1 \dots i_d} \left(\prod_{s=1}^d (Id - \mathbb{E}_{i_s}) f(X) \right)^2 \right)^{1/2} \\ &= \left(\mathbb{E}_{i_1 \dots i_d} \left(f(X) + \sum_{k=1}^d (-1)^k \sum_{1 \leq s_1 < \dots < s_k \leq d} \mathbb{E}_{i_{s_1} \dots i_{s_k}} f(X) \right)^2 \right)^{1/2}, \end{aligned}$$

Here, $\mathbb{E}_{i_1 \dots i_d}$ means taking the expectation with respect to X_{i_1}, \dots, X_{i_d} . For instance,

$$\mathbf{v}_{ij} f = \left(\mathbb{E}_{ij} (f - \mathbb{E}_i f - \mathbb{E}_j f + \mathbb{E}_{ij} f)^2 \right)^{1/2}, \quad 1 \leq i < j \leq n.$$

In particular, $\mathbf{v}_{ij} f \neq (\text{Var}_{ij} f)^{1/2}$. One major difference is that $\mathbf{v}_{ij} f$ annihilates first order Hoeffding terms, but $\text{Var}_{ij} f$ does not. Similar remarks hold for any $d \geq 2$.

Finally, in case of \mathfrak{d} , we define

$$(1.14) \quad \begin{aligned} \mathfrak{d}_{i_1 \dots i_d} f(X) &= \left(\frac{1}{2^d} \bar{\mathbb{E}}_{i_1 \dots i_d} \left(\prod_{s=1}^d (Id - T_{i_s}) f(X) \right)^2 \right)^{1/2} \\ &= \left(\frac{1}{2^d} \bar{\mathbb{E}}_{i_1 \dots i_d} \left(f(X) + \sum_{k=1}^d (-1)^k \sum_{1 \leq s_1 < \dots < s_k \leq d} T_{i_{s_1} \dots i_{s_k}} f(X) \right)^2 \right)^{1/2}. \end{aligned}$$

Here, $\bar{\mathbb{E}}_{i_1 \dots i_d}$ means taking the expectation with respect to $\bar{X}_{i_1}, \dots, \bar{X}_{i_d}$, recalling that $T_{i_1 \dots i_d} = T_{i_1} \circ \dots \circ T_{i_d}$. For \mathfrak{d}^\pm , a variant of (1.14) holds by setting

$$\mathfrak{d}_{i_1 \dots i_d}^\pm f(X) = \left(\frac{1}{2^d} \bar{\mathbb{E}}_{i_1 \dots i_d} \left(\prod_{s=1}^d (Id - T_{i_s}) f(X) \right)_\pm^2 \right)^{1/2}.$$

In the same way as in (1.3), we may define d -th order hyper-matrices with respect to any of the difference operators introduced above, e. g.

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

The hyper-matrices $\mathfrak{D}^{(d)}f$, $\mathfrak{d}^{(d)}f$ and $\mathfrak{d}^{\pm(d)}f$ are defined analogously. As in case of $\mathfrak{h}^{(d)}f$, we equip these hyper-matrices with the respective Hilbert–Schmidt type norms. We are now ready to formulate the following generalization of (1.12).

Theorem 1.8 (Higher Order Efron–Stein Inequality). *Let X_1, \dots, X_n be independent random variables, and assume that $f(X) \in L^2(\mathbb{P})$ admits a Hoeffding decomposition of type $f = \mathbb{E}f + \sum_{k=d}^n f_k$ for some $1 \leq d \leq n$. Then*

$$\text{Var}f \leq \frac{1}{d!} \mathbb{E}|\mathbf{v}^{(d)}f|^2, \quad \text{Var}f \leq \frac{1}{d!} \mathbb{E}|\mathfrak{D}^{(d)}f|^2, \quad \text{Var}f \leq \frac{1}{d!} \mathbb{E}|\mathfrak{d}^{(d)}f|^2.$$

Moreover,

$$\text{Var}f \leq \frac{2}{d!} \mathbb{E}|\mathfrak{d}^{+(d)}f|^2 \quad \text{and} \quad \text{Var}f \leq \frac{2}{d!} \mathbb{E}|\mathfrak{d}^{-(d)}f|^2.$$

Equality holds if and only if the Hoeffding decomposition of f consists of the expected value and the d -th order term only, i. e. $f = \mathbb{E}f + f_d$.

In particular, Theorem 1.8 yields the following formula for the variance of an arbitrary function $f = f(X) \in L^2(\mathbb{P})$ with Hoeffding decomposition $f = \sum_{k=0}^n f_k$:

$$(1.15) \quad \text{Var}f = \sum_{k=1}^n \frac{1}{k!} \mathbb{E}|\mathbf{v}^{(k)}f_k|^2 = \sum_{k=1}^n \frac{1}{k!} \mathbb{E}|\mathfrak{D}^{(k)}f_k|^2 = \sum_{k=1}^n \frac{1}{k!} \mathbb{E}|\mathfrak{d}^{(k)}f_k|^2.$$

This result is related to the work of Houdré [Hou], who studied iterations of the Efron–Stein inequality for symmetric functions in the context of the jackknife estimate of the variance. In particular, he obtained formulas for the variance in terms of certain higher order difference operators adapted to this situation. Following the lines of our proofs, it is possible to extend his results to arbitrary functions of independent random variables. To provide an example, we may prove that

$$(1.16) \quad \text{Var}f = \sum_{k=1}^n \frac{(-1)^{k+1}}{k!} \mathbb{E}|\mathbf{v}^{(k)}f|^2,$$

which is an extension of (1.3) from [Hou]. As always, here the difference operator \mathbf{v} can be replaced by \mathfrak{D} , \mathfrak{d} and (up to a factor 2) \mathfrak{d}^{\pm} .

1.3. Differentiable Functions. In the following we shall develop higher order concentration in the setting of smooth functions on \mathbb{R}^n . Here we may derive similar results in the spirit of Adamczak and Wolff [A-W], when the underlying probability measure satisfies a logarithmic Sobolev inequality. Let us recall that a Borel probability measure μ on an open set $G \subset \mathbb{R}^n$ is said to satisfy a *Poincaré-type* and respectively a *logarithmic Sobolev inequality* with constant $\sigma^2 > 0$, if for any bounded smooth function f on G with gradient ∇f , respectively

$$(1.17) \quad \text{Var}_{\mu}(f) \leq \sigma^2 \int |\nabla f|^2 d\mu,$$

$$(1.18) \quad \text{Ent}_\mu(f^2) \leq 2\sigma^2 \int |\nabla f|^2 d\mu.$$

Here, $\text{Var}_\mu(f) = \int f^2 d\mu - (\int f d\mu)^2$ is the variance, and $\text{Ent}_\mu(f^2) = \int f^2 \log f^2 d\mu - \int f^2 d\mu \log \int f^2 d\mu$ is the entropy functional. Logarithmic Sobolev inequalities are stronger than Poincaré inequalities, in the sense that (1.18) implies (1.17).

Given a function $f \in C^d(G)$, we define $f^{(d)}$ to be the (hyper-) matrix whose entries

$$(1.19) \quad f_{i_1 \dots i_d}^{(d)}(x) = \partial_{i_1 \dots i_d} f(x), \quad d = 1, 2, \dots$$

represent the d -fold (continuous) partial derivatives of f at $x \in G$. By considering $f^{(d)}(x)$ as a symmetric multilinear d -form, we define operator-type norms by

$$(1.20) \quad |f^{(d)}(x)|_{\text{Op}} = \sup \{f^{(d)}(x)[v_1, \dots, v_d] : |v_1| = \dots = |v_d| = 1\}.$$

For instance, $|f^{(1)}(x)|_{\text{Op}}$ is the Euclidean norm of the gradient $\nabla f(x)$, and $|f^{(2)}(x)|_{\text{Op}}$ is the operator norm of the Hessian $f''(x)$. Furthermore, similarly to (1.4), we will use the short-hand notation

$$(1.21) \quad \|f^{(d)}\|_{\text{Op},p} = \left(\int_G |f^{(d)}|_{\text{Op}}^p d\mu \right)^{1/p}, \quad p \in (0, \infty].$$

We now have the following results, assuming that μ is a probability measure on G satisfying a logarithmic Sobolev inequality with constant $\sigma^2 > 0$.

Theorem 1.9. *Let $f: G \rightarrow \mathbb{R}$ be a C^d -smooth function with $\int_G f d\mu = 0$. If*

$$(1.22) \quad \|f^{(k)}\|_{\text{Op},2} \leq \min(1, \sigma^{d-k}) \quad \forall k = 1, \dots, d-1$$

and

$$(1.23) \quad \|f^{(d)}\|_{\text{Op},\infty} \leq 1,$$

then with some universal constant $c > 0$ we have

$$\int_G \exp \left\{ \frac{c}{\sigma^2} |f|^{2/d} \right\} d\mu \leq 2.$$

Here, a possible choice is $c = 1/(8e)$. If f has centered partial derivatives of order up to $d-1$, it is possible to replace (1.22) by a possibly simpler condition. To this end, as in the previous section, we need to involve Hilbert–Schmidt-type norms $|f^{(d)}(x)|_{\text{HS}}$ which are defined by taking the Euclidean norm of $f^{(d)}(x) \in \mathbb{R}^{n^d}$. As in (1.4), $\|f^{(d)}\|_{\text{HS},2}$ then denotes the L^2 -norm of $|f^{(d)}|_{\text{HS}}$. In detail:

Theorem 1.10. *Let $f: G \rightarrow \mathbb{R}$ be a C^d -smooth function such that $\int_G f d\mu = 0$ and $\int_G \partial_{i_1 \dots i_k} f d\mu = 0$ for all $k = 1, \dots, d-1$ and $1 \leq i_1, \dots, i_k \leq n$. Assume that*

$$\|f^{(d)}\|_{\text{HS},2} \leq 1 \quad \text{and} \quad \|f^{(d)}\|_{\text{Op},\infty} \leq 1.$$

Then, there exists some universal constant $c > 0$ such that

$$\int_G \exp \left\{ \frac{c}{\sigma^2} |f|^{2/d} \right\} d\mu \leq 2.$$

Here again, a possible choice is $c = 1/(8e)$. Note that, by partial integration, if μ is the standard Gaussian measure, the conditions $\int_G f d\mu = 0$ and $\int_G \partial_{i_1 \dots i_k} f d\mu = 0$ are satisfied if f is orthogonal to all polynomials of (total) degree at most $d-1$.

As in case of functions of independent random variables, it is possible to refine the tail estimates implied by Theorem 1.9:

Corollary 1.11. *Let $f: G \rightarrow \mathbb{R}$ be a \mathcal{C}^d -smooth function such that $\int_G f d\mu = 0$. For any $t \geq 0$, putting*

$$\eta_f(t) = \min \left(\frac{t^{2/d}}{\sigma^2 \|f^{(d)}\|_{\text{Op},\infty}^{2/d}}, \min_{k=1,\dots,d-1} \frac{t^{2/k}}{\sigma^2 \|f^{(k)}\|_{\text{Op},2}^{2/k}} \right),$$

we have

$$\mu(|f| \geq t) \leq e^2 \exp\{-\eta_f(t)/(de)^2\}.$$

Note that for $d = 2$ and functions $f(X) = \sum_{i,j} a_{ij} X_i X_j$, where X_1, \dots, X_n are independent with mean zero, this yields Hanson–Wright type inequalities.

Possible applications of Theorem 1.9 include functionals of the eigenvalues of random matrices. As in [G-S], we consider two situations. First, let $\{\xi_{jk}\}_{1 \leq j \leq k \leq N}$ be a family of independent random variables, and assume that the distributions of the ξ_{jk} 's all satisfy a (one-dimensional) logarithmic Sobolev inequality (1.18) with common constant σ^2 . Putting $\xi_{jk} = \xi_{kj}$ for $k < j$, consider a symmetric $N \times N$ random matrix $\Xi = (\xi_{jk}/\sqrt{N})_{1 \leq j, k \leq N}$ and denote by $\mu^{(N)} = \mu$ the joint distribution of its ordered eigenvalues $\lambda_1 \leq \dots \leq \lambda_N$ on \mathbb{R}^N (note that $\lambda_1 < \dots < \lambda_N$ a.s.).

Secondly, we consider 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

$$(1.24) \quad \mu(d\lambda) = \frac{1}{Z_N} e^{-\beta N \mathcal{H}(\lambda)} d\lambda, \quad \mathcal{H}(\lambda) = \frac{1}{2} \sum_{k=1}^N V(\lambda_k) - \frac{1}{N} \sum_{1 \leq k < l \leq N} \log(\lambda_l - \lambda_k)$$

for $\lambda = (\lambda_1, \dots, \lambda_N)$, $\lambda_1 < \dots < \lambda_N$. Here, $V: \mathbb{R} \rightarrow \mathbb{R}$ is a strictly convex \mathcal{C}^2 -smooth function, and Z_N is a normalization constant. 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 (1.24) as particle systems on the real line with Coulomb interactions.

In both cases, the probability measure μ satisfies a logarithmic Sobolev inequality with constant of order $1/N$ (see [G-S] for details). Throughout the rest of this section, we consider the probability space $(\mathbb{R}^N, \mathbb{B}^N, \mu)$, where μ is one of the two probability measures introduced above, supported on the set $\lambda_1 < \dots < \lambda_N$.

In [G-S], we studied concentration bounds for linear and quadratic eigenvalue statistics. Those results may be reproved (up to constants) using Theorem 1.10, and by Corollary 1.11 it is moreover possible to give slightly more accurate estimates for the tails. In the sequel, we will rather study a related problem, namely multilinear polynomials in the eigenvalues $\lambda_1, \dots, \lambda_N$. That is, we consider functionals of type

$$(1.25) \quad \sum_{i_1 \neq \dots \neq i_d} a_{i_1 \dots i_d} \lambda_{i_1} \cdots \lambda_{i_d}.$$

Here, $a_{i_1 \dots i_d}$ are real numbers such that for any permutation $\sigma \in S^d$, $a_{\sigma(i_1) \dots \sigma(i_d)} \equiv a_{i_1 \dots i_d}$, and $a_{i_1 \dots i_d} = 0$ whenever the indexes i_1, \dots, i_d are not pairwise different. This gives rise to a hypermatrix $A = (a_{i_1 \dots i_d}) \in \mathbb{R}^{n^d}$, whose Euclidean norm we denote by $\|A\|_{\text{HS}}$. Moreover, set $\|A\|_{\infty} = \max_{i_1 < \dots < i_d} |a_{i_1 \dots i_d}|$.

According to the framework sketched in Theorem 1.10, we shall not only center around the expected value of (1.25) but also around some “lower order” terms in order to arrive at centered derivatives of order up to $d - 1$. We work out details for $d = 1, \dots, 4$. To facilitate notation, let us introduce the following conventions: by $\mu[\cdot]$, we denote integration with respect to the measure μ . Moreover, set $\tilde{\lambda}_i = \lambda_i - \mu[\lambda_i]$. For any subset $\{i_1, \dots, i_d\} \subset \{1, \dots, N\}$, write $\tilde{\lambda}_{i_1 \dots i_d} = \tilde{\lambda}_{i_1} \cdots \tilde{\lambda}_{i_d}$. Now (similarly to Theorem 1.4 in [G-S-S]) define the functions

$$\begin{aligned} f_1(\lambda) &= \sum_{i=1}^N a_i \tilde{\lambda}_i, \\ f_2(\lambda) &= \sum_{i \neq j} a_{ij} (\tilde{\lambda}_{ij} - \mu[\tilde{\lambda}_{ij}]), \\ f_3(\lambda) &= \sum_{i \neq j \neq k} a_{ijk} (\tilde{\lambda}_{ijk} - \mu[\tilde{\lambda}_{ijk}] - 3\tilde{\lambda}_i \mu[\tilde{\lambda}_{jk}]), \\ f_4(\lambda) &= \sum_{i \neq j \neq k \neq l} a_{ijkl} (\tilde{\lambda}_{ijkl} - \mu[\tilde{\lambda}_{ijkl}] - 4\tilde{\lambda}_i \mu[\tilde{\lambda}_{jkl}] - 6\tilde{\lambda}_{ij} \mu[\tilde{\lambda}_{kl}] + 6\mu[\tilde{\lambda}_{ij}] \mu[\tilde{\lambda}_{kl}]). \end{aligned}$$

Applying Theorem 1.10 and recalling that the Sobolev constant of μ is of order $1/N$ immediately yields the following result.

Proposition 1.12. *Let μ be the joint distribution of the ordered eigenvalues of Ξ or the distribution defined in (1.24). For the functions f_d , $d = 1, \dots, 4$, defined above, with some constant $c > 0$, we have*

$$\int \exp \left\{ \frac{cN}{\|A\|_{\text{HS}}^{2/d}} |f_d|^{2/d} \right\} d\mu \leq 2$$

and moreover

$$\int \exp \left\{ \frac{c}{\|A\|_{\infty}^{2/d}} |f_d|^{2/d} \right\} d\mu \leq 2.$$

If μ is the eigenvalue distribution of Ξ , c depends on the logarithmic Sobolev constant σ^2 , and if μ is the β -ensemble distribution (1.24), c depends on β and the potential function V . In particular,

$$\mu(|f_d| \geq t) \leq 2 \exp \left\{ - \frac{cNt^{2/d}}{\|A\|_{\text{HS}}^{2/d}} \right\} \leq 2 \exp \left\{ - \frac{ct^{2/d}}{\|A\|_{\infty}^{2/d}} \right\}.$$

The bounds may be somewhat sharpened by applying Corollary 1.11. We omit details. In particular, Proposition 1.12 implies that if we “recenter”

$$\sum_{i_1 \neq \dots \neq i_d} \lambda_{i_1} \cdots \lambda_{i_d} \quad (d = 1, \dots, 4)$$

in such a way that all derivatives of order up to $d - 1$ are centered (cf. the definition of the functions f_d given above), we obtain exponential concentration bounds which yield fluctuations of order $\mathcal{O}_P(1)$. For $d = 2$, we thus get back a result shown in Proposition 1.12 from [G-S]. These bounds may be extended to higher orders $d \geq 5$. In some sense, this may be seen as an extension of the self-normalizing property of linear eigenvalue statistics for a special class of higher order polynomials.

1.4. Functions on the unit sphere. As a particular case, one may consider (real-valued) functions defined in some open neighbourhood G of the unit sphere

$$S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}, \quad n \geq 2,$$

which we equip with the uniform, or normalized Lebesgue measure σ_{n-1} . Since any \mathcal{C}^d -smooth function on S^{n-1} can be extended to a \mathcal{C}^d -smooth function on $\mathbb{R}^n \setminus \{0\}$, this means no loss of generality. We then restrict the usual (Euclidean) derivatives of f to S^{n-1} , which allows to use the definitions of the hyper-matrices (1.19) with operator norms (1.20), together with the L^p -norms $\|f^{(d)}\|_{\text{Op},p}$ in (1.21) taken with respect to σ_{n-1} . This yields the following analogue of Theorem 1.9.

Theorem 1.13. *Let f be a \mathcal{C}^d -smooth function on some open neighbourhood of S^{n-1} with $\int_{S^{n-1}} f d\sigma_{n-1} = 0$. Assume that*

$$\|f^{(k)}\|_{\text{Op},2} \leq n^{-(d-k)/2} \quad \forall k = 1, \dots, d-1$$

and $|f^{(d)}(\theta)|_{\text{Op}} \leq 1$ for all $\theta \in S^{n-1}$. Then, for some universal constant $c > 0$,

$$\int_{S^{n-1}} \exp\{(n-1)|f|^{2/d}/(8e)\} d\sigma_{n-1} \leq 2.$$

Moreover, an analogue of Theorem 1.10 also holds, which is particularly interesting for the class of p -homogeneous functions. Recall that a function f on $\mathbb{R}^n \setminus \{0\}$ is p -homogeneous for some $p \in \mathbb{R}$, if $f(\lambda x) = \lambda^p f(x)$ for all $x \neq 0$ and $\lambda > 0$.

Theorem 1.14. *Suppose that a \mathcal{C}^d -smooth function f on $\mathbb{R}^n \setminus \{0\}$ is p -homogeneous for some real number $p > d-3$ and is orthogonal in $L^2(\sigma_{n-1})$ to all polynomials of total degree at most $d-1$. Moreover, assume that*

$$(1.26) \quad \|f^{(d)}\|_{\text{HS},2} \leq 1 \quad \text{and} \quad \|f^{(d)}\|_{\text{Op},\infty} \leq 1.$$

Then, with some universal constant $c > 0$

$$\int_G \exp\left\{\frac{c}{\sigma^2}|f|^{2/d}\right\} d\mu \leq 2.$$

A possible choice is $c = 1/(8e)$. The same holds for $p \leq d-3$, if $n > d-p-1$.

Recall that the Hilbert space $L^2(S^{n-1})$ can be decomposed into a sum of orthogonal subspaces H_d , $d = 0, 1, 2, \dots$, consisting of all d -homogeneous harmonic polynomials (in fact, restrictions of such polynomials to the sphere). This fact is mirrored in the orthogonality assumptions from Theorem 1.14. If f is not a homogeneous function, the bounds from Theorem 1.14 remain valid assuming (1.26), but instead of orthogonality to polynomials of lower degree we have to require that f and all its partial derivatives of order up to $d-1$ are centered with respect to σ_{n-1} .

In Theorem 1.13, we use the usual (Euclidean) derivatives of functions defined in an open neighbourhood of the unit sphere. In applications, this is usually sufficient. There is also a notion of intrinsic (spherical) derivatives (cf. Section 5), and it is possible to obtain an analogue of Theorem 1.13 for these derivatives as well.

To fix some notation, denote by $\nabla_S f$ the spherical gradient of a differentiable function $f: S^{n-1} \rightarrow \mathbb{R}$ and write $D_i f = \langle \nabla_S f, e_i \rangle$, $i = 1, \dots, n$, for the spherical partial derivatives of f . Here, e_i denotes the i -th standard unit vector in \mathbb{R}^n . Higher order spherical partial derivatives are defined by iteration, e. g. $D_{ij} f = \langle \nabla_S \langle \nabla_S f, e_j \rangle, e_i \rangle$

for any $1 \leq i, j \leq n$. Note that in general, $D_{ij}f \neq D_{ji}f$. If $f \in \mathcal{C}^d(S^{n-1})$, we denote by $D^d f(\theta)$ the hyper-matrix of the spherical partial derivatives of order d , i. e.

$$(1.27) \quad (D^{(d)}f(\theta))_{i_1 \dots i_d} = D_{i_1 \dots i_d}f(\theta), \quad \theta \in S^{n-1}.$$

Similarly to (1.20), let $|D^{(d)}f(\theta)|_{\text{Op}}$ be the operator norm of $D^{(d)}f(\theta)$. Finally, write

$$(1.28) \quad \|D^{(d)}f\|_{\text{Op},p} = \left(\int_{S^{n-1}} |D^{(d)}f|_{\text{Op}}^p d\sigma_{n-1} \right)^{1/p}, \quad p \in (0, \infty].$$

We have the following ‘‘intrinsic’’ version of Theorem 1.13.

Theorem 1.15. *Let f be a \mathcal{C}^d -smooth function on S^{n-1} such that $\int_{S^{n-1}} f d\sigma_{n-1} = 0$. Assume that*

$$\|D^{(k)}f\|_{\text{Op},2} \leq n^{-(d-k)/2}, \quad k = 1, \dots, d-1,$$

and $|D^{(d)}f(\theta)|_{\text{Op}} \leq 1$ for all $\theta \in S^{n-1}$. Then

$$\int_{S^{n-1}} \exp\{(n-1)|f|^{2/d}/(8e)\} d\sigma_{n-1} \leq 2.$$

1.5. Outline. In Section 2, we give the proofs of the theorems and corollaries from Section 1.1. We briefly discuss the notion of difference operators. The main tool is a recursion inequality for the L^p -norms of the function f and the Hilbert–Schmidt norms of $|\mathfrak{h}^{(k)}f|$. In Section 3, Theorem 1.8 is proven. This includes a number of relations between the difference operators introduced in Definition 1.7. In Section 4, we prove Theorems 1.9 and 1.10 as well as Corollary 1.11 by adapting the main steps of the proof of Theorem 1.1. In Section 5, the proofs of Theorems 1.13, 1.14 and 1.15 are given; in particular, we introduce some facts about spherical calculus which allow us to proceed in a similar way as in case of functions on \mathbb{R}^n . Finally, in Section 6, we illustrate Theorem 1.13 on the example of polynomials and the problem of Edgeworth approximations for symmetric functions on the sphere. For additional applications we refer to [G-S].

2. FUNCTIONS OF INDEPENDENT RANDOM VARIABLES: PROOFS

Let $X = (X_1, \dots, X_n)$ be a vector of independent random variables on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. By a ‘‘difference operator’’ we mean an \mathbb{R}^n -valued functional Γ defined on $L^\infty(\mathbb{P})$ such that the following two conditions hold:

Conditions 2.1. (i) $\Gamma f(X) = (\Gamma_1 f(X), \dots, \Gamma_n f(X))$, where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ may be any Borel measurable function such that $f(X) \in L^\infty(\mathbb{P})$.
(ii) $|\Gamma_i(a f(X) + b)| = a |\Gamma_i f(X)|$ for all $a > 0$, $b \in \mathbb{R}$ and $i = 1, \dots, n$.

We also call Γ a gradient operator or simply gradient. We do not suppose Γ to satisfy any sort of ‘‘Leibniz rule’’. Clearly, the difference operator \mathfrak{h} from (1.1) and any of the difference operators introduced in Definition 1.7 satisfy Conditions 2.1.

For the proof of Theorem 1.1 we will need several lemmas. As before, let $T_i f = f(X_1, \dots, X_{i-1}, \bar{X}_i, X_{i+1}, \dots, X_n)$ with $\bar{X}_1, \dots, \bar{X}_n$ an independent copy of X . As a first step, the Hilbert–Schmidt norms of the derivatives of consecutive orders are related in the following way:

Lemma 2.2. *For any $d \geq 2$,*

$$(2.1) \quad |\mathfrak{h}|\mathfrak{h}^{(d-1)}f(X)|_{\text{HS}} \leq |\mathfrak{h}^{(d)}f(X)|_{\text{HS}}.$$

Proof. First let $d = 2$. Using $T_i|\mathfrak{h}f| = |T_i\mathfrak{h}f|$ and the triangle inequality, we have

$$(2.2) \quad (\mathfrak{h}_i|\mathfrak{h}f|)^2 = \frac{1}{4} \|\|\mathfrak{h}f| - |T_i\mathfrak{h}f|\|_{i,\infty}^2 \leq \frac{1}{4} \|\|\mathfrak{h}f - T_i\mathfrak{h}f\|_{i,\infty}^2 = \frac{1}{4} \|\|\mathfrak{h}f - T_i\mathfrak{h}f|^2\|_{i,\infty}.$$

Here, $\mathfrak{h}f - T_i\mathfrak{h}f$ is defined componentwise. Since $T_i\mathfrak{h}_if = \mathfrak{h}_if$, we obtain

$$(2.3) \quad \begin{aligned} |\mathfrak{h}f - T_i\mathfrak{h}f|^2 &= \sum_{j \neq i} \frac{1}{4} (\|f - T_jf\|_{j,\infty} - \|T_if - T_{ij}f\|_{j,\infty})^2 \\ &\leq \sum_{j \neq i} \frac{1}{4} \|f - T_jf - T_if + T_{ij}f\|_{j,\infty}^2, \end{aligned}$$

where the last inequality follows from the reverse triangular inequality again (for the pseudo-norm $\|\cdot\|_{j,\infty}$). Combining (2.2) and (2.3) yields

$$\begin{aligned} (\mathfrak{h}_i|\mathfrak{h}f|)^2 &\leq \left\| \frac{1}{16} \sum_{j \neq i} \|f - T_jf - T_if + T_{ij}f\|_{j,\infty}^2 \right\|_{i,\infty} \\ &\leq \frac{1}{16} \sum_{j \neq i} \|f - T_jf - T_if + T_{ij}f\|_{i,j,\infty}^2. \end{aligned}$$

Summing over $i = 1, \dots, n$ we arrive at the result in the case $d = 2$.

For $d \geq 3$, note that $T_i\mathfrak{h}_{i_1 \dots i_{d-1}}f = \mathfrak{h}_{i_1 \dots i_{d-1}}f$ whenever $i \in \{i_1, \dots, i_{d-1}\}$. The claim then follows in the same way as above. \square

Corresponding results in the setting of differentiable functions (see Lemma 4.1) suggest to replace the Hilbert–Schmidt norms in Lemma 2.2 by operator type norms (1.20). In Boucheron, Bousquet, Lugosi and Massart [B-B-L-M], Theorem 14, iterations of (2.6) are sketched to study applications for Rademacher chaos type functions. Unfortunately, working out the arguments in the proof of Theorem 14 we seemed to need Hilbert–Schmidt instead of operator norms. Already in second order statistics of Rademacher variables like $\sum_{i=1}^n X_i X_{i+1}$ (setting $X_{n+1} = X_1$), an analogue of (2.1) for operator norms cannot be true. Similar remarks hold for any of the difference operators introduced in Definition 1.7 (cf. Remark 3.1).

Our results will follow from certain moment inequalities for functions of independent random variables. In [B-B-L-M], cf. Theorem 2, the following moment bounds are shown

$$\|(f - \mathbb{E}f)_+\|_p \leq \sqrt{2\kappa p} \|V^+(f)\|_p, \quad \|(f - \mathbb{E}f)_-\|_p \leq \sqrt{2\kappa p} \|V^-(f)\|_p$$

in terms of the conditional expectations

$$V^+(f) = \mathbb{E} \left(\sum_{i=1}^n (f - T_if)_+^2 | X \right) \quad V^-(f) = \mathbb{E} \left(\sum_{i=1}^n (f - T_if)_-^2 | X \right),$$

where $\kappa = \frac{\sqrt{e}}{2(\sqrt{e}-1)} < 1.271$. Note that, in our notations according to Definition 1.7,

$$V^+(f) = 2|\mathfrak{d}^+f|^2 \quad \text{and} \quad V^-(f) = 2|\mathfrak{d}^-f|^2.$$

For iterating these inequalities however, we had to bypass the problem that $\mathfrak{d}_{ii} = \mathfrak{d}_i$ respectively $\mathfrak{d}_{ii}^+ = \mathfrak{d}_i^+$ (up to constant), which would introduce additional lower

order differences on the right-hand side of (2.1). This motivated us to introduce the following related quantities adapted to the framework of L^∞ -bounds. For $i = 1, \dots, n$, introduce

$$(2.4) \quad \mathfrak{h}_i^+ f(X) = \frac{1}{2} \|(f(X) - T_i f(X))_+\|_{i, \infty}, \quad \mathfrak{h}^+ f = (\mathfrak{h}_1^+ f, \dots, \mathfrak{h}_n^+ f),$$

$$(2.5) \quad \mathfrak{h}_i^- f(X) = \frac{1}{2} \|(f(X) - T_i f(X))_-\|_{i, \infty}, \quad \mathfrak{h}^- f = (\mathfrak{h}_1^- f, \dots, \mathfrak{h}_n^- f),$$

which are clearly difference operators in the sense of Conditions 2.1. Using the relations $V^+(f) \leq 4 |\mathfrak{h}^+ f|^2$ and $V^-(f) \leq 4 |\mathfrak{h}^- f|^2$, we get from the [B-B-L-M]-result the following somewhat weaker bounds in terms of the L^p -norms as in (1.4).

Theorem 2.3. *With the same constant $\kappa = \frac{\sqrt{e}}{2(\sqrt{e}-1)}$, for any real $p \geq 2$,*

$$(2.6) \quad \|(f - \mathbb{E}f)_+\|_p \leq \sqrt{8 \kappa p} \|\mathfrak{h}^+ f\|_p,$$

$$(2.7) \quad \|(f - \mathbb{E}f)_-\|_p \leq \sqrt{8 \kappa p} \|\mathfrak{h}^- f\|_p.$$

For the reader's convenience, let us give a self-contained proof of Theorem 2.3. It is sufficient to derive (2.6), since (2.7) follows from (2.6) by considering $-f$. The key step are the following two lemmas.

Lemma 2.4. *Assume $\mathbb{E}f = 0$. Then,*

$$(2.8) \quad \|f\|_2 \leq \sqrt{2} \|\mathfrak{h}f\|_2,$$

$$(2.9) \quad \|f_+\|_2 \leq 2 \|\mathfrak{h}^+ f\|_2.$$

Proof. By the Efron–Stein inequality (1.10), $\mathbb{E}f^2 \leq \mathbb{E}|\mathfrak{d}f|^2$ and $\mathbb{E}f^2 \leq 2 \mathbb{E}|\mathfrak{d}^+ f|^2$, while $|\mathfrak{d}^+ f|^2 \leq 2 |\mathfrak{h}^+ f|^2$ and $|\mathfrak{d}f|^2 \leq 2 |\mathfrak{h}f|^2$ (cf. Remark 3.1 (v)). \square

The next lemma provides a moment recursion similarly to [B-B-L-M], Lemma 3.

Lemma 2.5. *For any real $p \geq 2$,*

$$(2.10) \quad \|f_+\|_p^p \leq \|f_+\|_{p-1}^p + 4(p-1) \|\mathfrak{h}^+ f\|_p^2 \|f_+\|_p^{p-2}.$$

Proof. First assume $n = 1$, i. e. $f = f(X)$ for a random variable X and $Tf = f(\bar{X})$, where \bar{X} is an independent copy of X . Using the notation $f_+^{p-1} \equiv (f_+)^{p-1}$, we have

$$\frac{1}{2} \mathbb{E}[(f_+^{p-1} - T f_+^{p-1})(f_+ - T f_+)] - \|f_+\|_p^p = -\|f_+\|_{p-1}^{p-1} \|f_+\|_1 \geq -\|f_+\|_{p-1}^p.$$

Thus, by symmetry in X and \bar{X} , and since $(f - Tf)_+ \leq 2 \mathfrak{h}^+ f$,

$$\begin{aligned} \|f_+\|_p^p &\leq \|f_+\|_{p-1}^p + \mathbb{E}[(f_+^{p-1} - T f_+^{p-1})_+ (f_+ - T f_+)] \\ &\leq \|f_+\|_{p-1}^p + (p-1) \mathbb{E}[(f - Tf)_+^2 f_+^{p-2}]. \end{aligned}$$

Using Hölder's inequality, the last expectation may be bounded by

$$(2.11) \quad 4 \mathbb{E} [|\mathfrak{h}^+ f|^2 f_+^{p-2}] \leq 4 \| |\mathfrak{h}^+ f|^2 \|_{p/2} \|f_+\|_p^{p-2} = 4 \|\mathfrak{h}^+ f\|_p^2 \|f_+\|_p^{p-2}.$$

This completes the proof in case $n = 1$.

For $n \geq 1$, we use a tensorization argument: For any $g \in L^q$, $q \in (1, 2]$,

$$(2.12) \quad \mathbb{E} |g|^q - (\mathbb{E} |g|)^q \leq \mathbb{E} \sum_{i=1}^n \left(\mathbb{E}_i |g|^q - (\mathbb{E}_i |g|)^q \right),$$

where \mathbb{E}_i denotes expectation with respect to X_i . Applying this inequality to $g = f_+^{p-1}$ with $q = p/(p-1)$, similarly to the case of $n = 1$ we obtain

$$\begin{aligned} \|f_+\|_p^p - \|f_+\|_{p-1}^p &\leq \mathbb{E} \sum_{i=1}^n \left(\mathbb{E}_i f_+^p - (\mathbb{E}_i f_+^{p-1})^{p/(p-1)} \right) \\ &\leq (p-1) \sum_{i=1}^n \mathbb{E} \mathbb{E}_i \bar{\mathbb{E}}_i [(f - T_i f)_+^2 f_+^{p-2}] \\ &\leq 4(p-1) \sum_{i=1}^n \mathbb{E} [|\mathfrak{h}_i^+ f|^2 f_+^{p-2}] = 4(p-1) \mathbb{E} [|\mathfrak{h}^+ f|^2 f_+^{p-2}]. \end{aligned}$$

As in (2.11), the last expectation is bounded by $\|\mathfrak{h}^+ f\|_p^2 \|f_+\|_p^{p-2}$ using Hölder's inequality, which gives the desired result.

It remains to prove (2.12). Let us mention that the tensorization of functionals $L(g) = \mathbb{E} \Psi(g) - \Psi(\mathbb{E} g)$ was proposed in the mid 1990's by Bobkov, as explained in [L1], Proposition 4.1. This property is actually equivalent to the convexity of L in g , and can be explicitly expressed in terms of R (convexity of Ψ and $-1/\Psi''$; see also [L-O]). For completeness of exposition let us include here a direct argument for the power functions $\Psi(x) = x^q$. By induction, it suffices to consider $n = 2$; we use the representation

$$(2.13) \quad L(|g|) = \sup_{h \in L^q} \left\{ q \mathbb{E} [|g|(|h|^{q-1} - (\mathbb{E} |h|)^{q-1})] - (q-1) (\mathbb{E} |h|^q - (\mathbb{E} |h|)^q) \right\}.$$

Indeed, by the arithmetic-geometric inequality, $\frac{1}{q} \mathbb{E} |g|^q + \frac{q-1}{q} \mathbb{E} |h|^q \leq \mathbb{E} |g| |h|^{q-1}$, which we rewrite as

$$\mathbb{E} |g|^q \leq q \mathbb{E} |g| |h|^{q-1} - (q-1) \mathbb{E} |h|^q.$$

We may assume $\mathbb{E} |g| = 1$; therefore, subtracting $(\mathbb{E} |g|)^q = 1$ on both sides,

$$L(|g|) \leq q \mathbb{E} [|g|(|h|^{q-1} - (\mathbb{E} |h|)^{q-1})] - (q-1) (\mathbb{E} |h|^q - (\mathbb{E} |h|)^q) + R(\mathbb{E} |h|)$$

with $R(x) = qx^{q-1} - (q-1)x^q - 1$ for $x \geq 0$. Since $R(x) \leq 0$, while equality holds if $h = g$, we arrive at (2.13). By Fubini's theorem and applying (2.13), we now get

$$\begin{aligned} \mathbb{E}_2 [(\mathbb{E}_1 |g|)^q - (\mathbb{E} |g|)^q] &= \mathbb{E}_2 (\mathbb{E}_1 |g|)^q - (\mathbb{E}_2 \mathbb{E}_1 |g|)^q \\ &= \sup_{h(X_2) \in L^q} \left\{ q \mathbb{E}_2 [(\mathbb{E}_1 |g|)(|h|^{q-1} - (\mathbb{E}_2 |h|)^{q-1})] - (q-1) (\mathbb{E}_2 |h|^q - (\mathbb{E}_2 |h|)^q) \right\} \\ &= \sup_{h(X_2) \in L^q} \left\{ \mathbb{E}_1 [q (\mathbb{E}_2 |g|)(|h|^{q-1} - (\mathbb{E}_2 |h|)^{q-1}) - (q-1) (\mathbb{E}_2 |h|^q - (\mathbb{E}_2 |h|)^q)] \right\} \\ &\leq \mathbb{E}_1 \left[\sup_{h(X_2) \in L^q} \left\{ q \mathbb{E}_2 [|g|(|h|^{q-1} - (\mathbb{E}_2 |h|)^{q-1})] - (q-1) (\mathbb{E}_2 |h|^q - (\mathbb{E}_2 |h|)^q) \right\} \right] \\ &= \mathbb{E}_1 [\mathbb{E}_2 |g|^q - (\mathbb{E}_2 |g|)^q]. \end{aligned}$$

As a consequence, by Fubini's theorem again,

$$\begin{aligned} \mathbb{E} |g|^q - (\mathbb{E} |g|)^q &= \mathbb{E}_2 [\mathbb{E}_1 |g|^q - (\mathbb{E}_1 |g|)^q] + \mathbb{E}_2 [(\mathbb{E}_1 |g|)^q - (\mathbb{E} |g|)^q] \\ &\leq \mathbb{E}_2 [\mathbb{E}_1 |g|^q - (\mathbb{E}_1 |g|)^q] + \mathbb{E}_1 [\mathbb{E}_2 |g|^q - (\mathbb{E}_2 |g|)^q]. \end{aligned}$$

□

Following the arguments in [B-B-L-M], we may now prove Theorem 2.3:

Proof of Theorem 2.3. It suffices to prove (2.6) assuming $\mathbb{E}f = 0$. To this end, by induction on k , we show that for all $k \in \mathbb{N}$ and all $p \in (k, k + 1]$,

$$(2.14) \quad \|f_+\|_p \leq \sqrt{8\kappa_p p} \|\mathfrak{h}^+ f\|_{p\sqrt{2}} \quad \text{with} \quad \kappa_p = \frac{1}{2} \left(1 - \left(1 - \frac{1}{p}\right)^{p/2}\right)^{-1}.$$

These constants are strictly increasing in p , $\kappa_1 = 1/2$ and $\lim_{p \rightarrow \infty} \kappa_p = \kappa = \frac{\sqrt{e}}{2(\sqrt{e}-1)}$.

For $k = 1$ and $p \in (1, 2]$, by (2.9) and the fact that $\kappa_p p \geq 1/2$, we have

$$\|f_+\|_p \leq \|f_+\|_2 \leq 2 \|\mathfrak{h}^+ f\|_2 \leq \sqrt{8\kappa_p p} \|\mathfrak{h}^+ f\|_2.$$

To make an induction step, fix an integer $k > 1$ and assume that (2.14) holds for all real $p \in [1, k]$. Now, consider the values $p \in (k, k + 1]$. Set

$$x_p = \|f_+\|_p^p 8^{-p/2} \kappa_p^{-p/2} p^{-p/2} \|\mathfrak{h}^+ f\|_{p\sqrt{2}}^{-p},$$

so that it suffices to prove that $x_p \leq 1$. In terms of x_p , (2.10) implies that

$$\begin{aligned} & x_p 8^{p/2} \kappa_p^{p/2} p^{p/2} \|\mathfrak{h}^+ f\|_p^p \\ & \leq x_{p-1}^{p/(p-1)} 8^{p/2} \kappa_{p-1}^{p/2} (p-1)^{p/2} \|\mathfrak{h}^+ f\|_{p-1}^p \\ & \quad + 4(p-1) \|\mathfrak{h}^+ f\|_p^2 x_p^{1-2/p} 8^{p/2-1} \kappa_p^{p/2-1} p^{p/2-1} \|\mathfrak{h}^+ f\|_p^{p-2} \\ & \leq x_{p-1}^{p/(p-1)} 8^{p/2} \kappa_p^{p/2} (p-1)^{p/2} \|\mathfrak{h}^+ f\|_p^p + \frac{1}{2} x_p^{1-2/p} 8^{p/2} \kappa_p^{p/2-1} p^{p/2} \|\mathfrak{h}^+ f\|_p^p. \end{aligned}$$

Here we have used the fact that $\kappa_{p-1} \leq \kappa_p$. Simplifying and using that by induction, $x_{p-1} \leq 1$, it follows that

$$x_p \leq x_{p-1}^{p/(p-1)} \left(1 - \frac{1}{p}\right)^{p/2} + \frac{1}{2\kappa_p} x_p^{1-2/p} \leq \left(1 - \frac{1}{p}\right)^{p/2} + \frac{1}{2\kappa_p} x_p^{1-2/p}.$$

Now note that the function

$$u_p(x) = \left(1 - \frac{1}{p}\right)^{p/2} + \frac{1}{2\kappa_p} x^{1-2/p} - x$$

is concave on \mathbb{R}_+ and positive at $x = 0$. Since $u_p(1) = 0$ and $u_p(x_p) \geq 0$, we may conclude that $x_p \leq 1$. □

Corollary 2.6. *Given $f = f(X_1, \dots, X_n)$ in $L^\infty(\mathbb{P})$, for all $p \geq 2$,*

$$(2.15) \quad \|f\|_p \leq \|f\|_2 + \sqrt{32\kappa p} \|\mathfrak{h}f\|_p.$$

If additionally $\mathbb{E}f = 0$,

$$(2.16) \quad \|f\|_p \leq \sqrt{32\kappa p} \|\mathfrak{h}f\|_p.$$

Proof. By Theorem 2.3,

$$\begin{aligned} \|f - \mathbb{E}f\|_p & \leq \|(f - \mathbb{E}f)_+\|_p + \|(f - \mathbb{E}f)_-\|_p \\ & \leq \sqrt{8\kappa p} \|\mathfrak{h}^+ f\|_p + \sqrt{8\kappa p} \|\mathfrak{h}^- f\|_p \leq 2\sqrt{8\kappa p} \|\mathfrak{h}f\|_p, \end{aligned}$$

which proves (2.16). Moreover, by the triangle inequality,

$$\|f - \mathbb{E}f\|_p \geq \|f\|_p - |\mathbb{E}f| \geq \|f\|_p - \|f\|_2,$$

so that we obtain (2.15). \square

We shall now prove Theorem 1.1. Recall that if the relation of the form

$$(2.17) \quad \|f\|_k \leq \gamma k \quad (k \in \mathbb{N})$$

holds true with some constant $\gamma > 0$, then f has sub-exponential tails, i. e. $\mathbb{E}e^{cf|} \leq 2$ for some constant $c = c(\gamma) > 0$, e. g. $c = \frac{1}{2\gamma e}$. Indeed, using $k! \geq (\frac{k}{e})^k$, we have

$$\mathbb{E} \exp(c|f|) = 1 + \sum_{k=1}^{\infty} c^k \frac{\mathbb{E}|f|^k}{k!} \leq 1 + \sum_{k=1}^{\infty} (c\gamma)^k \frac{k^k}{k!} \leq 1 + \sum_{k=1}^{\infty} (c\gamma e)^k = 2.$$

Proof of Theorem 1.1. Put $A = \sqrt{32\kappa p}$. Using (2.15) with f replaced by $|\mathfrak{h}^{(k-1)}f|_{\text{HS}}$ for $k = 1, \dots, d$, and applying Lemma 2.2, we get

$$\begin{aligned} \|\mathfrak{h}^{(k-1)}f\|_{\text{HS},p} &\leq \|\mathfrak{h}^{(k-1)}f\|_{\text{HS},2} + A \|\mathfrak{h}|\mathfrak{h}^{(k-1)}f|_{\text{HS}}\|_p \\ &\leq \|\mathfrak{h}^{(k-1)}f\|_{\text{HS},2} + A \|\mathfrak{h}^{(k)}f\|_{\text{HS},p}. \end{aligned}$$

Consequently, using (2.16) and then iterating (2.15), we arrive at

$$(2.18) \quad \|f\|_p \leq \sum_{k=1}^{d-1} A^k \|\mathfrak{h}^{(k)}f\|_{\text{HS},2} + A^d \|\mathfrak{h}^{(d)}f\|_{\text{HS},p}.$$

Now, since $\|\mathfrak{h}^{(k)}f\|_{\text{HS},2} \leq 1$ for $k \leq d-1$ and $\|\mathfrak{h}^{(d)}f\|_{\text{HS},\infty} \leq 1$ by assumption, we get

$$\|f\|_p \leq \sum_{k=1}^d A^k = \frac{A^{d+1} - 1}{A - 1} - 1 \leq \left(\frac{A}{A-1}\right)^d$$

for all $p \geq 2$, where $A/(A-1) \leq 1.12$. We now arrive at $1.12A \leq \sqrt{Cp}$, where the best constant corresponds to $p = 2$, and then we find that $C < 52$. Hence, we obtain the bound

$$(2.19) \quad \|f\|_p \leq (52p)^{d/2}, \quad p \geq 2.$$

As for $0 < p < 2$, one may find $\|f\|_p \leq \|f\|_2 \leq (104)^{d/2}$, and thus, for all $k \geq 1$,

$$\| |f|^{2/d} \|_k = \|f\|_{2k/d}^{2/d} \leq \gamma k,$$

as in (2.17), with constant $\gamma = 104$. \square

Proof of Proposition 1.2. First note that since X_i 's are centered, we have $\alpha_0 = 0$, and the Hoeffding decomposition of f is given by the polynomials

$$h_{i_1 \dots i_d}(X_{i_1}, \dots, X_{i_d}) = \alpha_{i_1 \dots i_d} X_{i_1} \cdots X_{i_d} \quad (i_1 < \dots < i_d, \quad d = 1, \dots, n).$$

It is now easily seen that for any $1 \leq k \leq d$ and $1 \leq j_1 \neq \dots \neq j_k \leq n$,

$$\begin{aligned} \mathfrak{h}_{j_1 \dots j_k} f(X) &= \mathfrak{h}_{X_{j_1}} \cdots \mathfrak{h}_{X_{j_k}} \left| \sum_{\substack{i_1 < \dots < i_d \\ \ni j_1, \dots, j_k}} \alpha_{i_1 \dots i_d} \prod_{\substack{\nu \in \{i_1, \dots, i_d\} \\ \setminus \{j_1, \dots, j_k\}}} X_\nu \right. \\ &\quad \left. + \sum_{\substack{i_1 < \dots < i_{d+1} \\ \ni j_1, \dots, j_k}} \alpha_{i_1 \dots i_{d+1}} \prod_{\substack{\nu \in \{i_1, \dots, i_{d+1}\} \\ \setminus \{j_1, \dots, j_k\}}} X_\nu + \dots \right|. \end{aligned}$$

Here, $\mathfrak{h}X_{j_\nu}$ is understood according to (1.1) as the difference $\mathfrak{h}g$ for the function $g(X_{j_\nu}) = X_{j_\nu}$ (i. e. in dimension $n = 1$). Hence, for any $k = 0, \dots, d$,

$$\begin{aligned} \|\mathfrak{h}^{(k)}f(X)\|_{\text{HS},2}^2 &= \sum_{i_1 < \dots < i_d} \alpha_{i_1 \dots i_d}^2 \sum_{\substack{j_1 \neq \dots \neq j_k \\ \in \{i_1, \dots, i_d\}}} (\mathfrak{h}X_{j_1})^2 \cdots (\mathfrak{h}X_{j_k})^2 \\ &+ \sum_{i_1 < \dots < i_{d+1}} \alpha_{i_1 \dots i_{d+1}}^2 \sum_{\substack{j_1 \neq \dots \neq j_k \\ \in \{i_1, \dots, i_{d+1}\}}} (\mathfrak{h}X_{j_1})^2 \cdots (\mathfrak{h}X_{j_k})^2 + \dots \end{aligned}$$

As a consequence, since $\mathfrak{h}X_i \geq 1$ for all $i = 1, \dots, n$,

$$(2.20) \quad \|f\|_{\text{HS},2} \leq \|\mathfrak{h}^{(1)}f(X)\|_{\text{HS},2} \leq \dots \leq \|\mathfrak{h}^{(d)}f(X)\|_{\text{HS},2} \leq \|\mathfrak{h}^{(d)}f(X)\|_{\text{HS},\infty}.$$

□

Corollary 1.3 can now be obtained along the lines of the proof of Theorem 1.1, when the random variables X_1, \dots, X_n are Rademacher variables.

Proof of Corollary 1.3. Without loss of generality we assume h to be symmetric (i. e. invariant under permutations). Hence f can be rewritten as

$$f(X_1, \dots, X_n) = \frac{1}{\binom{n}{d}} \sum_{i_1 < \dots < i_d} h(X_{i_1}, \dots, X_{i_d}).$$

Introduce a set of independent Rademacher variables $\varepsilon_1, \dots, \varepsilon_n$ which are independent of the random variables X_1, \dots, X_n and consider

$$f^\varepsilon(X, \varepsilon) = \frac{1}{\binom{n}{d}} \sum_{i_1 < \dots < i_d} \varepsilon_{i_1} \cdots \varepsilon_{i_d} h(X_{i_1}, \dots, X_{i_d}).$$

Denote by $\mathfrak{h}_i^\varepsilon$, $\mathfrak{h}^{\varepsilon(d)}$ and by similar expressions differences of f^ε with respect to the Rademacher variables ε_i conditionally on X .

Note that conditionally on X , $f^\varepsilon(X, \varepsilon)$ has Fourier–Walsh expansion consisting of the d -th order term only. Hence, we may use (2.20) with $\|\mathfrak{h}^{(d)}f(X)\|_{\text{HS},\infty}$ replaced by $\|\mathfrak{h}^{(d)}f(X)\|_{\text{HS},p}$ in (2.18). Arguing as in (2.19), conditionally on X we get

$$\mathbb{E}_\varepsilon |f^\varepsilon(X, \varepsilon)|^p \leq (52p)^{pd/2} \mathbb{E}_\varepsilon |\mathfrak{h}^{\varepsilon(d)}f^\varepsilon(X, \varepsilon)|_{\text{HS}}^p$$

for all $p \geq 2$. Hence, taking expectations with respect to X on both sides, we have

$$(2.21) \quad \mathbb{E} |f^\varepsilon(X, \varepsilon)|^p \leq (52p)^{pd/2} \mathbb{E} |\mathfrak{h}^{\varepsilon(d)}f^\varepsilon(X, \varepsilon)|_{\text{HS}}^p.$$

It follows from a result by de la Peña and Giné [D-G], Theorem 3.5.3 (also see Joly and Lugosi [J-L], Theorem 8) that

$$(2.22) \quad \mathbb{E} |f(X)|^p \leq \tilde{c}^p \mathbb{E} |f^\varepsilon(X, \varepsilon)|^p$$

with some constant \tilde{c} depending on d only. Moreover, for any $i_1 \neq \dots \neq i_d$, it is not hard to see that

$$\mathfrak{h}_{i_1 \dots i_d}^\varepsilon f^\varepsilon(X, \varepsilon) = |h(X_{i_1}, \dots, X_{i_d})|$$

(cf. the proof of Proposition 1.2 and note that $\mathfrak{h}\varepsilon_i = 1$). Consequently,

$$(2.23) \quad |\mathfrak{h}^{\varepsilon(d)}f^\varepsilon(X, \varepsilon)|_{\text{HS}} \leq C_d(f),$$

where

$$C_d = \frac{1}{\binom{n}{d}} \left(\sum_{i_1 \neq \dots \neq i_d} \|h(X_1, \dots, X_d)\|_\infty^2 \right)^{1/2}.$$

Applying (2.22) and (2.23) on (2.21) and taking p -th roots, we arrive at

$$\|f\|_p \leq \tilde{c} (52p)^{d/2} C_d(f).$$

From here on, the proof is similar to the proof of Theorem 1.1 if we normalize f such that $C_d(f) \leq 1$. However, it follows from the assumptions on h that $C_d(f) \leq \hat{c}n^{-d/2}$ for some numerical constant \hat{c} depending on d and M only. Hence we arrive at the normalization $\hat{c}^{-1}n^{d/2}f$ which yields Corollary 1.3. \square

Proof of Corollary 1.4. First note that, by Chebychev's inequality, for any $p \geq 1$

$$(2.24) \quad \mathbb{P}(|f| \geq e \|f\|_p) \leq e^{-p}.$$

Moreover, if $p \geq 2$, it follows from (2.18) that

$$e \|f\|_p \leq e \left(\sum_{k=1}^{d-1} (41p)^{k/2} \|\mathfrak{h}^{(k)} f\|_{\text{HS},2} + (41p)^{d/2} \|\mathfrak{h}^{(d)} f\|_{\text{HS},\infty} \right).$$

Here we have used that $32\kappa < 41$. Assuming $\eta_f(t) \geq 2 \cdot 41$, we therefore arrive that

$$e \|f\|_{\eta_f(t)/41} \leq e \left(\sum_{k=1}^{d-1} t + t \right) = (de)t.$$

Hence, applying (2.24) to $p = \eta_f(t)/41$ (if $p \geq 2$) yields

$$\mu(|f| \geq (de)t) \leq \mu(|f| \geq e \|f\|_{\eta_f(t)/41}) \leq \exp\{-\eta_f(t)/41\}.$$

Using a trivial estimate in case of $p < 2$, we also obtain that

$$\mu(|f| \geq (de)t) \leq e^2 \exp\{-\eta_f(t)/41\}.$$

The proof is now easily completed by rescaling f and using $\eta_{(de)f}(t) \geq \eta_f(t)/(de)^2$. \square

Proof of Theorem 1.5. First note that

$$\mathfrak{h}_i \sup_{f \in \mathfrak{F}} |f(X)| \leq \sup_{f \in \mathfrak{F}} \mathfrak{h}_i f(X)$$

by the reverse triangular inequality, and therefore (writing $\mathfrak{h}^*(\mathfrak{F}) \equiv \mathfrak{h}^{*(1)}(\mathfrak{F})$)

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

In a similar way, we may also prove an analogue of (2.1), i. e.

$$|\mathfrak{h}|\mathfrak{h}^{*(d-1)}(\mathfrak{F})|_{\text{HS}} \leq |\mathfrak{h}^{*(d)}(\mathfrak{F})|_{\text{HS}}.$$

To see this, note that $\sup_{f \in \mathfrak{F}} \|\cdot\|_{j,\infty}$ is a pseudo-norm. In view of these elementary facts, the proof of Theorem 1.5 is now similar to the proof of Theorem 1.1. \square

Proof of Theorem 1.6. The proof is obtained by calculating the differences of first and second order. To start, note that for any $\nu = 1, \dots, n$,

$$(\mathfrak{h}_\nu S_f(X))^2 = \frac{1}{4} \left\| \sum_{i \geq \nu} \left(f \left(\sum_{j=1}^i X_j \right) - f \left(\sum_{j=1}^i T_\nu X_j \right) \right) \right\|_\infty^2 \leq \|f\|_\infty^2 (n - \nu + 1)^2,$$

and consequently

$$|\mathfrak{h}S_f(X)|^2 \leq \|f\|_\infty^2 \sum_{\nu=1}^n (n-\nu+1)^2 = \frac{1}{6}n(n+1)(2n+1)\|f\|_\infty^2 \leq Cn^3\|f\|_\infty^2$$

for some constant $C > 0$. Moreover, for any $\nu \neq \mu$, $(\mathfrak{h}_{\nu\mu}S_f(X))^2$ is given by

$$\frac{1}{16} \left\| \sum_{i \geq \nu \vee \mu} \left(f\left(\sum_{j=1}^i X_j\right) - f\left(\sum_{j=1}^i T_\nu X_j\right) - f\left(\sum_{j=1}^i T_\mu X_j\right) + f\left(\sum_{j=1}^i T_{\nu\mu} X_j\right) \right) \right\|^2.$$

By similar arguments as above, this expression does not exceed $\|f\|_\infty^2(n-(\nu \vee \mu)+1)^2$, and therefore

$$\|\mathfrak{h}^{(2)}S_f(X)\|_{\text{HS}}^2 = \sum_{\nu \neq \mu} (\mathfrak{h}_{\nu\mu}S_f(X))^2 \leq Cn^4\|f\|_\infty^2.$$

Combining these arguments and applying Corollary 1.4 completes the proof. \square

3. HIGHER ORDER EFRON–STEIN INEQUALITY: PROOFS

Let us first collect some elementary facts about the difference operators introduced in Section 1. As before, assume that $X = (X_1, \dots, X_n)$ has independent components.

Remark 3.1. *Let $i = 1, \dots, n$.*

(i) *If $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ has independent Rademacher components, then $\mathfrak{D}_i f(\varepsilon) = \frac{1}{2}(f(\varepsilon) - f(\sigma_i \varepsilon))$, where $\sigma_i \varepsilon = (\varepsilon_1, \dots, -\varepsilon_i, \dots, \varepsilon_n)$. Moreover,*

$$\mathfrak{h}_i f(\varepsilon) = \mathfrak{v}_i f(\varepsilon) = \mathfrak{d}_i f(\varepsilon) = |\mathfrak{D}_i f(\varepsilon)|.$$

(ii) *If $f(X) \in L^2(\mathbb{P})$, then*

$$(\mathfrak{v}_i f(X))^2 = \mathbb{E}_i(\mathfrak{D}_i f(X))^2 \quad \text{and} \quad (\mathfrak{v}_i f(X))^2 = \mathbb{E}_i(\mathfrak{d}_i f(X))^2.$$

In particular, we immediately obtain (1.11), where the identities involving $\mathfrak{d}^\pm f$ follow from symmetry and Fubini's theorem.

(iii) *Let $f(X) \in L^2(\mathbb{P})$. Then, by independence, we can rewrite $\mathfrak{d}_i f(X)$ as*

$$\begin{aligned} \mathfrak{d}_i f(X) &= \left(\frac{1}{2} \left((f(X) - \mathbb{E}_i f(X))^2 + \mathbb{E}_i (f(X) - \mathbb{E}_i f(X))^2 \right) \right)^{1/2} \\ (3.1) \quad &= \left(\frac{1}{2} \left((\mathfrak{D}_i f(X))^2 + \mathbb{E}_i (\mathfrak{D}_i f(X))^2 \right) \right)^{1/2}. \end{aligned}$$

(iv) *By induction over n , f is bounded if and only if $|\mathfrak{D}f|$ is bounded. Using (3.1), the same holds for $|\mathfrak{d}f|$ and $|\mathfrak{h}f|$ instead of $|\mathfrak{D}f|$.*

(v) *We have $|\mathfrak{d}^+ f| \leq |\mathfrak{d}f|$, $|\mathfrak{d}^- f| \leq |\mathfrak{d}f|$, $|\mathfrak{d}^+ f| \leq \sqrt{2}|\mathfrak{h}^+ f|$, $|\mathfrak{d}^- f| \leq \sqrt{2}|\mathfrak{h}^- f|$ and $|\mathfrak{d}f| \leq \sqrt{2}|\mathfrak{h}f|$.*

The difference operator \mathfrak{D} is closely related to the Hoeffding decomposition (1.7). Indeed, the representation (1.7) follows by tensorizing the identity $\mathbb{E}_i + \mathfrak{D}_i = Id$, so

$$(3.2) \quad h_{i_1 \dots i_k}(X_{i_1}, \dots, X_{i_k}) = \left(\prod_{j \notin \{i_1, \dots, i_k\}} \mathbb{E}_j \prod_{l \in \{i_1, \dots, i_k\}} \mathfrak{D}_l \right) f(X_1, \dots, X_n).$$

Many of the relations described in Remark (3.1) extend to higher order differences. In particular, for any $i_1 < \dots < i_d$,

$$(3.3) \quad (\mathfrak{v}_{i_1 \dots i_d} f(X))^2 = \mathbb{E}_{i_1 \dots i_d} (\mathfrak{D}_{i_1 \dots i_d} f(X))^2 = \mathbb{E}_{i_1 \dots i_d} (\mathfrak{d}_{i_1 \dots i_d} f(X))^2.$$

Furthermore, similarly to (3.1), we may rewrite (1.14) as

$$\mathfrak{d}_{i_1 \dots i_d} f(X) = \left(\frac{1}{2^d} \left((\mathfrak{D}_{i_1 \dots i_d} f(X))^2 + \sum_{k=1}^d \sum_{1 \leq s_1 < \dots < s_k \leq d} \mathbb{E}_{i_{s_1} \dots i_{s_k}} (\mathfrak{D}_{i_1 \dots i_d} f(X))^2 \right) \right)^{1/2}.$$

(3.3) implies that we always have

$$\mathbb{E} (\mathfrak{D}_{i_1 \dots i_d} f)^2 = \mathbb{E} (\mathfrak{v}_{i_1 \dots i_d} f)^2 = \mathbb{E} (\mathfrak{d}_{i_1 \dots i_d} f)^2.$$

Moreover, by the symmetry and Fubini's theorem,

$$\mathbb{E} (\mathfrak{d}_{i_1 \dots i_d}^+ f)^2 = \mathbb{E} (\mathfrak{d}_{i_1 \dots i_d}^- f)^2 = \frac{1}{2} \mathbb{E} (\mathfrak{d}_{i_1 \dots i_d} f)^2.$$

In particular,

$$(3.4) \quad \mathbb{E} |\mathfrak{D}^{(d)} f|^2 = \mathbb{E} |\mathfrak{v}^{(d)} f|^2 = \mathbb{E} |\mathfrak{d}^{(d)} f|^2 = 2 \mathbb{E} |\mathfrak{d}^{+(d)} f|^2 = 2 \mathbb{E} |\mathfrak{d}^{-(d)} f|^2.$$

Finally, as in Remark 3.1 (i), we may conclude that even the identity

$$\mathfrak{h}_{i_1 \dots i_d} f(\varepsilon) = \mathfrak{v}_{i_1 \dots i_d} f(\varepsilon) = \mathfrak{d}_{i_1 \dots i_d} f(\varepsilon) = |\mathfrak{D}_{i_1 \dots i_d} f(\varepsilon)|$$

holds for independent Rademacher variables $(\varepsilon_1, \dots, \varepsilon_n) = \varepsilon$.

The proof of Theorem 1.8 is based on L^2 -identities together with some kind of “harmonic” analysis arguments on the symmetric group. To this end, we shall need specific (higher order) operators \mathfrak{L}_d we would call powers of “Laplacians”. Here we make use of the difference operators \mathfrak{D}_i . That is, we set

$$(3.5) \quad \mathfrak{L}_d = \sum_{1 \leq i_1 \neq i_2 \neq \dots \neq i_d \leq n} \mathfrak{D}_{i_1} \dots \mathfrak{D}_{i_d}, \quad d \in \mathbb{N}.$$

In case of $d = 1$ this just means summing over all $i = 1, \dots, n$.

To motivate the notation of \mathfrak{L}_d , recall that for the discrete hypercube $\{\pm 1\}^n$, $\mathfrak{L}_1 = \sum_{i=1}^n \mathfrak{D}_i$ is the usual graph Laplacian. The higher order operators \mathfrak{L}_d can be written as polynomials in \mathfrak{L}_1 of total degree d . Note that $(\mathfrak{L}_1)^d$ can be expressed as a sum of d -th order differences $\mathfrak{D}_{i_1 \dots i_d}$. Hence it easily follows that \mathfrak{L}_d can be expressed in terms of $(\mathfrak{L}_1)^d$ by removing all the differences in which some indexes appear more than once. This is in accordance with our definition of the hyper-matrices $\mathfrak{D}^{(d)} f$ (cf. the discussion of (1.3)). Relating the Hoeffding decomposition to the Laplacian \mathfrak{L}_d yields the following result.

Theorem 3.2. *Let $f = f(X_1, \dots, X_n)$ be in $L^1(\mathbb{P})$ with Hoeffding decomposition $f = \sum_{k=0}^n f_k$. Then,*

$$\mathfrak{L}_d f_k = (k)_d f_k,$$

where \mathfrak{L}_d is the d -th order Laplacian (3.5), and $(k)_d = k(k-1) \dots (k-d+1)$. Thus, the k -th Hoeffding term is an eigenfunction of \mathfrak{L}_d with eigenvalue $(k)_d$.

Consequently, there is an orthogonal decomposition of L^2 -functions $f(X)$ on which the Laplacian \mathfrak{L}_d operates diagonally, and the eigenvalues of the Hoeffding terms of order up to $d-1$ are 0.

Proof. Write $f_k(X_1, \dots, X_n) = \sum_{j_1 < \dots < j_k} h_{j_1 \dots j_k}(X_{j_1}, \dots, X_{j_k})$ as in (1.7). Fix $j_1 < \dots < j_k$. Then, we get

$$\mathbb{E}_i h_{j_1 \dots j_k}(X_{j_1}, \dots, X_{j_k}) = \begin{cases} 0, & i \in \{j_1, \dots, j_k\}, \\ h_{j_1 \dots j_k}(X_{j_1}, \dots, X_{j_k}), & i \notin \{j_1, \dots, j_k\}. \end{cases}$$

Therefore,

$$(3.6) \quad \mathfrak{D}_i f_d(X_1, \dots, X_n) = \sum_{\substack{j_1 < \dots < j_k \\ i \in \{j_1, \dots, j_k\}}} h_{j_1 \dots j_k}(X_{j_1}, \dots, X_{j_k})$$

and consequently, by iteration,

$$(3.7) \quad \mathfrak{D}_{i_1 \dots i_d} f_k(X_1, \dots, X_n) = \sum_{\substack{j_1 < \dots < j_k \\ i_1, \dots, i_d \in \{j_1, \dots, j_k\}}} h_{j_1 \dots j_d}(X_{j_1}, \dots, X_{j_d}).$$

It remains to check how often each term $h_{j_1 \dots j_d}(X_{j_1}, \dots, X_{j_d})$ appears in $\mathfrak{L}_d f_k = \sum_{i_1 \neq \dots \neq i_d} \mathfrak{D}_{i_1 \dots i_d} f_k$. As we just saw, each d -tuple $i_1 \neq \dots \neq i_d$ such that $i_1, \dots, i_d \in \{j_1, \dots, j_k\}$ replicates the summand $h_{j_1 \dots j_d}(X_{j_1}, \dots, X_{j_d})$ precisely once. As there are $k(k-1) \dots (k-d+1) = (k)_d$ such tuples, we arrive at the result. \square

There is a ‘‘partial integration’’ formula involving difference operators. Recall that by [G-S], Lemma 5.1, the difference operators \mathfrak{D}_i are self-adjoint in the sense that $\mathbb{E}(\mathfrak{D}_i f)g = \mathbb{E}f(\mathfrak{D}_i g) = \mathbb{E}(\mathfrak{D}_i f)(\mathfrak{D}_i g)$ whenever $f(X)$ and $g(X)$ are in $L^2(\mathbb{P})$. In particular, for the Laplacians \mathfrak{L}_d from (3.5), we have

$$(3.8) \quad \mathbb{E}(\mathfrak{L}_d f)g = \mathbb{E}f(\mathfrak{L}_d g) = \sum_{i_1 \neq \dots \neq i_d} \mathbb{E}(\mathfrak{D}_{i_1 \dots i_d} f)(\mathfrak{D}_{i_1 \dots i_d} g).$$

This identity can be used to obtain the following relation:

Proposition 3.3. *Let $f(X) \in L^2(\mathbb{P})$ have the Hoeffding decomposition $f = \sum_{m=d}^n f_m$. If $k \leq d \leq n$, then*

$$\mathbb{E}|\mathfrak{D}^{(k-1)} f|^2 \leq \frac{1}{d-k+1} \mathbb{E}|\mathfrak{D}^{(k)} f|^2.$$

Equality holds only if $f = f_d$.

Proof. First, let $f = f_m$. Then, applying (3.8) leads to

$$\mathbb{E}|\mathfrak{D}^{(k)} f_m|^2 = \sum_{i_1 \neq \dots \neq i_k} \mathbb{E}(\mathfrak{D}_{i_1 \dots i_k} f_m)(\mathfrak{D}_{i_1 \dots i_k} f_m) = \mathbb{E}f_m(\mathfrak{L}_k f_m).$$

Moreover, Theorem 3.2 yields $\mathfrak{L}_k f_m = (m)_k f_m$. Consequently,

$$\mathbb{E}|\mathfrak{D}^{(k)} f_m|^2 = (m)_k \mathbb{E}f_m^2. \quad (*)$$

The same argument with k replaced by $k-1$ yields

$$\mathbb{E}|\mathfrak{D}^{(k-1)} f_m|^2 = (m)_{k-1} \mathbb{E}f_m^2. \quad (**)$$

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

For functions with arbitrary Hoeffding decomposition we shall use the orthogonality of the terms in the Hoeffding decomposition and obtain

$$\mathbb{E}|\mathfrak{D}^{(k-1)} f|^2 = \sum_{m=d}^n \frac{1}{m-k+1} \mathbb{E}|\mathfrak{D}^{(k)} f_m|^2 \leq \frac{1}{d-k+1} \mathbb{E}|\mathfrak{D}^{(k)} f|^2. \quad \square$$

Proof of Theorem 1.8. Due to (3.4), it suffices to prove Theorem 1.8 for the difference operator \mathfrak{D} . In this case, iterating Proposition 3.3 yields the result. \square

Finally, we prove (1.16). By orthogonality and Proposition 3.3,

$$\mathbb{E} |\mathbf{v}^{(k)} f|^2 = \sum_{j=k}^n \mathbb{E} |\mathbf{v}^{(k)} f_j|^2 = \sum_{j=k}^n \frac{1}{(j-k)!} \mathbb{E} |\mathbf{v}^{(j)} f_j|^2$$

for any $k = 1, \dots, n$, which may be rewritten as

$$\mathbb{E} |\mathbf{v}^{(k)} f_k|^2 = \mathbb{E} |\mathbf{v}^{(k)} f|^2 - \sum_{j=k+1}^n \frac{1}{(j-k)!} \mathbb{E} |\mathbf{v}^{(j)} f_j|^2.$$

Iteratively plugging this into (1.15), we obtain that

$$\text{Var} f = \sum_{k=1}^n R_k \mathbb{E} |\mathbf{v}^{(k)} f|^2, \quad R_k = \sum_{j=0}^{k-1} R_j \frac{1}{(k-j)!}, \quad R_0 := 1.$$

It follows that for $k \geq 1$, $R_k = (-1)^{k+1}/k!$ which finishes the proof.

4. DIFFERENTIABLE FUNCTIONS: PROOFS

Given a continuous function on an open subset $G \subset \mathbb{R}^n$, the equality

$$(4.1) \quad |\nabla f(x)| = \limsup_{x \rightarrow y} \frac{|f(x) - f(y)|}{|x - y|}, \quad x \in G,$$

may be used as definition of the generalized modulus of the gradient of f . The function $|\nabla f|$ is Borel measurable, and if f is differentiable at x , $|\nabla f(x)|$ agrees with the Euclidean norm of the usual gradient. This operator preserves many identities from calculus in form of inequalities, such as a ‘‘chain rule inequality’’

$$(4.2) \quad |\nabla T(f)| \leq |T'(f)| |\nabla f|,$$

where $|T'|$ is understood according to (4.1) again.

Using the generalized modulus of the gradient, there is an analogue of Lemma 2.2 for the operator norms of the derivatives of consecutive orders:

Lemma 4.1. *Given a \mathcal{C}^d -smooth function $f: G \rightarrow \mathbb{R}$, $d \in \mathbb{N}$, at all points $x \in G$,*

$$|\nabla |f^{(d-1)}(x)|_{\text{Op}}| \leq |f^{(d)}(x)|_{\text{Op}}.$$

Proof. Indeed, for any $h \in \mathbb{R}^n$, by the triangle inequality,

$$\begin{aligned} & \left| |f^{(d-1)}(x+h)|_{\text{Op}} - |f^{(d-1)}(x)|_{\text{Op}} \right| \leq |f^{(d-1)}(x+h) - f^{(d-1)}(x)|_{\text{Op}} \\ & = \sup\{(f^{(d-1)}(x+h) - f^{(d-1)}(x))[v_1, \dots, v_{d-1}]: v_1, \dots, v_{d-1} \in S^{n-1}\}, \end{aligned}$$

while, by the Taylor expansion,

$$(f^{(d-1)}(x+h) - f^{(d-1)}(x))[v_1, \dots, v_{d-1}] = f^{(d)}(x)[v_1, \dots, v_{d-1}, h] + o(|h|)$$

as $h \rightarrow 0$. Here, the o -term can be bounded by a quantity which is independent of $v_1, \dots, v_{d-1} \in S^{n-1}$. As a consequence,

$$\begin{aligned} & \limsup_{h \rightarrow 0} \frac{||f^{(d-1)}(x+h)|_{\text{Op}} - |f^{(d-1)}(x)|_{\text{Op}}|}{|h|} \\ & \leq \sup\{f^{(d)}(x)[v_1, \dots, v_{d-1}, v_d]: v_1, \dots, v_d \in S^{n-1}\} = |f^{(d)}(x)|_{\text{Op}}. \end{aligned}$$

□

For higher order concentration we need to establish a recursion for the L^p -norms of the derivatives of f of consecutive orders similarly to (2.15). To this end, we recall a classical result on the moments of Lipschitz functions in the presence of a logarithmic Sobolev inequality which goes back to Aida and Stroock [A-S]. Namely, if a probability measure μ on G satisfies a logarithmic Sobolev inequality (1.18) with constant σ^2 , then, for any locally Lipschitz function $g: G \rightarrow \mathbb{R}$, and any $p > 2$,

$$(4.3) \quad \|g\|_p^2 \leq \|g\|_2^2 + \sigma^2(p-2) \|\nabla g\|_p^2.$$

For the reader's convenience, let us briefly recall the argument. We may assume g to be bounded, in which case the squares of the $L_p(\mu)$ -norms of g have derivatives

$$(4.4) \quad \frac{d}{dp} \|g\|_p^2 = \frac{2}{p^2} \|g\|_p^{2-p} \text{Ent}_\mu(|g|^p).$$

We apply this identity to the function $u = |g|^{p/2}$. By the chain rule inequality (4.2), $|\nabla u|^2 \leq \frac{p^2}{4} |g|^{p-2} |\nabla g|^2$. Hence, by Hölder's inequality,

$$\int |\nabla u|^2 d\mu \leq \frac{p^2}{4} \left(\int |g|^p d\mu \right)^{\frac{p-2}{p}} \left(\int |\nabla g|^p d\mu \right)^{\frac{2}{p}} = \frac{p^2}{4} \|g\|_p^{p-2} \|\nabla g\|_p^2.$$

Applying (1.18) to the function u , we therefore obtain

$$\text{Ent}_\mu(|g|^p) = \text{Ent}_\mu(u^2) \leq 2\sigma^2 \int |\nabla u|^2 d\mu \leq \frac{p^2\sigma^2}{2} \|g\|_p^{p-2} \|\nabla g\|_p^2.$$

Combining this with (4.4), we arrive at the differential inequality $\frac{d}{dp} \|g\|_p^2 \leq \sigma^2 \|\nabla g\|_p^2$. Integrating it from 2 to p yields (4.3).

Combining Lemma 4.1 and (4.3), we are now able to prove Theorem 1.9.

Proof of Theorem 1.9. Using (4.3) with f replaced by $|f^{(k-1)}|_{\text{Op}}$, $1 \leq k \leq d$, we get

$$(4.5) \quad \begin{aligned} \|f^{(k-1)}\|_{\text{Op},p}^2 &\leq \|f^{(k-1)}\|_{\text{Op},2}^2 + \sigma^2(p-2) \|\nabla |f^{(k-1)}|_{\text{Op}}\|_p^2 \\ &\leq \|f^{(k-1)}\|_{\text{Op},2}^2 + \sigma^2(p-2) \|f^{(k)}\|_{\text{Op},p}^2, \end{aligned}$$

where Lemma 4.1 was applied on the last step. Consequently, by iteration,

$$(4.6) \quad \begin{aligned} \|f\|_p^2 &\leq \|f\|_2^2 + \sum_{k=1}^{d-1} (\sigma^2(p-2))^k \|f^{(k)}\|_{\text{Op},2}^2 + (\sigma^2(p-2))^d \|f^{(d)}\|_{\text{Op},p}^2 \\ &\leq \sigma^2 \|\nabla f\|_2^2 + \sum_{k=1}^{d-1} (\sigma^2(p-2))^k \|f^{(k)}\|_{\text{Op},2}^2 + (\sigma^2(p-2))^d \|f^{(d)}\|_{\text{Op},p}^2 \\ &\leq \sum_{k=1}^{d-1} (\sigma^2 p)^k \|f^{(k)}\|_{\text{Op},2}^2 + (\sigma^2 p)^d \|f^{(d)}\|_{\text{Op},p}^2. \end{aligned}$$

Here, the second step was based on the Poincaré-type inequality. Since $\|f^{(k)}\|_{\text{Op},2}^2 \leq \min(1, \sigma^{2(d-k)})$ for all $k = 1, \dots, d-1$ and $\|f^{(d)}\|_{\text{Op},\infty} \leq 1$ by assumption, we obtain

$$(4.7) \quad \|f\|_p^2 \leq \sigma^{2d} \sum_{k=1}^d p^k \leq \frac{1}{1-p^{-1}} (\sigma^2 p)^d \leq 2(\sigma^2 p)^d$$

and therefore $\|f\|_p \leq (2\sigma^2 p)^{d/2}$ for all $p \geq 2$. Moreover, $\|f\|_p \leq \|f\|_2 \leq (4\sigma^2)^{d/2}$ for $p < 2$. It follows that $\| |f|^{2/d} \|_k \leq \gamma k$ for all $k \in \mathbb{N}$, i.e. (2.17) with $\gamma = 4\sigma^2$. \square

Proof of Theorem 1.10. Starting as in the proof of Theorem 1.9, we arrive at

$$(4.8) \quad \|f\|_p^2 \leq \sum_{k=1}^{d-1} (\sigma^2 p)^k \|f^{(k)}\|_{\text{HS},2}^2 + (\sigma^2 p)^d \|f^{(d)}\|_{\text{Op},p}^2,$$

where we used the property that the operator norms are dominated by the Hilbert–Schmidt norms. Moreover, since $\int_G \partial_{i_1 \dots i_k} f \, d\mu = 0$, by the Poincaré-type inequality,

$$\int_G (\partial_{i_1 \dots i_k} f)^2 \, d\mu \leq \sigma^2 \sum_{j=1}^n \int_G (\partial_{i_1 \dots i_k j} f)^2 \, d\mu$$

whenever $1 \leq i_1, \dots, i_k \leq n$, $k \leq d-1$. Summing over all $1 \leq i_1, \dots, i_k \leq n$, we get

$$(4.9) \quad \|f^{(k)}\|_{\text{HS},2}^2 = \int_G |f^{(k)}|_{\text{HS}}^2 \, d\mu \leq \sigma^2 \int_G |f^{(k+1)}|_{\text{HS}}^2 \, d\mu = \sigma^2 \|f^{(k+1)}\|_{\text{HS},2}^2.$$

Using (4.9) in (4.8) and iterating, we thus obtain

$$\|f\|_p^2 \leq \sum_{k=1}^{d-1} \sigma^{2d} p^k \|f^{(d)}\|_{\text{HS},2}^2 + (\sigma^2 p)^d \|f^{(d)}\|_{\text{Op},p}^2.$$

Noting that $\|f^{(d)}\|_{\text{HS},2} \leq 1$ and $\|f^{(d)}\|_{\text{Op},\infty} \leq 1$, we arrive at (4.7), from where we may proceed as in the proof of Theorem 1.9. \square

Proof of Corollary 1.11. For any $p \geq 2$, it follows from (4.6) that

$$e \|f\|_p \leq e \left(\sum_{k=1}^{d-1} (\sigma^2 p)^{k/2} \|f^{(k)}\|_{\text{Op},2} + (\sigma^2 p)^{d/2} \|f^{(d)}\|_{\text{Op},\infty} \right).$$

From here we may proceed as in the proof of Corollary 1.4. \square

5. FUNCTIONS ON THE SPHERE: PROOFS

First, let us recall some basic facts about the spherical calculus (cf. [S-W] or e.g. [B-C-G]). The normalized Lebesgue measure σ_{n-1} on the unit sphere S^{n-1} can be introduced as the distribution of $Z/|Z|$, assuming that the random vector Z has a standard normal distribution in \mathbb{R}^n . Using independence of $|Z|$ and $Z/|Z|$, this description implies that for any p -homogeneous function $f: \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}$,

$$(5.1) \quad \int_{\mathbb{R}^n} f(x) \, d\gamma_n(x) = \mathbb{E} |Z|^p \int_{S^{n-1}} f(\theta) \, d\sigma_{n-1}(\theta)$$

provided that all the integrals involved exist.

A function f on S^{n-1} is called \mathcal{C}^d -smooth if it can be extended to a \mathcal{C}^d -smooth function on some open subset G of \mathbb{R}^n containing the unit sphere. If f is \mathcal{C}^1 -smooth on S^{n-1} , then at every point $\theta \in S^{n-1}$ it admits the Taylor expansion

$$(5.2) \quad f(\theta') = f(\theta) + \langle v, \theta' - \theta \rangle + o(|\theta' - \theta|) \quad \text{as } \theta' \rightarrow \theta, \quad \theta' \in S^{n-1}$$

with some $v \in \mathbb{R}^n$. Among all the vectors v fulfilling (5.2), the one with smallest Euclidean norm represents the spherical derivative or gradient of f at θ and is denoted $\nabla_S f(\theta)$. Equivalently, in terms of the usual (Euclidean) gradient, we have

$$\nabla_S f(\theta) = P_{\theta^\perp} \nabla f(\theta) = \nabla f(\theta) - \langle \nabla f(\theta), \theta \rangle \theta,$$

where P_{θ^\perp} denotes the orthogonal projection from \mathbb{R}^n to the tangent space θ^\perp . In particular, $|\nabla_S f(\theta)| \leq |\nabla f(\theta)|$ for all $\theta \in S^{n-1}$. If f is a \mathcal{C}^1 -function on S^{n-1} , the norm of the spherical gradient $|\nabla_S f|$ coincides with the generalized modulus of the gradient (4.1) using either the geodesic or the induced Euclidean distance on S^{n-1} .

By a result of Mueller and Weissler [M-W], the uniform measure σ_{n-1} on S^{n-1} satisfies a logarithmic Sobolev inequality with constant $\sigma^2 = \frac{1}{n-1}$. In other words,

$$(5.3) \quad \text{Ent}_{\sigma_{n-1}}(f^2) \leq \frac{2}{n-1} \int |\nabla_S f|^2 d\sigma_{n-1}$$

for any smooth $f: S^{n-1} \rightarrow \mathbb{R}$. Therefore, considering any open neighbourhood G of S^{n-1} , we may regard σ_{n-1} as a Borel probability measure on G satisfying a logarithmic Sobolev inequality with constant $\sigma^2 = \frac{1}{n-1}$. Hence, Theorem 1.13 directly follows from Theorem 1.9. It remains to note that in the notation of Theorem 1.9,

$$\min(1, \sigma^{d-k}) \geq n^{-(k-d)/2},$$

hence arriving at the conditions used in Theorem 1.13.

In a similar way, we now prove Theorem 1.14.

Proof of Theorem 1.14. It follows from Theorem 1.10, once the partial derivatives up to order $d-1$ are centered under σ_{n-1} . Indeed, we may assume that f is defined on $\mathbb{R}^n \setminus \{0\}$ (cf. (5.4)). Fix any $i_1 \leq \dots \leq i_k$, $k \leq d-1$. Noting that $x_{i_1} \cdots x_{i_k} f(x)$ is $(p+k)$ -homogeneous, by (5.1) and a k -fold partial integration, we obtain

$$\begin{aligned} \int_{\mathbb{R}^n} \partial_{i_1 \dots i_k} f(x) d\gamma_n(x) &= \int_{\mathbb{R}^n} x_{i_1} \cdots x_{i_k} f(x) d\gamma_n(x) \\ &= \mathbb{E} |Z|^{p+k} \int_{S^{n-1}} \theta_{i_1} \cdots \theta_{i_k} f(\theta) d\sigma_{n-1}(\theta). \end{aligned}$$

On the other hand, since f is p -homogeneous, $\partial_{i_1 \dots i_k} f$ is $(p-k)$ -homogeneous. Therefore, applying (5.1) again,

$$\int_{\mathbb{R}^n} \partial_{i_1 \dots i_k} f(x) d\gamma_n(x) = \mathbb{E} |Z|^{p-k} \int_{S^{n-1}} \partial_{i_1 \dots i_k} f(\theta) d\sigma_{n-1}(\theta).$$

Hence, orthogonality to all polynomials of total degree at most $d-1$ implies that the partial derivatives up to order $d-1$ are centered (if the involved integrals exist). Since f is a \mathcal{C}^d -function (hence bounded on S^{n-1}), this holds true if $\mathbb{E} |Z|^{p-k} < \infty$ for all $k = 0, 1, \dots, d-1$, which in turn is satisfied iff $p - (d-1) + n > 0$. \square

To prove Theorem 1.15, we need some further details about spherical derivatives. First note that $\nabla_S f$ is a vector-valued function on S^{n-1} , and hence we may define spherical partial derivatives of first and higher orders as suggested in Section 1.4.

Any function f on S^{n-1} can be extended to a p -homogeneous function F on \mathbb{R}^n (where $p \in \mathbb{R}$) by putting

$$(5.4) \quad F(x) = \begin{cases} r^p f(\theta), & x \neq 0, \\ 0, & x = 0, \end{cases} \quad r = |x|, \quad \theta = x/|x|.$$

If f is \mathcal{C}^d -smooth, its p -homogeneous extension F will be \mathcal{C}^d -smooth on $\mathbb{R}^n \setminus \{0\}$.

The spherical derivative $\nabla_S f$ of a \mathcal{C}^1 -smooth function on S^{n-1} and the derivatives of its p -homogeneous extensions F are related by the identity

$$\nabla F(x) = r^{p-1} [pf(\theta)\theta + \nabla_S f(\theta)], \quad x \neq 0$$

(see [B-C-G], Proposition 12.1). In particular, for the 0-homogeneous extension $F^{(0)}(x) = f(\theta)$, $\nabla F^{(0)}$ is (-1) -homogeneous and is given by

$$(5.5) \quad \nabla F^{(0)}(x) = r^{-1} \nabla_S f(\theta),$$

so that $\nabla F^{(0)} = \nabla_S f$ on S^{n-1} . In other words, $\partial_i F^{(0)} = D_i f$ on S^{n-1} , where ∂_i and D_i denote the partial and spherical partial derivatives, respectively.

By iteration, we can retrieve spherical partial derivatives of any order from suitable homogeneous extensions. To start, note that $F^{(1)}(x) = r \nabla F^{(0)}(x)$ is a 0-homogeneous vector-valued function on $\mathbb{R}^n \setminus \{0\}$. It follows that for any $1 \leq i, j \leq n$, $\partial_i F_j^{(1)}$ is a (-1) -homogeneous function with

$$\partial_i F_j^{(1)}(x) = \partial_i (r \partial_j F^{(0)}(x)) = r^{-1} D_i \langle \nabla_S f(\theta), e_j \rangle = r^{-1} D_{ij} f(\theta),$$

so that $\partial_i F_j^{(1)} = D_{ij} f$ on S^{n-1} . For general $k \in \mathbb{N}$, we may therefore define

$$(5.6) \quad F^{(k)}(x) = r \nabla F^{(k-1)}(x),$$

which is a 0-homogeneous function on $\mathbb{R}^n \setminus \{0\}$ taking values in \mathbb{R}^{n^k} . Arguing as above, for any $1 \leq i_1, \dots, i_k \leq n$,

$$(5.7) \quad F_{i_1 \dots i_k}^{(k)} = D_{i_1 \dots i_k} f$$

on S^{n-1} , where $D_{i_1 \dots i_k} f$ denotes the k -th order spherical partial derivatives of f .

With the help of the 0-homogeneous functions $F^{(k)}$, $k = 0, 1, \dots, d$, it is now possible to adapt the key steps of the proof of Theorem 1.9.

Lemma 5.1. *If $f: S^{n-1} \rightarrow \mathbb{R}$ is \mathcal{C}^d -smooth, then for all $\theta \in S^{n-1}$,*

$$|\nabla |D^{(d-1)} f(\theta)|| \leq |D^{(d)} f(\theta)|.$$

Proof. Noting that $|F^{(d-1)}(x)|$ is the 0-homogeneous extension of $|D^{(d-1)} f(\theta)|$, we may perform similar arguments as in the proof of Lemma 4.1. \square

Proof of Theorem 1.15. Since the uniform distribution on the sphere satisfies a logarithmic Sobolev inequality (5.3), the moment recursion (4.3) from the proof of Theorem 1.9 remains valid for functions on the sphere with intrinsic derivatives. Hence, using (4.3) with f replaced by $|D^{(k-1)} f|_{\text{Op}}$ for $k = 1, \dots, d$, we obtain

$$\begin{aligned} \|D^{(d-1)} f\|_{\text{Op}, p}^2 &\leq \|D^{(d-1)} f\|_{\text{Op}, 2}^2 + \sigma^2(p-2) \|\nabla^* |D^{(d-1)} f|_{\text{Op}}\|_p^2 \\ &\leq \|D^{(d-1)} f\|_{\text{Op}, 2}^2 + \sigma^2(p-2) \|D^{(d)} f\|_{\text{Op}, p}^2. \end{aligned}$$

Here the last step relies upon Lemma 5.1. But this is the same inequality as (4.5). Therefore, the rest of the proof is analogous to the proof of Theorem 1.9. \square

6. POLYNOMIALS AND EDGEWORTH-TYPE EXPANSIONS ON THE SPHERE

For vectors $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ and an integer $d \geq 3$, consider the polynomials

$$Q_{d,a}(\theta) = \sum_{i=1}^n a_i \theta_i^d, \quad \theta \in S^{n-1},$$

extending them to \mathbb{R}^n by setting $Q_{d,a}(x) = \sum_{i=1}^n a_i x_i^d$. By an easy calculation,

$$\int_{S^{n-1}} \theta_i^{2p} d\sigma_{n-1}(\theta) = \frac{(2p-1)!!}{n(n+2)\cdots(n+2p-2)}, \quad i = 1, \dots, n, \quad p \in \mathbb{N}.$$

Therefore, differentiating $Q_{d,a}(x)$, it follows that, for any $k = 1, \dots, d-1$,

$$\|Q_{d,a}^{(k)}\|_{\text{HS},2}^2 = \frac{(2(d-k)-1)!! (d)_k^2}{n(n+2)\cdots(n+2(d-k-1))} \sum_{i=1}^n a_i^2,$$

where we used the notation $(d)_k = d(d-1)\cdots(d-k+1)$. Similarly,

$$|Q_{d,a}^{(d)}(\theta)|_{\text{HS}}^2 = (d!)^2 \sum_{i=1}^n a_i^2.$$

As a consequence, using the normalization $n^{-1} \sum_{i=1}^n a_i^2 = 1$ and choosing a suitable constant c_d , we see that the function $n^{-1/2} c_d Q_{d,a}$ satisfies the conditions of Theorem 1.13. Summarizing, we arrive at the following result.

Proposition 6.1. *Let $d \geq 3$ and let $n^{-1} \sum_{i=1}^n a_i^2 = 1$. There exists some constant $c_d > 0$ depending on d only such that*

$$\int_{S^{n-1}} \exp\left(c_d n^{(d-1)/d} |Q_{d,a} - \bar{Q}_{d,a}|^{2/d}\right) d\sigma_{n-1} \leq 2,$$

where $\bar{Q}_{d,a}$ denotes the σ_{n-1} -mean of $Q_{d,a}$. In particular, $Q_{d,a} - \bar{Q}_{d,a} = \mathcal{O}_{\sigma_{n-1}}(n^{-\frac{d-1}{2}})$.

Note though that if d is odd, $\bar{Q}_{d,a} = 0$, while for d even,

$$(6.1) \quad \bar{Q}_{d,a} = \frac{(d-1)!!}{(n+2)\cdots(n+d-2)} \bar{a}, \quad \bar{a} = n^{-1} \sum_{i=1}^n a_i.$$

In order to illustrate possible applications of these results, consider smooth function of the form

$$h_n(\theta) = \mathbb{E} H\left(\sum_{i=1}^n \theta_i X_i\right), \quad \theta \in S^{n-1},$$

with $X_i \in \mathbb{R}^k$ independent and such that $\text{Cov}(X_i) = \text{Id}$. For simplicity, let us assume that X_i is symmetric, i. e. $X_i = -X_i$ in distribution. It is known, cf. [G-H], that $h_n(\theta)$ may be approximated via Edgeworth expansions, in particular – by the polynomial $\Gamma_0 + \sum_{i=1}^n \Gamma_{4,i} \theta_i^4$, as long as $H \in \mathcal{C}^4(\mathbb{R}^k)$ and assuming that the quantity

$$M = \sup_{x \in \mathbb{R}^k} \left(|H(x)| + \sup_{|\alpha|=4} |\partial^\alpha H(x)| \right) (1 + |x|^6)^{-1}$$

is finite. The Γ -terms are the non-vanishing even Edgeworth expansion terms defined by $\Gamma_0 = \mathbb{E} H(N)$, where N is a standard normal random vector in \mathbb{R}^k , and by

$$\Gamma_{4,i} = \frac{1}{24} \left(\frac{\partial^4}{\partial \varepsilon^4} \Big|_{\varepsilon=0} \mathbb{E} H(N + \varepsilon X_j) - 3 \frac{\partial^4}{\partial \varepsilon_1^2 \partial \varepsilon_2^2} \Big|_{\varepsilon_1=\varepsilon_2=0} \mathbb{E} H(N + \varepsilon_1 X_i + \varepsilon_2 \bar{X}_i) \right),$$

where X_i, \bar{X}_i, N are independent and \bar{X}_i denotes an independent copy of X_i .

If furthermore $\rho_{p,i} = \mathbb{E} |X_i|^p < \infty$ for $p \leq 6$, then an inspection of the proof of [G-H], Theorem 3.6, yields an explicit bound for the error

$$(6.2) \quad R(\theta) = h_n(\theta) - \left(\Gamma_0 + \sum_{i=1}^n \Gamma_{4,i} \theta_i^4 \right),$$

namely

$$(6.3) \quad |R(\theta)| \leq c_M \left(\sum_{i=1}^n \rho_{6,i} \theta_i^6 + \left(\sum_{i=1}^n \rho_{3,i} |\theta_i|^3 \right)^4 \right) \leq c_M \sum_{i=1}^n (\rho_{6,i} + \rho_{3,i}^4) \theta_i^6$$

with some constant c_M depending on M only.

To study the asymptotic behaviour of $h_n(\theta)$ as a function of $\theta \in S^{n-1}$ as $n \rightarrow \infty$, we apply Proposition 6.1 together with (6.1). Here we are interested in concentration inequalities for $R(\theta)$, i. e. we do not only center around the constant term Γ_0 but also include the fourth order term $\sum_{i=1}^n \Gamma_{4,i} \theta_i^4$. Indeed, write

$$Q_{6,\rho}(\theta) = \sum_{i=1}^n (\rho_{6,i} + \rho_{3,i}^4) \theta_i^6,$$

so that $|R(\theta)| \leq c_M Q_{6,\rho}(\theta)$ by (6.3). Dividing $Q_{6,\rho}$ by $\rho_* = (\frac{1}{n} \sum_{i=1}^n (\rho_{6,i} + \rho_{3,i}^4))^2)^{1/2}$, we may apply Proposition 6.1 with $d = 6$, which yields

$$(6.4) \quad \int_{S^{n-1}} \exp \left(\frac{c_6}{\rho_*^{1/3}} n^{5/6} |Q_{6,\rho}(\theta) - \bar{Q}_{6,\rho}|^{1/3} \right) d\sigma_{n-1}(\theta) \leq 2$$

for some absolute constant $c_6 > 0$. Furthermore, by (6.1),

$$\bar{Q}_{6,\rho} = \frac{15}{(n+2)(n+4)} \bar{\rho}, \quad \bar{\rho} = \frac{1}{n} \sum_{i=1}^n (\rho_{6,i} + \rho_{3,i}^4).$$

In particular,

$$(6.5) \quad \int_{S^{n-1}} \exp \left(\frac{c_0}{\rho_*^{1/3}} n^{2/3} |\bar{Q}_{6,\rho}|^{1/3} \right) d\sigma_{n-1}(\theta) \leq 2$$

for some absolute constant $c_0 > 0$. Applying the Cauchy–Schwarz inequality together with (6.4) and (6.5), we therefore arrive at the concentration inequality

$$\int_{S^{n-1}} \exp \left(\frac{c}{\rho_*^{1/3}} n^{2/3} |R(\theta)|^{1/3} \right) d\sigma_{n-1}(\theta) \leq 2$$

for some absolute constant $c > 0$ depending on M only. One may take $c = c_M^{1/3} \min(c_0, c_6)/2$ with c_M from (6.3).

This example is related to a general framework of symmetric functions introduced by Götze, Naumov and Ulyanov [G-N-U]. Indeed, we may consider sequences of real functions $h_n(\theta_1, \dots, \theta_n)$, defined on \mathbb{R}^n such that

$$(6.6) \quad \begin{aligned} h_{n+1}(\theta_1, \dots, \theta_j, 0, \theta_{j+1}, \dots, \theta_n) &= h_n(\theta_1, \dots, \theta_j, \theta_{j+1}, \dots, \theta_n); \\ \frac{\partial}{\partial \theta_j} h_n(\theta_1, \dots, \theta_j, \dots, \theta_n) \Big|_{\theta_j=0} &= 0 \quad \forall j = 1, \dots, n; \\ h_n(\theta_{\pi(1)}, \dots, \theta_{\pi(n)}) &= h_n(\theta_1, \dots, \theta_n) \quad \forall \pi \in S_n, \end{aligned}$$

where S_n denotes the symmetric group. This model may be regarded as a general scheme which holds true in a lot of situations where asymptotic expansions are considered. As shown in [G-N-U], Conditions (6.6) ensure the existence of a “limit”

function $h_\infty(\sum_i \theta_i^2, \lambda_1, \dots, \lambda_s)$, $s \in \mathbb{N}_0$, together with “Edgeworth-type” asymptotic expansions. These expansions are given in terms of polynomials of $Q_d(\theta)$, $d \geq 3$, where $Q_d(\theta) = Q_{d,a^*}$ for $a^* = (1, \dots, 1)$, and coefficients given by the derivatives of the limit function h_∞ at $\lambda_1 = \dots = \lambda_s = 0$. In particular, if we assume $\theta \in S^{n-1}$, applying Proposition 6.1 yields higher order concentration results for $(h_n)_n$.

In a certain sense, this represents a higher order extension of the second order results by Klartag [K] for distribution functions of spherical weighted sums for log-concave measures. See also the results by Klartag and Sodin [K-S] for sums of independent random variables.

REFERENCES

- [A] Adamczak, R.: *Moment inequalities for U-statistics*. The Annals of Probability 34(6) (2006), 2288–2314.
- [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-S] Aida, S., Stroock, D.: *Moment estimates derived from Poincaré and logarithmic Sobolev inequalities*. Math. Res. Lett. 1(1) (1994), 75–86.
- [A-G] Arcones, M. A., Giné, E.: *Limit theorems for U-processes*. Ann. Prob. 21 (1993), 1494–1542.
- [B] Bobkov, S. G.: *The growth of L^p -norms in presence of logarithmic Sobolev inequalities*. Vestnik Syktyvkar Univ. 11(1) (2010), 92–111.
- [B-C-G] Bobkov, S. G., Chistyakov, G. P., Götze, F.: *Second Order Concentration on the Sphere*. Commun. Contemp. Math. 19(5) (2017), 1650058, 20 pp.
- [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-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.
- [B-B-L-M] Boucheron, S., Bousquet, O., Lugosi, G., Massart, P.: *Moment inequalities for functions of independent random variables*. The Annals of Probability 33(2) (2005), 514–560.
- [B-L-M] Boucheron, S., Lugosi, G., Massart, P.: *Concentration Inequalities. A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.
- [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.
- [E-S] Efron, B., Stein, C.: *The jackknife estimate of variance*. Ann. Statist. 9 (1981), 586–596.
- [G-H] Götze, F., Hipp, C.: *Asymptotic expansions in the central limit theorem under moment conditions*. Z. Wahrscheinlichkeitstheorie verw. Gebiete 42(1) (1978), 67–87.
- [G-N-U] Götze, F., Naumov, A., Ulyanov, V.: *Asymptotic analysis of symmetric functions*. Preprint, arXiv:1502.06267.
- [G-S] Götze, F., Sambale, H.: *Second order concentration via logarithmic Sobolev inequalities*. Preprint, arXiv:1605.08635.
- [G-S-S] Götze, F., Sambale, H., Sinulis, A.: *Higher order concentration for functions of weakly dependent random variables*. Preprint, arXiv:1801.06348.
- [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. Probab. 5 (2000), 119–136.

- [Hoe] Hoeffding, W.: *A class of statistics with asymptotically normal distribution*. Ann. Math. Statist. 19 (1948), 293–325.
- [Hou] Houdré, C.: *The iterated jackknife estimate of variance*. Stat. Probab. Lett. 32(2) (1997), 197–201.
- [J-L] Joly, E., Lugosi, G.: *Robust estimation of U -statistics*. Stoch. Proc. Appl. 126 (2016), 3760–3773.
- [K] Klartag, B.: *A Berry–Esseen type inequality for convex bodies with an unconditional basis*. Probab. Theory Related Fields 145(1–2) (2009), 1–33.
- [K-S] Klartag, B., Sodin, S.: *Variations on the Berry–Esseen theorem*. Teor. Veroyatn. Primen. 56(3) (2011), 514–533; translation in Theory Probab. Appl. 56(3) (2012), 403–419.
- [L-O] Latala, R.; Oleszkiewicz, K.: *Between Sobolev and Poincaré*. Geometric aspects of functional analysis, 147–168, Lecture Notes in Math., 1745, Springer, Berlin, 2000.
- [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-S] Milman, V. D.; Schechtman, Gideon Asymptotic theory of finite-dimensional normed spaces. With an appendix by M. Gromov. Lecture Notes in Mathematics, 1200. Springer-Verlag, Berlin, 1986. viii+156 pp.
- [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.
- [M-W] Mueller, C. E., Weissler, F. B.: *Hypercontractivity for the heat semigroup for ultraspherical polynomials and on the n -sphere*. J. Funct. Anal. 48 (1992), 252–283.
- [S-W] Stein, E. M., Weiss, G.: *Introduction to Fourier analysis on Euclidean spaces*. Princeton University Press, 1971.
- [T] Talagrand, M. Concentration of measure and isoperimetric inequalities in product spaces. Inst. Hautes Études Sci. Publ. Math. No. 81 (1995), 73–205.
- [W] Wolff, P.: *On some Gaussian concentration inequality for non-Lipschitz functions*. High Dimensional Probability VI, Progress in Probability 66, 103–110.

SERGEY G. BOBKOV, SCHOOL OF MATHEMATICS, UNIVERSITY OF MINNESOTA, MINNEAPOLIS, USA; NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS, RUSSIAN FEDERATION

E-mail address: bobkov@math.umn.edu

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