



Modified log-Sobolev inequalities and two-level concentration

Holger Sambale and Arthur Sinulis

Faculty of Mathematics, Bielefeld University,
Universitätsstraße 25, 33613 Bielefeld, Germany.
E-mail address: hsambale@math.uni-bielefeld.de, asinulis@math.uni-bielefeld.de

Abstract. We consider a generic modified logarithmic Sobolev inequality (mLSI) of the form $\text{Ent}_\mu(e^f) \leq \frac{\rho}{2} \mathbb{E}_\mu e^f \Gamma(f)^2$ for *some* difference operator Γ , and show how it implies two-level concentration inequalities akin to the Hanson–Wright or Bernstein inequality. This can be applied to the continuous (e. g. the sphere or bounded perturbations of product measures) as well as discrete setting (the symmetric group, finite measures satisfying an approximate tensorization property, ...).

Moreover, we use modified logarithmic Sobolev inequalities on the symmetric group S_n and for slices of the hypercube to prove Talagrand’s convex distance inequality, and provide concentration inequalities for locally Lipschitz functions on S_n . Some examples of known statistics are worked out, for which we obtain the correct order of fluctuations, which is consistent with central limit theorems.

1. Introduction

Concentration and one-sided deviation inequalities have become an indispensable tool of probability theory and its applications. A question that arises frequently is to bound the fluctuations of a function $f = f(X_1, \dots, X_n)$ of many random variables (or, equivalently, a function on a product space) around its mean, and often times it is possible to prove sub-Gaussian tail decay of the form

$$\mathbb{P}(f(X) - \mathbb{E} f(X) \geq t) \leq C \exp\left(-\frac{t^2}{2K^2}\right)$$

Received by the editors January 22th, 2020; accepted February 9th, 2021.

2010 *Mathematics Subject Classification.* 60E15, 20B30.

Key words and phrases. Bernstein inequality, concentration of measure phenomenon, convex distance inequality, Hanson–Wright inequality, modified logarithmic Sobolev inequality, symmetric group.

This research was supported by the German Research Foundation (DFG) via CRC 1283 “*Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications*”.

for some $C \geq 1$, $K^2 > 0$ and all $t \geq 0$. There are various ways to establish sub-Gaussian estimates, such as the *martingale method*, the *entropy method* and the *information-theoretic approach* as well as *Stein's method*, and we refer to the monograph [Boucheron et al. \(2013\)](#) for further details.

On the other hand, in some situations it is not possible to prove sub-Gaussian tails, and a suitable replacement might be Bernstein-type

$$\mathbb{P}(f(X) - \mathbb{E}f(X) \geq t) \leq C \exp\left(-\frac{t^2}{2(a+bt)}\right)$$

or Hanson–Wright-type inequalities

$$\mathbb{P}(f(X) - \mathbb{E}f(X) \geq t) \leq C \exp\left(-\min\left(\frac{t^2}{a}, \frac{t}{b}\right)\right).$$

As both inequalities show two different levels of tail decay (the Gaussian one for $t \leq ab^{-1}$ and an exponential one for $t > ab^{-1}$), we use the terminology of [Adamczak et al. \(2017, 2019\)](#) and call inequalities of these type *two-level deviation inequalities*. If a similar estimate holds for $-f(X)$ as well, we refer to these as two-level *concentration* inequalities. For another approach to two-level concentration involving Bernstein–Orlicz norms, see [van de Geer and Lederer \(2013\)](#).

The purpose of this note is to give a unified treatment of some of the existing literature on two-level deviation and concentration inequalities by showing that these are implied by a modified logarithmic Sobolev inequality (mLSI for short). We prove a general theorem providing two-level deviation and concentration inequalities in various frameworks. In particular, in Section 2, we get back and partially improve a number of earlier results like [Bobkov et al. \(2017\)](#) and [Götze and Sambale \(2020\)](#).

We work in a general framework which was introduced in [Bobkov and Götze \(1999\)](#). Consider a probability space $(\Omega, \mathcal{F}, \mu)$ and let $\mathbb{E}_\mu f$ denote the expectation of a random variable f with respect to μ . An operator Γ on a class \mathcal{A} of bounded, measurable functions is called a *difference operator*, if

- (1) for all $f \in \mathcal{A}$, $\Gamma(f)$ is a non-negative measurable function,
- (2) for all $f \in \mathcal{A}$ and $a \geq 0, b \in \mathbb{R}$ we have $af + b \in \mathcal{A}$ and $\Gamma(af + b) = a\Gamma(f)$.

At first reading, one can think of $\Gamma(f) = |\nabla f|$ in the setting $\Omega = \mathbb{R}^n$. However, we want to stress that we do not require Γ to satisfy a chain rule, and Γ does not need to be an operator in the language of functional analysis.

We say that μ satisfies a Γ -mLSI(ρ) for some $\rho > 0$, if for all $f \in \mathcal{A}$ we have

$$\text{Ent}_\mu(e^f) \leq \frac{\rho}{2} \mathbb{E}_\mu \Gamma(f)^2 e^f, \quad (1.1)$$

where $\text{Ent}_\mu(f) = \mathbb{E}_\mu f \log f - \mathbb{E}_\mu f \log(\mathbb{E}_\mu f)$ ($f \geq 0$) is the entropy functional. This functional inequality is well-known in the theory of concentration of measure and has been used in various works, see [Bobkov and Götze \(1999\)](#) and the references therein. It is well-known that if μ satisfies a Γ -mLSI(ρ), we have for any function $f \in \mathcal{A}$ such that $\Gamma(f) \leq 1$,

$$\mu(f - \mathbb{E}_\mu f \geq t) \leq \exp\left(-\frac{t^2}{2\rho}\right), \quad (1.2)$$

which is a classical first order concentration of measure result yielding subgaussian concentration (cf. (3.5)). It is not hard to see that the same holds for $-f$ if

$\Gamma(af) = |a|\Gamma(f)$ for all $a \in \mathbb{R}$. Furthermore, it is well-known that (1.1) implies a Poincaré inequality

$$\text{Var}_\mu(f) \leq \rho \mathbb{E}_\mu \Gamma(f)^2,$$

which in particular yields variance bounds for any $f \in \mathcal{A}$.

Our first goal is to establish second order analogues of (1.2).

1.1. *Two-level concentration inequalities.* Our first set of results are two-level deviation inequalities for probability measures satisfying a modified logarithmic Sobolev inequality.

Theorem 1.1. *Assume that μ satisfies a Γ -mLSI(ρ) for some difference operator Γ and $\rho > 0$. Let $f, g : \Omega \rightarrow \mathbb{R}$ be two measurable functions such that $\Gamma(f) \leq g$ and g is sub-Gaussian, i. e. for some $c > 0, C \geq 1, K > 0$ and for all $t \geq 0$*

$$\mu(g \geq c + t) \leq C \exp\left(-\frac{t^2}{2K^2}\right). \tag{1.3}$$

Then for all $t \geq 0$ it holds

$$\mu(f - \mathbb{E}_\mu f \geq t) \leq \frac{4C}{3} \exp\left(-\frac{1}{8} \min\left(\frac{t^2}{\rho c^2}, \frac{t}{\sqrt{\rho K}}\right)\right). \tag{1.4}$$

If moreover $\Gamma(af) = |a|\Gamma(f)$ for all $a \in \mathbb{R}$, we have

$$\mu(|f - \mathbb{E}_\mu f| \geq t) \leq 2C \exp\left(-\frac{1}{12} \min\left(\frac{t^2}{\rho c^2}, \frac{t}{\sqrt{\rho K}}\right)\right). \tag{1.5}$$

By elementary means (cf. (3.1)), the constant $4C/3$ can be replaced by any $C' > 1$ (at the cost of worsening the constant in the exponential). It is also possible to modify our proofs in order to apply [Klochkov and Zhivotovskiy \(2020, Lemma 1.3\)](#), which leads to an inequality of the form

$$\mu(f - \mathbb{E}_\mu f \geq t) \leq \exp\left(-c \min\left(\frac{t^2}{\rho(\mathbb{E}_\mu g)^2 + 2b^2\rho^2}, \frac{t}{\sqrt{2}\rho b}\right)\right)$$

for some absolute constant c (the same one as in [Klochkov and Zhivotovskiy, 2020](#)). However, this is at the cost of a weaker denominator in the Gaussian term as compared to (1.4), and so we choose to present it in the form of Theorem 1.1.

One possible way to show sub-Gaussian concentration for g in presence of a Γ -mLSI(ρ) is by the Herbst argument. This leads to the following corollary.

Corollary 1.2. *Assume that μ satisfies a Γ -mLSI(ρ) for some difference operator Γ and $\rho > 0$. Let f, g be two measurable functions such that $\Gamma(f) \leq g$ and $\Gamma(g) \leq b$. Then for all $t \geq 0$ we have*

$$\mu(f - \mathbb{E}_\mu f \geq t) \leq \frac{4}{3} \exp\left(-\frac{1}{8\rho} \min\left(\frac{t^2}{(\mathbb{E}_\mu g)^2}, \frac{t}{b}\right)\right).$$

If, again, $\Gamma(af) = |a|\Gamma(f)$ for all $a \in \mathbb{R}$, then the same bound holds for $-f$.

If the difference operator Γ satisfies a chain rule-type condition, we obtain the following result, especially improving some of the constants above:

Proposition 1.3. *Assume that μ satisfies a Γ -mLSI(ρ) for some $\rho > 0$ and some difference operator Γ which satisfies $\Gamma(g^2) \leq 2g\Gamma(g)$ for all positive functions g . Let $f \in \mathcal{A}$ be such that $\Gamma(f) \leq g$ and $\Gamma(g) \leq b$. For any $t \geq 0$ it holds*

$$\mu(f - \mathbb{E}_\mu f \geq t) \leq \exp\left(-\frac{1}{4\rho} \min\left(\frac{t^2}{2\mathbb{E}_\mu g^2}, \frac{t}{b}\right)\right).$$

If Γ satisfies $\Gamma(g^2) \leq 2|g|\Gamma(g)$ for any $g \in \mathcal{A}$, the same bound holds for $-f$.

We will see a number of examples of such difference operators all along this paper. Obviously, one example is the usual gradient, but also many difference operators involving a positive part satisfy the property in question.

In all the above results, a possible choice of g is usually given by $g = \Gamma(f)$, resulting in $\mathbb{E}_\mu \Gamma(f)$ in the denominator of the Gaussian term. In this case, the second condition reads as $\Gamma(\Gamma(f)) \leq b$, which can be understood as a condition on an iterated (and thus second order) difference of f .

In fact, Theorem 1.1 can be understood as a Bernstein-type concentration inequality. Indeed, it is easy to see that for all $a, b > 0$ and $t \geq 0$ we have

$$\frac{t^2}{a^2 + bt} \leq \min\left(\frac{t^2}{a^2}, \frac{t}{b}\right) \leq \frac{2t^2}{a^2 + bt}.$$

In particular, in the situation of Theorem 1.1, for all $t \geq 0$ we have

$$\mu(f - \mathbb{E}_\mu f \geq t) \leq \frac{4C}{3} \exp\left(-\frac{t^2}{8(\rho c^2 + \sqrt{\rho}Kt)}\right).$$

If $\Gamma(af) = |a|\Gamma(f)$ for all $a \in \mathbb{R}$, the same bound holds with f replaced by $-f$.

Let us remark that the use of modified LSIs allows us to prove results for some classes of measures we could not address in previous work (e.g. Götze et al., 2020), e.g. weakly dependent measures which might not have a finite number of atoms. Note also that whenever we have a concentration bound like (1.5), it is possible to derive variance estimates by using that $\text{Var}_\mu(f) = \int_0^\infty \mu(|f - \mathbb{E}_\mu(f)| \geq \sqrt{t})dt$.

Next, we show similar deviation inequalities for an important class of functions, namely *self-bounded functions*. In our framework, for a difference operator Γ we say that $f \geq 0$ is a $\Gamma - (a, b)$ -self-bounded function, if

$$\Gamma(f)^2 \leq af + b$$

for some constants $a, b \geq 0$. For a product measure μ , there are various sources that provide deviation or concentration inequalities for self-bounded functions, see e.g. Boucheron et al. (2000, Theorem 2.1), Rio (2001, Théorème 3.1), Boucheron et al. (2003, Theorem 5), Boucheron et al. (2005, Corollary 1), Chatterjee (2005, Theorem 3.9), McDiarmid and Reed (2006, Theorem 1) and Boucheron et al. (2009, Theorem 1). As many of the proofs rely on the entropy method, it is not hard to adapt them to obtain Bernstein-type deviation inequalities only requiring an mLSI, which includes many more types of measures also allowing for dependencies:

Proposition 1.4. *Assume that μ satisfies a Γ -mLSI(ρ) and let $f \geq 0$ be a $\Gamma - (a, b)$ -self-bounded function. Then for all $t \geq 0$ we have*

$$\mu(f - \mathbb{E}_\mu f \geq t) \leq \exp\left(-\frac{t^2}{2\rho(2a\mathbb{E}_\mu f + 2b + \frac{1}{3}at)}\right).$$

If, additionally, $\Gamma(\lambda f) = |\lambda|\Gamma(f)$ for all $\lambda \in \mathbb{R}$, then for all $t \in [0, \mathbb{E}_\mu f]$ it holds

$$\mu(\mathbb{E}_\mu f - f \geq t) \leq \exp\left(-\frac{t^2}{2\rho(2a\mathbb{E}_\mu f + 2b + \frac{1}{3}at)}\right).$$

As we show in Proposition 2.18, product measures always satisfy an mLSI with respect to a certain L^2 -type difference operator, which was also used in the works mentioned above. This is a well-known fact and was first proven in Massart (2000).

1.2. *The symmetric group.* One example we especially discuss in this note is the symmetric group S_n equipped with the uniform measure. To this end, we need some notations. We write the group operation on S_n as $\tau\sigma$ for $\tau, \sigma \in S_n$, and denote by τ_{ij} the transposition of i and j . We define two difference operators (on $\mathcal{A} = L^\infty(\pi_n) = \mathbb{R}^{S_n}$) via

$$\begin{aligned} \Gamma(f)(\sigma)^2 &= \frac{1}{n} \sum_{i,j=1}^n (f(\sigma) - f(\sigma\tau_{ij}))^2, \\ \Gamma^+(f)(\sigma)^2 &= \frac{1}{n} \sum_{i,j=1}^n (f(\sigma) - f(\sigma\tau_{ij}))_+^2. \end{aligned}$$

For our results, we will need that the symmetric group satisfies modified logarithmic Sobolev inequalities with respect to the two difference operators defined above:

Proposition 1.5. *Let (S_n, π_n) be the symmetric group equipped with the uniform measure. Then a Γ -mLSI(1) and a Γ^+ -mLSI(2) hold.*

To formulate our next result, let us recall the notion of observable diameter. In the context of S_n equipped with any metric d , we define it by

$$\text{ObsDiam}(S_n, d) := \max_{\sigma \in S_n} n^{-1} \sum_{i,j} d(\sigma, \sigma\tau_{ij})^2.$$

For some metrics, this expression can be simplified. We say that a metric is *right invariant*, if for any $\pi, \sigma, \tau \in S_n$ we have $d(\pi, \sigma) = d(\pi\tau, \sigma\tau)$, and *left invariant* if $d(\pi, \sigma) = d(\tau\pi, \tau\sigma)$. It is *bi-invariant*, if it is right and left invariant. Assuming that d is left (or right) invariant, we have

$$\text{ObsDiam}(S_n, d) = n^{-1} \sum_{i,j} d(\text{id}, \tau_{ij})^2.$$

We call a function $f : S_n \rightarrow \mathbb{R}$ *locally Lipschitz* with respect to d , if for all $\sigma \in S_n$ and $i, j \in \{1, \dots, n\}$ we have $|f(\sigma) - f(\sigma\tau_{ij})| \leq d(\sigma, \sigma\tau_{ij})$.

Theorem 1.6. *Let (S_n, d) be the symmetric group equipped with a metric d and π_n be the uniform distribution on S_n . Assume that $f : S_n \rightarrow \mathbb{R}$ is locally Lipschitz with respect to d . For all $t \geq 0$ it holds*

$$\pi_n(|f - \mathbb{E}_{\pi_n} f| \geq t) \leq 2 \exp\left(-\frac{t^2}{2\text{ObsDiam}(S_n, d)}\right). \tag{1.6}$$

As a consequence, we have

$$\text{Var}_{\pi_n}(f) \leq 4\text{ObsDiam}(S_n, d).$$

For example, Theorem 1.6 can easily recover concentration inequalities for locally Lipschitz functions with respect to the normalized Hamming distance $d_H(\sigma, \pi) = n^{-1} \sum_{i=1}^n \mathbb{1}_{\sigma(i) \neq \pi(i)}$. In this case, $\text{ObsDiam}(S_n, d_H) \leq 4$. We work out further examples in Subsection 2.1.

Finally, we give a proof of Talagrand’s famous concentration inequality for the convex distance for random permutations by similar means as used in the proofs of the upper results. To this end, recall that for any measurable space Ω and any $\omega = (\omega_1, \dots, \omega_n) \in \Omega^n$, we may define the convex distance of ω to some measurable set $A \subset \Omega^n$ by

$$d_T(\omega, A) := \sup_{\alpha \in \mathbb{R}^n: |\alpha|_2=1} d_\alpha(\omega, A),$$

where

$$d_\alpha(\omega, A) := \inf_{\omega' \in A} d_\alpha(\omega, \omega') := \inf_{\omega' \in A} \sum_{i=1}^n |\alpha_i| \mathbb{1}_{\omega_i \neq \omega'_i}.$$

Proposition 1.7. *For any $A \subseteq S_n$ it holds*

$$\pi_n(A) \mathbb{E}_{\pi_n} \exp\left(\frac{d_T(\cdot, A)^2}{144}\right) \leq 1. \tag{1.7}$$

As compared to Talagrand’s original formulation (see Talagrand, 1995, Theorem 5.1), (1.7) has a weaker absolute constant 144 instead of 16. It is possible to improve our own constant a bit by invoking slightly more subtle estimates but we do not seem to arrive at 16. For product measures, an inequality similar to (1.7) was deduced in Talagrand (1995), a form of which with a weaker constant was proven in Boucheron et al. (2009) with the help of the entropy method. This was extended to weakly dependent random variables in Paulin (2014). However, it does not seem possible to adjust the method therein to the case of the symmetric group, and so we are not aware of any proof of either of the inequalities for the symmetric group using the entropy method. In Samson (2017) the author has proven the convex distance inequality for the symmetric group using weak transport inequalities.

It is possible to prove a weaker version of (1.7) with a somewhat better constant:

Proposition 1.8. *Let S_n be the symmetric group and π_n be the uniform distribution on S_n . For any set $A \subseteq S_n$ with $\pi_n(A) \geq 1/2$ and all $t \geq 0$ we have*

$$\pi_n(d_T(\cdot, A) \geq t) \leq 2 \exp\left(-\frac{t^2}{64}\right). \tag{1.8}$$

In fact, (1.7) implies (1.8) with a constant of 144 instead of 64.

1.3. *Slices of the hypercube.* Finally, let us discuss another model for which we are able to prove a convex distance inequality similar to (1.7). Given two natural numbers n, r such that $r \leq n$, consider the corresponding slice of the hypercube $C_{n,r} := \{\eta \in \{0, 1\}^n : \sum_i \eta_i = r\}$, and denote by $\mu_{n,r}$ the uniform measure on $C_{n,r}$. On $C_{n,r}$, we define the difference operators

$$\begin{aligned} \Gamma(f)(\eta)^2 &= \frac{2}{n} \sum_{i < j} (f(\eta) - f(\tau_{ij}\eta))^2 = \frac{2}{n} \sum_{i,j=1}^n \eta_i(1 - \eta_j)(f(\eta) - f(\tau_{ij}\eta))^2, \\ \Gamma^+(f)(\sigma)^2 &= \frac{2}{n} \sum_{i < j} (f(\eta) - f(\tau_{ij}\eta))_+^2 = \frac{2}{n} \sum_{i,j=1}^n \eta_i(1 - \eta_j)(f(\eta) - f(\tau_{ij}\eta))_+^2. \end{aligned}$$

Here, $\tau_{ij}\eta$ switches the i -th and the j -th coordinate of the configuration η . Up to the scaling of $2/n$, $\Gamma(f)^2$ is the generator of the so-called Bernoulli–Laplace model.

As in the previous section, a modified logarithmic Sobolev inequality holds:

Proposition 1.9. *For $(C_{n,r}, \mu_{n,r})$ as above, a Γ -mLSI(1) and a Γ^+ -mLSI(2) hold.*

Using this, we may establish a convex distance inequality by means of the entropy method again:

Proposition 1.10. *For any $A \subseteq C_{n,r}$ it holds*

$$\mu_{n,r}(A) \mathbb{E}_{\mu_{n,r}} \exp\left(\frac{d_T(\cdot, A)^2}{544}\right) \leq 1.$$

1.4. *Outline.* In Section 2 we provide various applications and concentration inequalities. This includes examples of functions on the symmetric group (Section 2.1), concentration inequalities for multilinear polynomials in $[0, 1]$ -valued random variables (Section 2.2), as well as consequences of Theorem 1.1 for the Euclidean sphere and measures on \mathbb{R}^n satisfying a logarithmic Sobolev inequality (Section 2.3) and for probability measures (on general spaces) satisfying an mLSI with respect to some “ L^2 difference operator” (see Section 2.4). Moreover, in Section 2.5 we recover and extend the classical Bernstein inequality for independent random variables (up to constants).

Section 3 contains all the proofs, both of the results mentioned in this section as well as in Section 2.

2. Applications

Let us now describe various situations which give rise to mLSIs with respect to “natural” difference operators, and show some consequences of the main results.

2.1. *Symmetric group.* The aim of this subsection is to show how the results from Section 1 can be used to easily obtain concentration inequalities for functions on the symmetric group. In particular, we calculate many examples of statistics for which central limit theorems were proven, and show that the variance proxy of the sub-Gaussian estimate and the true variance agree (up to a constant independent of the dimension). This provides non-asymptotic concentration results, which are consistent with the limit theorems.

First, let us introduce the following natural metrics on S_n :

$$H(\pi, \sigma) = \sum_{i=1}^n \mathbb{1}_{\pi(i) \neq \sigma(i)}$$

$$D(\pi, \sigma) = \sum_{i=1}^n |\pi(i) - \sigma(i)|$$

$$S(\pi, \sigma) = \left(\sum_{i=1}^n |\pi(i) - \sigma(i)|^2\right)^{1/2}$$

$$I(\pi, \sigma) = \min\{k \geq 0 : \exists k \text{ adjacent transpositions from } \sigma^{-1} \text{ to } \pi^{-1}\}.$$

Table 2.1 collects some basic properties of H , D , S^2 and I .

function d	invariance	mean $\mathbb{E} d(\text{id}, \cdot)$	$\text{Var}(d(\text{id}, \cdot))$	limit theorem
H	bi-invariant	$n - 1$	1	$n - H \Rightarrow \text{Poi}(1)$
D	right invariant	$\frac{n^2-1}{3}$	$\frac{(n+1)(2n^2+7)}{45}$	CLT
S^2	right invariant	$\frac{n(n^2-1)}{6}$	$\frac{n^2(n-1)(n+1)^2}{36}$	CLT
I	right invariant	$\frac{n(n-1)}{4}$	$\frac{n(n-1)(2n+5)}{72}$	CLT

TABLE 2.1. Invariance and probabilistic properties of the four functions H (Hamming distance), D (Spearman's footrule), S^2 (Spearman's rank correlation) and I (Kendall's τ). This table has been extracted from information in [Diaconis \(1988, Chapter 6\)](#).

Example 2.1. In this example, we calculate the observable diameters of the metrics on the symmetric group introduced above. By [Theorem 1.6](#), this yields concentration properties for (locally) Lipschitz functions.

- (1) For the Hamming distance H it is clear that $H(\sigma, \sigma\tau_{ij}) = 2$, which implies $\text{ObsDiam}(S_n, H) = 4(n - 1)$. So, [Theorem 1.6](#) recovers a concentration result from [Maurey \(1979\)](#).

The resulting variance estimate is not always sharp; for example, if we consider the function $H(\cdot, \text{id})$, the variance is 1 and not of order n . On the other hand, the function $G = n - H(\text{id}, \cdot)$ is a locally Lipschitz function with respect to H , which converges weakly to a Poisson random variable. As a consequence, there cannot be an n -independent sub-Gaussian estimate in the class of all locally Lipschitz functions.

- (2) If we define for $p \in [1, \infty)$ a distance d_p on S_n by the induced ℓ^p norm

$$d_p(\sigma, \pi) = \left(\sum_{k=1}^n |\sigma(k) - \pi(k)|^p \right)^{\frac{1}{p}},$$

this yields $d_p(\sigma, \sigma\tau_{ij}) = 2^{1/p} |\sigma(i) - \sigma(j)|$. Consequently, recalling that

$$\sum_{i \neq j} (\sigma(i) - \sigma(j))^2 = \frac{n^2(n^2 - 1)}{6}$$

for any $\sigma \in S_n$, we have

$$\text{ObsDiam}(S_n, d_p) = \frac{2^{2/p}}{6} n(n^2 - 1).$$

The case $p = 1$ gives Spearman's footrule and $p = 2$ Spearman's rank correlation.

- (3) Considering Kendall's τ , we can readily see that for two indices i, j and any $\sigma \in S_n$ it holds $I(\sigma, \sigma\tau_{ij}) \leq 2|\sigma(i) - \sigma(j)|$, since $\tau_{ij}\sigma^{-1}$ can be brought to σ^{-1} by first taking $\sigma^{-1}(i)$ to its place, and then $\sigma^{-1}(j)$. So, as above, this leads to

$$\text{ObsDiam}(S_n, I) \leq \frac{2}{3} n(n^2 - 1).$$

- (4) In a more general setting, let $\rho : S_n \rightarrow \text{GL}(V)$ be a faithful, unitary representation of S_n and let $\|\cdot\|$ be a unitarily invariant norm on $\text{GL}(V)$. Then

$d_\rho(\sigma, \tau) := \|\rho(\sigma) - \rho(\tau)\|$ defines a bi-invariant metric on S_n , and in this case we have

$$\text{ObsDiam}(S_n, d_\rho) = n^{-1} \sum_{i,j} \|\text{Id} - \rho(\tau_{ij})\|.$$

Example 2.2. Define the random variable $f(\sigma) = S^2(\sigma, \text{id}) = \sum_{i=1}^n (\sigma(i) - i)^2$. We have

$$\Gamma(f)^2(\sigma) = n^{-1} \sum_{i,j=1}^n (f(\sigma) - f(\sigma\tau_{ij}))^2 = 4n^{-1} \sum_{i,j=1}^n (\sigma(i) - \sigma(j))^2 (i - j)^2.$$

If we define the matrix $A(\sigma) = (a_{ij}(\sigma))_{i,j}$ via $a_{ij}(\sigma) = (\sigma(i) - \sigma(j))(i - j)$, then the right hand side is (up to the factor $4n^{-1}$) the squared Hilbert–Schmidt norm of $A(\sigma)$. It is clear that $|A(\sigma)|_{\text{HS}} = |A(\sigma^{-1})|_{\text{HS}}$, and one can also easily see that it is invariant under right multiplication with any transposition τ_{kl} . As any permutation can be written as a product of transpositions, we can evaluate it at the identity element. Consequently,

$$\Gamma(f)^2(\sigma) = 4n^{-1} \sum_{i,j=1}^n (i - j)^4 \leq \frac{4}{15} n^5.$$

Using (1.2), this leads to the concentration inequality

$$\pi_n(|f - \mathbb{E}_{\pi_n} f| \geq t) \leq 2 \exp\left(-\frac{15t^2}{8n^5}\right).$$

Actually, the term n^5 is natural, as the variance of f is of order n^5 (see the table above). Incorporating the variance of f into the inequality above leads to

$$\pi_n(|f - \mathbb{E}_{\pi_n} f| \geq \text{Var}(f)^{1/2}t) \leq 2 \exp\left(-\frac{t^2}{19.2}\right),$$

which yields the correct tail behavior.

Example 2.3. Let us consider the 1-Lipschitz function $f(\sigma) = I(\sigma, \text{id})$. For any $t \geq 0$ we have by (1.6), $\text{Var}_{\pi_n}(f) = n(n - 1)(2n + 5)/72$ and Example 2.1 (3)

$$\pi_n(|f - \mathbb{E}_{\pi_n} f| \geq \text{Var}_{\pi_n}(f)^{1/2}t) \leq 2 \exp\left(-\frac{t^2}{48}\right),$$

which is consistent with the central limit theorem for f .

Example 2.4. We define the number of ascents $f(\sigma) = \sum_{j=1}^{n-1} \mathbb{1}_{\sigma(j+1) > \sigma(j)}$. It can be easily shown that for any $i \neq j$ the number of ascents is not sensitive to transpositions in the sense that $|f(\sigma) - f(\sigma\tau_{ij})| \leq 2$. Consequently, this leads to $\Gamma(f)^2 \leq 4(n - 1)$, implying the concentration inequality

$$\pi_n(|f - \mathbb{E}_{\pi_n} f| \geq t) \leq 2 \exp\left(-\frac{t^2}{8(n - 1)}\right),$$

again using (1.2). Alternatively, this also follows from Example 2.1 (1). Again, the variance term of order \sqrt{n} is of the right order, as in Carlitz et al. (1972) the authors have shown a central limit theorem for the number of ascents. More precisely, the sequence $g_n = (f - \mathbb{E}_{\pi_n} f)/(\sqrt{(n + 1)/12})$ converges to a standard normal distribution. The above calculations lead to

$$\pi_n(|g_n| \geq t) \leq 2 \exp\left(-\frac{t^2}{96}\right).$$

Example 2.5. A closely related statistic is given by the sum of the ascents defined as $f(\sigma) = \sum_{j=1}^{n-1} (\sigma_{i+1} - \sigma_i)_+$. A short calculation shows

$$\Gamma(f)^2 = n^{-1} \sum_{i \neq j} (f(\sigma) - f(\sigma\tau_{ij}))^2 \leq 4(n-1)^2 n^{-1} \sum_{i \neq j} 1 = 4(n-1)^3.$$

Indeed, if we let $\Delta_{i,j} := (\sigma(i) - \sigma(j))_+$, then

$$\begin{aligned} & (f(\sigma) - f(\sigma\tau_{ij}))^2 \\ &= (\Delta_{i,i-1} + \Delta_{i+1,i} + \Delta_{j,j-1} + \Delta_{j+1,j} - \Delta_{j,i-1} - \Delta_{i+1,j} - \Delta_{i,j-1} - \Delta_{j+1,i})^2 \\ &\leq \max\left(\Delta_{i,i-1} + \Delta_{i+1,i} + \Delta_{j,j-1} + \Delta_{j+1,j}, \Delta_{j,i-1} + \Delta_{i+1,j} + \Delta_{i,j-1} + \Delta_{j+1,i}\right)^2. \end{aligned}$$

Now each of the terms $\Delta_{i,i-1} + \Delta_{i+1,i}$, $\Delta_{j,j-1} + \Delta_{j+1,j}$ is less than $(n-1)$, and the same holds true for the two other sums. Therefore this yields

$$\pi_n(|f - \mathbb{E}_{\pi_n} f| \geq t) \leq 2 \exp\left(-\frac{t^2}{8(n-1)^3}\right).$$

Clark (2009) has calculated the variance of the sum of ascents, and it is of order n^3 , which is in good accordance with the concentration inequality (again, up to the factor).

Example 2.6. Given a matrix $a = (a_{ij})$ of real numbers satisfying $a_{ij} \in [0, 1]$, define $f(\sigma) = \sum_{i=1}^n a_{i,\sigma(i)}$. By elementary computations one can show $\Gamma(f)^2 \leq 4f + 4\mathbb{E}_{\pi_n} f$, i. e. f is self-bounding. As a consequence, Proposition 1.4 leads to

$$\pi_n(|f - \mathbb{E}_{\pi_n} f| \geq t) \leq 2 \exp\left(-\frac{t^2}{32\mathbb{E}_{\pi_n} f + 8t/3}\right).$$

Concentration inequalities for f have been proven using the exchangeable pair approach in Chatterjee (2005, Proposition 3.10) (see also Chatterjee, 2007, Theorem 1.1), with the denominator being $4\mathbb{E}_{\pi_n} f + 2t$.

For example, if a is the identity matrix, f is the number of fixed points of a random permutation, which satisfies $\mathbb{E}_{\pi_n} f = 1$ for all $n \in \mathbb{N}$. In this case, f converges to a Poisson distribution with mean 1 as $n \rightarrow \infty$ (see e.g. Diaconis, 1988).

Example 2.7. Finally, consider the random variable $f(\sigma) = g(\sigma) + g(\sigma^{-1})$, where $g(\sigma) = \sum_{i=1}^{n-1} \mathbb{1}_{\sigma(i+1) > \sigma(i)}$ is the number of descents. In Chatterjee and Diaconis (2017) the authors calculated the expectation and variance of f and proved a central limit theorem. As in the above example one can easily see that $\Gamma(g)^2 \leq 4(n-1)$, as well as $\Gamma(g \circ \text{inv})^2 \leq 4(n-1)$, where $\text{inv} : S_n \rightarrow S_n$ denotes the inverse map. Since $\Gamma(h_1 + h_2)^2 \leq 2\Gamma(h_1)^2 + 2\Gamma(h_2)^2$ holds true for any functions h_1, h_2 , we also have $\Gamma(f)^2 \leq 16(n-1)$, implying for any $t \geq 0$

$$\pi_n(|f - \mathbb{E} f| \geq t) \leq 2 \exp\left(-\frac{t^2}{32(n-1)}\right).$$

Again, the variance is of order \sqrt{n} , so that it is consistent with the CLT.

2.2. *Multilinear polynomials in $[0, 1]$ -random variables.* The aim of this section is to show Bernstein-type concentration inequalities for a class of polynomials in independent random variables with values in $[0, 1]$. The functions we consider are constructed as follows: Let $H = (V, E, (w_e)_{e \in E})$ be a weighted hypergraph, such that every $e \in E$ consists of at most k vertices, assume that $(X_v)_{v \in V}$ are independent, $[0, 1]$ -valued random variables, and set

$$f(X) = f((X_v)_{v \in V}) = \sum_{e \in E} w_e \prod_{v \in e} X_v = \sum_{e \in E} w_e X_e. \tag{2.1}$$

Define the maximum first order partial derivative $\text{ML}(f)$ as

$$\text{ML}(f) := \sup_{v \in V} \sup_{x \in [0, 1]^V} \partial_v f(x). \tag{2.2}$$

Proposition 2.8. *Let $(X_v)_{v \in V}$ be independent, $[0, 1]$ -valued random variables and $f : [0, 1]^V \rightarrow \mathbb{R}$ given as in (2.1). Assume that $w_e \geq 0$ and $|e| \leq k$ for all $e \in E$. We have for any $t \geq 0$*

$$\mathbb{P}(f(X) - \mathbb{E} f(X) \geq t) \leq \exp\left(-\frac{t^2}{2k\text{ML}(f)(\mathbb{E} f(X) + t/2)}\right). \tag{2.3}$$

Furthermore, for $t \in [0, \mathbb{E} f]$ it holds

$$\mathbb{P}(\mathbb{E} f(X) - f(X) \geq t) \leq \exp\left(-\frac{t^2}{2k\text{ML}(f)}\right).$$

A slight modification of the proof of Proposition 2.8 also allows to prove deviation inequalities for suprema of such homogeneous polynomials. For example, this can be used to prove the following concentration inequalities for maxima or l^p norms of linear forms.

Proposition 2.9. *Let $(X_v)_{v \in V}$ be independent, $[0, 1]$ -valued random variables, $\mathcal{F} \subset \{a \in \mathbb{R}^V : a_i \in [0, 1]^n\}$ and define $f_{\mathcal{F}}(X) := \sup_{a \in \mathcal{F}} \sum_{i \in V} a_i X_i$. For any $t \geq 0$ we have*

$$\mathbb{P}(f_{\mathcal{F}}(X) - \mathbb{E} f_{\mathcal{F}}(X) \geq t) \leq \exp\left(-\frac{t^2}{2 \sup_{a \in \mathcal{F}} \|a\|_{\infty} (\mathbb{E} f_{\mathcal{F}}(X) + t/2)}\right).$$

In particular, for any $p \in [1, \infty]$ it holds

$$\mathbb{P}(\|X\|_p - \mathbb{E}\|X\|_p \geq t) \leq \exp\left(-\frac{t^2}{2(\mathbb{E}\|X\|_p + t/2)}\right). \tag{2.4}$$

Note that we may also derive inequalities similar to (2.4) by convex concentration as first established in Talagrand (1988). Indeed, it is easy to see that the function $x \mapsto \|x\|_p$ has Lipschitz seminorm $\max(1, |V|^{1/p-1/2})$ (for $p < 2$, this follows from Hölder’s inequality for the uniform measure on V , while the case $p \geq 2$ is trivial). Therefore, applying Ledoux (1995/97, Corollary 1.3), we arrive at the bound

$$\mathbb{P}(\|X\|_p - \mathbb{E}\|X\|_p \geq t) \leq \exp\left(-\frac{t^2}{2 \max(1, |V|^{2/p-1})}\right),$$

which is weaker than (2.4) if $0 \leq t \leq 2(\max(1, |V|^{2/p-1}) - \mathbb{E}\|X\|_p)$.

One possible application of Proposition 2.8 is to understand the finite n concentration properties of the so-called *d-runs on the line*.

Proposition 2.10. *Let $d \in \mathbb{N}, n > d$, $(X_i)_{i=1, \dots, n}$ be independent, identically distributed random variables with values in $[0, 1]$ and mean $\eta := \mathbb{E} X_1 > 0$. Define the random variable $f_d(X) := \sum_{i=1}^n X_i \cdots X_{i+d-1}$, where the indices are to be understood modulo n . For any $t \geq 0$ it holds*

$$\mathbb{P} \left(f_d(X) - \mathbb{E} f_d(X) \geq \sqrt{n\eta^d t} \right) \leq 2 \exp \left(- \frac{t^2}{2d^2(1 + t/\sqrt{n\eta^d})} \right). \tag{2.5}$$

In Reinert and Röllin (2009, Theorem 4.1), the authors prove a CLT for the d -runs on the line for Bernoulli random variables X_i with success probability p , by normalizing f by $n^{-1/2}p^{-d/2}$. This is also the reason for the choice $n^{1/2}\eta^{d/2}t$ in inequality (2.5). In other words, under the assumption $n\eta^d \rightarrow \infty$ as $n \rightarrow \infty$, Proposition 2.10 yields sub-Gaussian tails for $n^{-1/2}\eta^{-d/2}f$. This is in good accordance with the aforementioned CLT.

Moreover, note that in this example, our methodology leads to better results than the usual bounded difference inequality. Indeed, the latter only yields

$$\mathbb{P}(|f_d(X) - \mathbb{E} f_d(X)| \geq t) \leq 2 \exp \left(- \frac{2t^2}{nd^2} \right),$$

suggesting an (inaccurate) normalization of $f(X)$ by $n^{-1/2}$.

Example 2.11. If $(X_v)_{v \in E(K_n)}$ is the Erdős–Rényi model with parameter p , for any fixed graph H with $|V|$ vertices and $|E|$ edges, the subgraph counting statistic T_H can be written in the form (2.1) with $w_e = 1$, and $k = |E|$. Furthermore, it is easy to see that $\text{ML}(f) \leq n^{\Delta-1}$ for the maximum degree Δ , so that Proposition 2.8 yields

$$\mathbb{P}(T_H(X) - \mathbb{E} T_H(X) \geq \varepsilon \mathbb{E} T_H(X)) \leq \exp \left(- C_{k,\varepsilon} n^{|V|-\Delta+1} p^{|E|} \right).$$

For example, this gives nontrivial bounds in the triangle case whenever $n^2 p^3 \rightarrow \infty$ as $n \rightarrow \infty$. This bound is suboptimal, as the optimal decay is known to be $np \rightarrow \infty$, see Chatterjee (2012); DeMarco and Kahn (2012). However, it is better than the bound obtained by the bounded differences inequality. In general, if we consider subgraph counting statistics for some subgraph H with v vertices and e edges on an Erdős–Rényi model $(X_v)_{v \in E(K_n)}$, the bounded difference inequality yields the estimate

$$\mathbb{P}(f(X) - \mathbb{E} f(X) \geq \varepsilon \mathbb{E} f(X)) \leq \exp \left(- C_{\varepsilon,H} \frac{n^{2|V|} p^{2|E|}}{n^2 n^{2\Delta-2}} \right).$$

Thus, to obtain non-trivial estimates in the limit $n \rightarrow \infty$, one has to assume that $n^{|V|-\Delta} p^{|E|} \rightarrow \infty$. With the above inequality, this can be weakened to $n^{|V|-\Delta+1} p^{|E|} \rightarrow \infty$.

2.3. Derivations. If Γ satisfies the chain rule, i. e. for all differentiable $u : \mathbb{R} \rightarrow \mathbb{R}$ and $f \in \mathcal{A}$ such that $u \circ f \in \mathcal{A}$ we have $\Gamma(u \circ f) = |u' \circ f| \Gamma(f)$, then (1.1) is equivalent to the usual logarithmic Sobolev inequality (in short: Γ -LSI(ρ))

$$\text{Ent}_\mu(f^2) \leq 2\rho \mathbb{E}_\mu \Gamma(f)^2.$$

Using this, one can derive second order concentration inequalities similar to the ones given in Bobkov et al. (2017) from Proposition 1.3. Let $S^{n-1} := \{x \in \mathbb{R}^n : |x| = 1\}$

be the unit sphere equipped with the uniform measure σ_{n-1} . It is known that for $\rho_n := (n - 1)^{-1}$

$$\text{Ent}_{\sigma_{n-1}}(e^f) \leq \frac{\rho_n}{2} \mathbb{E}_{\sigma_{n-1}} |\nabla_S f|^2 e^f \tag{2.6}$$

holds for all Lipschitz functions f and the spherical gradient $\nabla_S f$ (see [Bobkov et al. \(2017, Formula \(3.1\)\)](#) for the logarithmic Sobolev inequality, from which the modified one follows as above). To state our next result, we introduce the following notation (which we will stick to for the rest of this paper): if A is an $n \times n$ matrix, we denote by $\|A\|_{\text{HS}}$ its Hilbert–Schmidt and by $\|A\|_{\text{op}}$ its operator norm.

Proposition 2.12. *Consider S^{n-1} equipped with the uniform measure σ_{n-1} and let $f : S^{n-1} \rightarrow \mathbb{R}$ be a C^2 function satisfying $\sup_{\theta \in S^{n-1}} \|f''_S(\theta)\|_{\text{op}} \leq 1$. For any $t \geq 0$*

$$\sigma_{n-1}(|f - \mathbb{E}_{\sigma_{n-1}} f| \geq t) \leq 2 \exp\left(-\frac{1}{4\rho_n} \min\left(\frac{t^2}{2\mathbb{E}_{\sigma_{n-1}} |\nabla_S f|^2}, t\right)\right).$$

This follows immediately from Proposition 1.3 and the inequality $|\nabla_S |\nabla_S f|| \leq \|f''_S\|_{\text{op}}$ proven in [Bobkov et al. \(2017, Lemma 3.1\)](#). Now, if f is C^2 and orthogonal to all affine functions (in $L^2(\sigma_{n-1})$), [Bobkov et al. \(2017, Proposition 5.1\)](#) shows $\mathbb{E}_{\sigma_{n-1}} |\nabla_S f|^2 \leq \rho_n \mathbb{E}_{\sigma_{n-1}} \|f''_S\|_{\text{HS}}^2$. So, if we additionally have $\mathbb{E}_{\sigma_{n-1}} \|f''\|_{\text{HS}}^2 \leq b^2$, the estimate

$$\sigma_{n-1}((n - 1)|f - \mathbb{E}_{\sigma_{n-1}} f| \geq t) \leq 2 \exp\left(-\frac{1}{4} \min\left(\frac{t^2}{2b^2}, t\right)\right) \tag{2.7}$$

follows.

In a similar manner, one may address open subsets of \mathbb{R}^n equipped with some probability measure μ satisfying a logarithmic Sobolev inequality (with respect to the usual gradient ∇). This situation has been sketched in [Bobkov et al. \(2017, Remark 5.3\)](#) and was discussed in more detail in [Götze and Sambale \(2020\)](#). Here we easily obtain the following result:

Proposition 2.13. *Let $G \subseteq \mathbb{R}^n$ be an open set, equipped with a probability measure μ which satisfies a ∇ -LSI(ρ), and let $f : G \rightarrow \mathbb{R}$ be a C^2 function satisfying $\sup_{x \in G} \|f''(x)\|_{\text{op}} \leq 1$. For any $t \geq 0$*

$$\mu(|f - \mathbb{E}_\mu f| \geq t) \leq 2 \exp\left(-\frac{1}{4\rho} \min\left(\frac{t^2}{2\mathbb{E}_\mu |\nabla f|^2}, t\right)\right).$$

For the proof it only remains to note that $|\nabla |\nabla f|| \leq \|f''\|_{\text{op}}$, cf. [Götze and Sambale \(2020, Lemma 7.2\)](#). As above, if we require the first order partial derivatives $\partial_i f$ to be centered (which translates into orthogonality to linear functions if μ is the standard Gaussian measure, for instance), a simple application of the Poincaré inequality yields $\mathbb{E}_\mu |\nabla f|^2 \leq \rho \mathbb{E}_\mu \|f''\|_{\text{HS}}^2$. In particular, we have the following corollary which immediately follows from Proposition 1.3 and the Poincaré inequality.

Corollary 2.14. *Let $G \subseteq \mathbb{R}^n$ be an open set, equipped with a probability measure μ satisfying a ∇ -LSI(ρ), and $f : G \rightarrow \mathbb{R}$ be a C^2 function with*

$$\sup_{x \in \text{supp}(\mu)} \|f''(x)\|_{\text{op}} \leq b \quad \text{and} \quad \int \|f''(x)\|_{\text{HS}}^2 d\mu(x) \leq a^2.$$

For any $t \geq 0$ we have

$$\mu(|f(x) - \mathbb{E}_\mu f(x) - \langle x - \mathbb{E}_\mu(x), \mathbb{E}_\mu \nabla f(x) \rangle| \geq t) \leq 2 \exp\left(-\frac{1}{4} \min\left(\frac{t^2}{2\rho^2 a^2}, \frac{t}{\rho b}\right)\right).$$

Thus, if we recenter a function and its derivatives, the two conditions on the Hessian ensure two-level concentration inequalities. For functions $f(X, Y)$ of independent Gaussian vectors, two-level concentration inequalities have been studied in Wolff (2013) using the Hoeffding decomposition instead of a recentering of the partial derivatives.

Note that (2.7) and Corollary 2.14 do not only recover Bobkov et al. (2017, Theorem 1.1) and Götze and Sambale (2020, Theorem 1.4), but even strengthen these results by providing two-level bounds. To illustrate this, we discuss one of the examples from Götze and Sambale (2020) in more detail.

Example 2.15 (Eigenvalues of Wigner matrices). Let $\{\xi_{jk}, 1 \leq j \leq k \leq N\}$ be a family of independent real-valued random variables whose distributions all satisfy a ∇ -LSI(ρ) for a fixed $\rho > 0$. Putting $\xi_{jk} = \xi_{kj}$ for $1 \leq k < j \leq N$, we define the random matrix $\Xi = (\xi_{jk}/\sqrt{N})_{1 \leq j, k \leq N}$. Then, by a simple argument using the Hoffman–Wielandt theorem, the joint distribution $\mu^{(N)} = \mu$ of its ordered eigenvalues $\lambda_1 \leq \dots \leq \lambda_N$ on \mathbb{R}^N (in fact, $\lambda_1 < \dots < \lambda_N$ a.s.) satisfies a ∇ -LSI(ρ_N) with constant $\rho_N = 2\rho/N$ (see for instance Anderson et al., 2010, Theorem 2.3.5).

Now consider a \mathcal{C}^2 -smooth function $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ with first order (partial) derivatives in $L^1(\mu)$ and second order derivatives bounded by some constant γ . Considering a quadratic statistic $\sum_{j \neq k} g(\lambda_j, \lambda_k)$ and recentering according to Corollary 2.14, we shall study

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

where g_x, g_y denote partial derivatives. For instance, if $g(x, y) := xy$, we have $Q_N = \sum_{j \neq k} (\lambda_j - \mu[\lambda_j])(\lambda_k - \mu[\lambda_k])$. Simple calculations show that $\|Q_N''\|_{\text{op}} \leq c\gamma N$ as well as $\int \|Q_N\|_{\text{HS}}^2 d\mu \leq c\gamma^2 N^3$. Here, by $c > 0$ we denote suitable absolute constants which may vary from line to line. Following Götze and Sambale (2020, Proposition 8.5), this leads to the exponential moment bound

$$\int \exp\left(\frac{c}{\rho\gamma N^{1/2}} |Q_N|\right) d\mu \leq 2.$$

By Chebyshev’s inequality, $\mu(|Q_N| \geq t) \leq 2 \exp(-ct/(\rho\gamma N^{1/2}))$ for all $t \geq 0$, thus yielding subexponential fluctuations of order $\mathcal{O}_P(N^{1/2})$.

By contrast, Corollary 2.14 leads to

$$\mu(|Q_N| \geq t) \leq 2 \exp\left(-c \min\left(\frac{t^2}{\rho^2 \gamma^2 N}, \frac{t}{\rho \gamma}\right)\right),$$

which is much better for large t . In particular, the fluctuations in the subexponential regime are of order $\mathcal{O}_P(1)$ now. This can be interpreted as an extension of the self-normalizing property of linear eigenvalue statistics to a second order situation

on the level of fluctuations (cf. the discussion of [Götze and Sambale, 2020](#), Proposition 8.5). Note that in [Götze and Sambale \(2020\)](#), a comparable result could be achieved for the special case of $g(x, y) := xy$ only.

2.4. Weakly dependent measures. To continue the discussion of the previous section for a larger class of measures, we will now consider applications of Theorem 1.1 for functions of weakly dependent random variables (which, in our case, essentially means that a certain mLSI with respect to a suitable difference operator is satisfied). Throughout this section, we shall consider probability measures μ on a product of Polish spaces $\mathcal{X} = \otimes_{i=1}^n \mathcal{X}_i$. For a vector $x = (x_i)_{i \in I}$ and $j \in I$ we let $x_{i^c} = (x_j)_{j \in I \setminus \{i\}}$, and for $y \in \mathbb{R}$ we write $y_+ = \max(y, 0)$. Now we define difference operators on $L^\infty(\mu)$ via

$$\begin{aligned} |\mathfrak{d}f|(x) &= \left(\sum_{i=1}^n \int (f(x) - f(x_{i^c}, x'_i))^2 d\mu(x'_i | x_{i^c}) \right)^{1/2}, \\ |\mathfrak{d}^+f|(x) &= \left(\sum_{i=1}^n \int (f(x) - f(x_{i^c}, x'_i))_+^2 d\mu(x'_i | x_{i^c}) \right)^{1/2}, \\ |\mathfrak{h}f|(x) &= \left(\sum_{i=1}^n \sup_{x_i, x'_i} (f(x) - f(x_{i^c}, x'_i))^2 \right)^{1/2}, \\ |\mathfrak{h}^+f|(x) &= \left(\sum_{i=1}^n \sup_{x'_i} (f(x) - f(x_{i^c}, x'_i))_+^2 \right)^{1/2}. \end{aligned}$$

Here, the suprema over x'_i (and x_i) are to be understood with respect to the support of μ . Clearly, $|\mathfrak{d}f| \leq |\mathfrak{h}f|$ and $|\mathfrak{d}^+f| \leq |\mathfrak{h}^+f|$. Moreover, we need a second order version of the difference operator \mathfrak{h} . To this end, for any $i \neq j$, define

$$\mathfrak{h}_{ij}f(x) = \sup_{x_i, x'_i, x_j, x'_j} |f(x) - f(x_{i^c}, x'_i) - f(x_{j^c}, x'_j) - f(x_{\{i,j\}^c}, x'_i, x'_j)|$$

and let $\mathfrak{h}^{(2)}f(x)$ be the matrix (“Hessian”) with zero diagonal and entries $\mathfrak{h}_{ij}f(x)$ on the off-diagonal.

We now have the following second order result in presence of a \mathfrak{d} -mLSI:

Proposition 2.16. *Let μ be a probability measure on a product of Polish spaces $\mathcal{X} = \otimes_{i=1}^n \mathcal{X}_i$ satisfying a \mathfrak{d} -mLSI(σ^2), and let $f : \mathcal{X} \rightarrow \mathbb{R}$ be a bounded measurable function. If $|\mathfrak{d}^+|\mathfrak{d}f| \leq b$, we have for any $t \geq 0$*

$$\mu(|f - \mathbb{E}_\mu f| \geq t) \leq 2 \exp \left(- \frac{1}{12\sigma^2} \min \left(\frac{t^2}{(\mathbb{E}_\mu |\mathfrak{d}f|)^2}, \frac{t}{2b} \right) \right). \tag{2.8}$$

On the other hand, if $\|\mathfrak{h}^{(2)}f\|_{\text{op}} \leq b$ for all $x \in \mathcal{X}$, we have for all $t \geq 0$

$$\mu(|f - \mathbb{E}_\mu f| \geq t) \leq 2 \exp \left(- \frac{1}{12\sigma^2} \min \left(\frac{t^2}{(\mathbb{E}_\mu |\mathfrak{h}f|)^2}, \frac{t}{b} \right) \right). \tag{2.9}$$

Proposition 2.16 implies many second order results from previous articles. For instance, it is well-known (and we will check again below) that any product probability measure μ satisfies a \mathfrak{d} -mLSI(1). Therefore, from (2.8) it is easily possible to obtain results similar to [Götze and Sambale \(2020, Theorem 1.2\)](#). To see this, it suffices to note that for functions with Hoeffding decomposition $f = \sum_{k=2}^n f_k$, one may apply [Götze and Sambale \(2020, Proposition 5.2\)](#) to upper bound $\mathbb{E}_\mu |\mathfrak{d}f|^2$

by $\mathbb{E}_\mu \|\mathfrak{d}^{(2)} f\|_{\text{HS}}^2$. Unlike in [Götze and Sambale \(2020\)](#), Proposition 2.16 yields two-level (or Bernstein-type) inequalities, which can be regarded as an advantage of the present approach.

Similarly, we may retrieve (and sharpen) some of the results from further articles like e.g. [Götze et al. \(2019\)](#) for $d = 2$. On the other hand, it seems that requiring modified logarithmic Sobolev inequalities instead of usual logarithmic Sobolev inequalities extends the class of measures to which our results apply, in particular in non-independent situations. We will discuss the \mathfrak{d} -mLSI property and provide some sufficient conditions in more detail below.

For some classes of functions, we can obtain variants of Proposition 2.16 which are especially adapted to the properties of the functions under consideration. In particular, we may show deviation inequalities for suprema of quadratic forms in the spirit of [Klochkov and Zhivotovskiy \(2020\)](#) for the weakly dependent case.

Proposition 2.17. *Let μ be supported in $[-1, +1]^n$ and satisfy a \mathfrak{d} -mLSI(σ^2). Let \mathcal{B} be a countable class of symmetric matrices, bounded in operator norm and with zeroes on its diagonal. Define $h(x) := \sup_{B \in \mathcal{B}} \langle x, Bx \rangle$, $f_{\mathcal{B}}(x) := \sup_{B \in \mathcal{B}} \|Bx\|$ and $\Sigma := \sup_{B \in \mathcal{B}} \|B\|_{\text{op}}$. We have for any $t > 0$*

$$\mu(h - \mathbb{E}_\mu h \geq t) \leq \frac{4}{3} \exp\left(-\frac{1}{128\sigma^2} \min\left(\frac{t^2}{2(\mathbb{E}_\mu f_{\mathcal{B}})^2}, \frac{t}{\Sigma}\right)\right). \tag{2.10}$$

Note that while in general, we only obtain deviation inequalities here, for a single symmetric matrix B with zeroes on its diagonal and the quadratic form $f(x) = \langle x, Bx \rangle$ similar arguments as in the proof of Proposition 2.16 do lead to concentration inequalities for f .

If μ is a product measure, the result of Proposition 2.17 is well-known and has been proven various times, see for example [Talagrand \(1996, Theorem 1.2\)](#) for concentration inequalities in Rademacher random variables, [Ledoux \(1995/97, Theorem 3.1\)](#) for the upper tail inequalities and random variables satisfying $|X_i| \leq 1$, [Boucheron et al. \(2003, Theorem 17\)](#) for the upper bound and Rademacher random variables and [Boucheron et al. \(2005, Corollary 4\)](#). More recent results include [Hsu et al. \(2012\)](#); [Rudelson and Vershynin \(2013\)](#); [Adamczak \(2015\)](#); [Adamczak et al. \(2019\)](#); [Klochkov and Zhivotovskiy \(2020\)](#); [Götze et al. \(2020\)](#).

To understand which classes of measures may be addressed by Propositions 2.16 and 2.17, let us study the \mathfrak{d} -mLSI property in more detail. First, we show that it is implied by another functional inequality. Assume that a probability measure μ on a product of Polish spaces $\mathcal{X} = \otimes_{i=1}^n \mathcal{X}_i$ satisfies

$$\text{Ent}_\mu(e^f) \leq \sigma^2 \sum_{i=1}^n \int \text{Cov}_{\mu(\cdot | x_{i^c})}(f(x_{i^c}, \cdot), e^{f(x_{i^c}, \cdot)}) d\mu(x), \tag{2.11}$$

where $\mu(\cdot | x_{i^c})$ denotes the regular conditional probability. This functional inequality is (also) known as a modified logarithmic Sobolev inequality in the framework of Markov processes, and it is equivalent to exponential decay of the relative entropy along the Glauber semigroup, see for example [Bobkov and Tetali \(2006\)](#) or [Caputo et al. \(2015\)](#).

Proposition 2.18. *If μ satisfies (2.11), then a \mathfrak{d} -mLSI(σ^2) and a \mathfrak{d}^+ -mLSI($2\sigma^2$) hold. Consequently, for any $f : \mathcal{X} \rightarrow \mathbb{R}$ and any $\alpha > \sigma^2/2$ we have*

$$\mathbb{E}_\mu \exp\left(f - \mathbb{E}_\mu f\right) \leq \left(\mathbb{E}_\mu \exp\left(\alpha |\mathfrak{d}f|^2\right)\right)^{\frac{\sigma^2}{2\alpha - \sigma^2}}. \tag{2.12}$$

The same is true for \mathfrak{d}^+ with σ^2 replaced by $2\sigma^2$. This especially holds for product measures $\mu = \mu_1 \otimes \dots \otimes \mu_n$ with $\sigma^2 = 1$.

Here, choosing $\alpha = \sigma^2$ or $\alpha = 2\sigma^2$ respectively leads to the exponential inequalities

$$\mathbb{E}_\mu \exp(f) \leq \mathbb{E}_\mu \exp\left(\sigma^2 |\mathfrak{d}f|^2\right) \quad \text{and} \quad \mathbb{E}_\mu \exp(f) \leq \mathbb{E}_\mu \exp\left(2\sigma^2 |\mathfrak{d}^+ f|^2\right).$$

The inequality involving $|\mathfrak{d}f|^2$ is well-known in the case of the discrete cube, cf. [Bobkov and Götze \(1999, Corollary 2.4\)](#) with a better constant. On the other hand, the proof presented herein is remarkably short and does not rely on some special properties of the measure μ , but can be derived under [\(2.11\)](#).

[Proposition 2.18](#) implies [Boucheron et al. \(2003, Theorem 2\)](#), as product measures satisfy [\(2.11\)](#) with $\sigma^2 = 1$. Indeed, taking the logarithms on both sides of [\(2.12\)](#) gives for any $\alpha > 1$ and $\lambda \geq 0$

$$\log \mathbb{E}_\mu \exp\left(\lambda(f - \mathbb{E}_\mu f)\right) \leq \frac{1}{\alpha - 1} \log \mathbb{E}_\mu \exp\left(\lambda^2 \alpha |\mathfrak{d}^+ f|^2\right).$$

It remains to choose some fixed $\theta > 0$ and set $\alpha = (\lambda\theta)^{-1}$.

The property [\(2.11\)](#) is satisfied for a large class containing non-product measures. Note that a sufficient condition (due to Jensen’s inequality) for [\(2.11\)](#) is the *approximate tensorization property*

$$\text{Ent}_\mu(e^f) \leq \sigma^2 \sum_{i=1}^n \int \text{Ent}_{\mu(\cdot|x_{i^c})}(e^{f(x_i, \cdot)}) d\mu(x). \tag{2.13}$$

Establishing [\(2.13\)](#) is subject to ongoing research, and we especially want to highlight two possible approaches.

The first one is akin to the perturbation argument of Holley and Stroock as outlined in [Holley and Stroock \(1987\)](#) (see also [Royer, 2007, Proposition 3.1.18](#) for a similar reasoning). Assume that $d\mu = Z^{-1}e^f d\nu$, where $f : \mathcal{X} \rightarrow \mathbb{R}$ is a measurable function, $\nu = \otimes_{i=1}^n \nu_i$ is *some* product measure and $Z = \mathbb{E}_\nu e^f$. If we require f to be bounded, we clearly have $\text{osc}(f) < \infty$ for its (maximal) oscillation $\text{osc}(f) = \sup_{x \in \mathcal{X}} f(x) - \inf_{x \in \mathcal{X}} f(x)$. Under these assumptions, μ satisfies [\(2.13\)](#) with $\sigma^2 = \exp(2\text{osc}(f))$.

Furthermore, under weak dependence conditions on the local specifications of some measure μ on a product space \mathcal{X} , [\(2.13\)](#) was proven in [Marton \(2013, 2019\)](#); [Caputo et al. \(2015\)](#).

2.5. Bernstein inequality. As a final application, let us demonstrate how to recover the classical Bernstein inequality for independent bounded random variables by means of [Theorem 1.1](#) (up to constants). In fact, as in some previous works we may remove the boundedness assumption.

There are various extensions of Bernstein’s inequality to unbounded random variables. For instance, [Adamczak \(2008, Theorem 4\)](#) proves deviation inequalities for empirical processes in independent random variables with finite Ψ_α norm for some $\alpha \in (0, 1]$, which in particular includes concentration inequalities for sums of random variables with finite Ψ_α norm. Moreover, [Boucheron et al. \(2013, Theorem 2.10\)](#) requires a certain control of the moments of the random variables, which is in essence a condition on the Ψ_1 norms. Thirdly, [Vershynin \(2018, Theorem 2.8.1\)](#)

provides a Bernstein inequality for random variables with bounded Ψ_1 norms. However, note that the Gaussian term in the last two mentioned works is a sum of the Ψ_1 norm instead of the variance. By our methods, we obtain a version of Bernstein’s inequality for sub-Gaussian random variables with the variance of the sum in the Gaussian term, with a reasonable constant.

Theorem 2.19. *There exists an absolute constant $c' > 0$ such that the following holds. For any set of independent random variables X_1, \dots, X_n satisfying $\|X_i\|_{\Psi_2} < \infty$, we have for any $t \geq 0$*

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i - \mathbb{E} X_i\right| \geq t\right) \leq 2 \exp\left(-\min\left(\frac{t^2}{80\text{Var}(\sum_i X_i)}, \frac{t}{c'\|\max_i |X_i|\|_{\Psi_2}}\right)\right). \tag{2.14}$$

In particular, if $|X_i| \leq M$ almost surely for all $i \in \{1, \dots, n\}$ and some $M > 0$, then for all $t \geq 0$ it holds

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i - \mathbb{E} X_i\right| \geq t\right) \leq 2 \exp\left(-\min\left(\frac{t^2}{80\text{Var}(\sum_i X_i)}, \frac{t}{c'M}\right)\right).$$

We want to give three concluding remarks on Theorem 2.19. Firstly, note that is not possible to prove an inequality

$$\mathbb{P}\left(\sum_{i=1}^n X_i - \mathbb{E} X_i \geq t\right) \leq 2 \exp\left(-c \frac{t^2}{\text{Var}(\sum_i X_i)}\right)$$

for some absolute constant $c > 0$ in the class of all sub-Gaussian random variables. This can be easily seen in the case $n = 1$ and by choosing $X \sim \text{Ber}(p)$ for $p \rightarrow 0$. Thus, to obtain a sub-Gaussian tail with the variance parameter, one has to limit the range of t for which one can expect sub-Gaussian behaviour.

Secondly, one cannot replace $\|\max_i |X_i|\|_{\Psi_2}$ by $\max_i \|X_i\|_{\Psi_2}$ in (2.14), i. e. there cannot be an inequality of the form

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i - \mathbb{E} X_i\right| \geq t\right) \leq 2 \exp\left(-c \min\left(\frac{t^2}{\text{Var}(\sum_i X_i)}, \frac{t}{\max_i \|X_i\|_{\Psi_2}}\right)\right).$$

This, again, follows by choosing $X_i \sim \text{Ber}(p)$ for $p = \lambda/n$, $\lambda > 0$. In this case, the sum converges (weakly) to a Poisson random variable, whereas the sub-Gaussian range extends to $t \in \mathbb{R}_+$ for $n \rightarrow \infty$, giving a contradiction.

Thirdly, it is well known that the Ψ_2 norm of the maximum of Ψ_2 random variables (bounded by some constant, say K) grows at most logarithmically in the dimension. For example, if we consider i. i. d. random variables X_i with unit variance, we have the sub-Gaussian estimate for t of order (at least) $n/\log(n)$.

3. Proofs and auxiliary results

We begin by proving Theorem 1.1. Before we start, let us recall [Bobkov and Götze \(1999, Theorem 2.1\)](#), relating the exponential moments of $f - \mathbb{E}_\mu f$ to those of $\Gamma(f)^2$.

Theorem 3.1. *Assume that (Ω, μ, Γ) satisfies (1.1) with constant $\rho > 0$. Then for any $f \in \mathcal{A}$ and any $\alpha > \frac{\rho}{2}$ we have*

$$\mathbb{E}_\mu \exp(f - \mathbb{E}_\mu f) \leq \left(\mathbb{E}_\mu \exp(\alpha \Gamma(f)^2)\right)^{\frac{\rho}{2\alpha - \rho}}.$$

Note that formally, Theorem 3.1 and our own results like Theorem 1.1 are valid for bounded functions f only, since Γ was defined on a subset of bounded functions. However, it is not hard to see that our proofs can usually be extended to a suitable larger class of functions $\tilde{\mathcal{A}} \supset \mathcal{A}$. One possible approach is first to truncate the random variable f under consideration, and then prove bounds which are independent of the truncation level. As this is somewhat situational and depends on the difference operator Γ , we stick to the boundedness assumption for the sake of a clearer presentation of the arguments. Nevertheless, we can prove Theorem 1.1 under the assumption that Γ can be suitably defined for the function f at hand, and that $\Gamma(f) \leq g$ for some sub-Gaussian function.

Furthermore, we need an elementary inequality to adjust the constants in concentration or deviation inequalities: for any two constants $c_1 > c_2 > 1$ we have for all $r \geq 0$ and $c > 0$

$$c_1 \exp(-cr) \leq c_2 \exp\left(-\frac{\log(c_2)}{\log(c_1)} cr\right) \tag{3.1}$$

whenever the left hand side is smaller or equal to 1.

Proof of Theorem 1.1: Assume that $\rho = 1$, which can always be achieved by defining a new difference operator $\Gamma_\rho(f) = \sqrt{\rho}\Gamma(f)$. The general inequality follows by straightforward modifications from the $\rho = 1$ case.

Making use of Theorem 3.1 in the first and $a^2 \leq 2(a - b)_+^2 + 2b^2$ for any $a, b \geq 0$ in the second inequality, we obtain for all $\lambda \geq 0$

$$\mathbb{E}_\mu \exp(\lambda(f - \mathbb{E}_\mu f)) \leq \mathbb{E}_\mu \exp(\lambda^2 \Gamma(f)^2) \leq \exp(2\lambda^2 c^2) \mathbb{E}_\mu \exp(2\lambda^2 (g - c)_+^2).$$

The sub-Gaussian condition (1.3) leads to

$$\int \exp(2\lambda^2 (g - c)_+^2) d\mu \leq 1 + \int_0^\infty \exp\left(-t\left(\frac{1}{4\lambda^2 K^2} - 1\right)\right) dt \leq \frac{C}{1 - 4\lambda^2 K^2}$$

whenever $4\lambda^2 K^2 < 1$. Consequently, for all $\lambda \in [0, (2K)^{-1})$ we obtain by Markov's inequality

$$\mu(f - \mathbb{E}_\mu f \geq t) \leq \frac{C}{1 - 4\lambda^2 K^2} \exp(-\lambda t + 2\lambda^2 c^2). \tag{3.2}$$

Now we distinguish the two cases $t \leq \frac{c^2}{K}$ and $t > \frac{c^2}{K}$. In the first case, set $\lambda := \frac{t}{4c^2}$ (which implies $4\lambda^2 K^2 \leq 1/4$ and thus is in the range) to obtain

$$\frac{C}{1 - 4\lambda^2 K^2} \exp(-\lambda t + 2\lambda^2 c^2) \leq \frac{4C}{3} \exp\left(-\frac{t^2}{4c^2} + \frac{t^2}{8c^2}\right) = \frac{4C}{3} \exp\left(-\frac{t^2}{8c^2}\right), \tag{3.3}$$

using the monotonicity of $\frac{1}{1-x}$. In the second case, we simply set $\lambda := \frac{1}{4K}$ (implying $\lambda^2 K^2 = 1/4$) and observe that

$$\frac{C}{1 - 4\lambda^2 K^2} \exp(-\lambda t + 2\lambda^2 c^2) \leq \frac{4C}{3} \exp\left(-\frac{t}{4K} + \frac{c^2}{8K^2}\right) \leq \frac{4C}{3} \exp\left(-\frac{t}{8K}\right). \tag{3.4}$$

Combining (3.3) and (3.4) finishes the proof of (1.4).

Finally, (1.5) follows by considering $-f$ instead of f , which yields

$$\mu(|f - \mathbb{E}_\mu f| \geq t) \leq \frac{8C}{3} \exp\left(-\frac{1}{8} \min\left(\frac{t^2}{c^2}, \frac{t}{K}\right)\right).$$

The constant can be adjusted using (3.1). □

Proof of Corollary 1.2: Using the Γ -mLSI(ρ), by applying Theorem 3.1 to $f := \lambda g$, Markov's inequality and optimizing it can be shown that for all $t \geq 0$

$$\mu(g - \mathbb{E}_\mu g \geq t) \leq \exp\left(-\frac{t^2}{2\rho b^2}\right). \quad (3.5)$$

Here, to obtain the factor 2 in the denominator, one has to let $\alpha \rightarrow \infty$ in Theorem 3.1. Thus, the corollary follows easily from Theorem 1.1. \square

Proof of Proposition 1.3: We assume $b = 1$ which can be done by rescaling.

First, observe that Bobkov and Götze (1999, equation (2.4)) holds for any positive function g , since the inequality $\Gamma(g^2) \leq 2g\Gamma(g)$ is sufficient to apply the argument given therein. Thus, for any positive function g satisfying $\Gamma(g) \leq 1$ it holds for $\lambda \in [0, (2\rho)^{-1}]$

$$\mathbb{E}_\mu \exp(\lambda g^2) \leq \exp\left(\frac{\lambda}{1 - 2\rho\lambda} \mathbb{E}_\mu g^2\right). \quad (3.6)$$

So, by applying Theorem 3.1 (with $\alpha = \rho$) we have

$$\mathbb{E}_\mu \exp\left(\frac{1}{2\rho}(f - \mathbb{E}_\mu f)\right) \leq \mathbb{E}_\mu \exp\left(\frac{1}{4\rho}\Gamma(f)^2\right) \leq \mathbb{E}_\mu \exp\left(\frac{1}{4\rho}g^2\right) \leq \exp\left(\frac{1}{2\rho}\mathbb{E}_\mu g^2\right),$$

which can also be applied to λf and λg instead of f and g , for $\lambda \in [0, 1]$. Thus, by Markov's inequality, for any $\lambda \in [0, 1]$

$$\mu(f - \mathbb{E}_\mu f \geq t) \leq \exp\left(-\frac{\lambda t}{2\rho} + \frac{\lambda^2}{2\rho}\mathbb{E}_\mu g^2\right).$$

The claim follows by putting $\lambda = \min(\frac{t}{2\mathbb{E}_\mu g^2}, 1)$ and noting that if $t/(2\mathbb{E}_\mu g^2) \geq 1$, we have $t - \mathbb{E}_\mu g^2 \geq t/2$. \square

Proof of Proposition 1.4: Choosing $\alpha = \rho$ in Theorem 3.1, applying the inequality to λf and using the monotonicity leads to

$$\mathbb{E}_\mu \exp(\lambda(f - \mathbb{E}_\mu f)) \leq \exp(\lambda^2\rho(b + a\mathbb{E}_\mu f)) \mathbb{E}_\mu \exp(\lambda^2\rho a(f - \mathbb{E}_\mu f)).$$

Thus for $\lambda \in (0, (a\rho)^{-1})$, by Jensen's inequality (applied to the concave function $x \mapsto x^{\lambda\rho a}$) we have

$$(1 - \lambda\rho a) \log(\mathbb{E}_\mu \exp(\lambda(f - \mathbb{E}_\mu f))) \leq \lambda^2\rho(b + a\mathbb{E}_\mu f).$$

Finally, Markov's inequality and Boucheron et al. (2003, Lemma 11) yield the first inequality.

To see the second inequality, note that for any $\lambda > 0$ such that $\lambda a\rho < 1$, by Theorem 3.1 and concavity of $x \mapsto x^{\lambda a\rho}$, it holds

$$\begin{aligned} \mathbb{E}_\mu \exp(\lambda(\mathbb{E}_\mu f - f)) &\leq \mathbb{E}_\mu \exp(\rho\Gamma(-\lambda f)^2) = \mathbb{E}_\mu \exp(\lambda^2\rho\Gamma(f)^2) \\ &\leq \mathbb{E}_\mu \exp(\lambda^2\rho(a f + b)) \\ &= \exp(\lambda^2\rho(a\mathbb{E}_\mu f + b)) \mathbb{E}_\mu \exp(\lambda^2\rho a(f - \mathbb{E}_\mu f)) \\ &\leq \exp(\lambda^2\rho(a\mathbb{E}_\mu f + b)) \left(\mathbb{E}_\mu \exp(\lambda(f - \mathbb{E}_\mu f))\right)^{\lambda\rho a}. \end{aligned}$$

Finally, applying the estimates from the first part we obtain

$$\mathbb{E}_\mu \exp(\lambda(\mathbb{E}_\mu f - f)) \leq \exp\left(\frac{\lambda^2\rho}{1 - \lambda\rho a}(a\mathbb{E}_\mu f + b)\right).$$

The concentration inequality follows as in the first part. \square

Proof of Proposition 1.5: Using and rewriting Gao and Quastel (2003, Theorem 1) we obtain for any $f : S_n \rightarrow \mathbb{R}$

$$\text{Ent}(e^f) \leq \frac{1}{2nn!} \sum_{i,j=1}^n \sum_{\sigma \in S_n} (f(\sigma\tau_{ij}) - f(\sigma))(e^{f(\sigma\tau_{ij})} - e^{f(\sigma)}).$$

Now, the inequality $(a - b)(e^a - e^b) \leq \frac{1}{2}(e^a + e^b)(a - b)^2$ and the fact that $\sigma \mapsto \sigma\tau_{ij}$ is an automorphism of S_n leads to the Γ -mLSI(1). The Γ^+ -mLSI(2) follows in the same manner from the inequality $(a - b)_+(e^a - e^b) \leq (a - b)_+^2 e^a$. \square

Proof of Theorem 1.6: By Proposition 1.5 and Theorem 3.1 we have for any $f : S_n \rightarrow \mathbb{R}$, any $\lambda \in \mathbb{R}$ and any $\alpha > 1/2$ the inequality

$$\mathbb{E}_{\pi_n} \exp(\lambda(f - \mathbb{E}_{\pi_n} f)) \leq \left(\mathbb{E}_{\pi_n} \exp(\alpha\lambda^2\Gamma(f)^2) \right)^{\frac{1}{2\alpha-1}}.$$

If f is locally Lipschitz with respect to d , an easy calculation shows that we can upper bound $\Gamma(f)^2 \leq \text{ObsDiam}(S_n, d)$, so that from the above inequality in combination with $\alpha \rightarrow \infty$ we get

$$\mathbb{E}_{\pi_n} \exp(\lambda(f - \mathbb{E}_{\pi_n} f)) \leq \exp(\lambda^2 \text{ObsDiam}(S_n, d)/2).$$

The sub-Gaussian estimate follows by Markov's inequality and the variance bound from integration by parts. \square

In order to prove Proposition 1.7, we first need to establish the following lemma:

Lemma 3.2. *Let $f : S_n \rightarrow \mathbb{R}$ be a non-negative function such that*

- (1) $\Gamma^+(f)^2 \leq f$,
- (2) $|f(\sigma) - f(\sigma\tau_{ij})| \leq 1$ for all σ, i, j .

Then for all $t \in [0, \mathbb{E}_{\pi_n} f]$ we have

$$\pi_n(\mathbb{E}_{\pi_n} f - f \geq t) \leq \exp\left(-\frac{t^2}{8\mathbb{E}_{\pi_n} f}\right)$$

Epecially we have

$$\pi_n(f = 0) \exp\left(\frac{\mathbb{E}_{\pi_n} f}{8}\right) \leq 1.$$

In particular, this holds for $f(\sigma) = \frac{1}{16}d_T(\sigma, A)^2$, where $A \subset S_n$ is any set.

Proof of Lemma 3.2: Rewriting Gao and Quastel (2003, Theorem 1), we have that for any positive function g ,

$$\begin{aligned} \text{Ent}_{\pi_n}(g) &\leq \frac{1}{n!} \frac{1}{2n} \sum_{i,j} \sum_{\sigma \in S_n} (g(\sigma\tau_{ij}) - g(\sigma))(\log g(\sigma\tau_{ij}) - \log g(\sigma)) \\ &= \frac{1}{n!} \frac{1}{n} \sum_{i,j} \sum_{\sigma \in S_n} (g(\sigma\tau_{ij}) - g(\sigma))(\log g(\sigma\tau_{ij}) - \log g(\sigma))_+. \end{aligned}$$

Using this, we obtain for any $\lambda \in [0, 1]$

$$\begin{aligned} \text{Ent}_{\pi_n}(e^{-\lambda f}) &\leq \frac{\lambda}{n} \mathbb{E}_{\pi_n} \sum_{i,j} (f(\sigma) - f(\sigma\tau_{ij}))_+ (\exp(-\lambda f(\sigma\tau_{ij})) - \exp(-\lambda f(\sigma))) \\ &\leq \frac{\lambda}{n} \mathbb{E}_{\pi_n} \sum_{i,j} (f(\sigma) - f(\sigma\tau_{ij}))_+ (\exp(\lambda(f(\sigma) - f(\sigma\tau_{ij}))) - 1) e^{-\lambda f(\sigma)} \\ &\leq \frac{\lambda}{n} \mathbb{E}_{\pi_n} \sum_{i,j} (f(\sigma) - f(\sigma\tau_{ij}))_+ \Psi(\lambda(f(\sigma) - f(\sigma\tau_{ij}))) e^{-\lambda f(\sigma)}, \end{aligned}$$

where $\Psi(x) := e^x - 1$. By a Taylor expansion it can easily be seen that $\Psi(x) \leq 2x$ for all $x \in [0, 1]$, so that (recall that by (2) we have $f(\sigma) - f(\sigma\tau_{ij}) \leq 1$, and $f(\sigma) - f(\sigma\tau_{ij}) \geq 0$ due to the positive part)

$$\begin{aligned} \text{Ent}_{\pi_n}(e^{-\lambda f}) &\leq \frac{2\lambda^2}{n} \mathbb{E}_{\pi_n} \sum_{i,j} (f(\sigma) - f(\sigma\tau_{ij}))_+^2 e^{-\lambda f(\sigma)} \\ &= 2\lambda^2 \mathbb{E}_{\pi_n} \Gamma^+(f)^2 e^{-\lambda f} \leq 2\lambda^2 \mathbb{E}_{\pi_n} f e^{-\lambda f}. \end{aligned}$$

Chebyshev’s association inequality yields

$$\text{Ent}_{\pi_n}(e^{-\lambda f}) \leq 2\lambda^2 \mathbb{E}_{\pi_n} f \mathbb{E}_{\pi_n} e^{-\lambda f}.$$

In other terms, if we set $h(\lambda) := \mathbb{E}_{\pi_n} e^{-\lambda f}$, we have

$$\left(\frac{\log h(\lambda)}{\lambda}\right)' \leq 2 \mathbb{E}_{\pi_n} f,$$

which by the fundamental theorem of calculus implies for all $\lambda \in [0, 1]$

$$\mathbb{E}_{\pi_n} \exp\left(\lambda(\mathbb{E}_{\pi_n} f - f)\right) \leq \exp\left(2\lambda^2 \mathbb{E}_{\pi_n} f\right).$$

So, for any $t \in [0, \mathbb{E}_{\pi_n} f]$, by Markov’s inequality and setting $\lambda = \frac{t}{4 \mathbb{E}_{\pi_n} f}$

$$\pi_n(\mathbb{E}_{\pi_n} f - f \geq t) \leq \exp\left(-\lambda t + 2\lambda^2 \mathbb{E}_{\pi_n} f\right) = \exp\left(-\frac{t^2}{8 \mathbb{E}_{\pi_n} f}\right).$$

The second part follows by nonnegativity and $t = \mathbb{E}_{\pi_n} f$.

It remains to show that $f(\sigma) = \frac{1}{16} d_T(\sigma, A)^2$ satisfies the two conditions of this lemma. To this end, we first need to show that $\Gamma^+(d_T(\cdot, A))^2 \leq 4$. Writing $g(\sigma) := d_T(\sigma, A)$, it is well known (see [Boucheron et al., 2003](#)) that we have

$$g(\sigma) = \inf_{\nu \in \mathcal{M}(A)} \sup_{\alpha \in \mathbb{R}^n: |\alpha|_2=1} \sum_{k=1}^n \alpha_k \nu(\sigma' : \sigma'_k \neq \sigma_k), \tag{3.7}$$

where $\mathcal{M}(A)$ is the set of all probability measures on A . To estimate $\Gamma^+(g)^2(\sigma)$, one has to compare $g(\sigma)$ and $g(\sigma\tau_{ij})$. To this end, for any $\sigma \in S_n$ fixed, let $\tilde{\alpha}, \tilde{\nu}$ be parameters for which the value $g(\sigma)$ is attained, and let $\hat{\nu} = \hat{\nu}_{ij}$ be a minimizer of $\inf_{\nu \in \mathcal{M}(A)} \sum_{k=1}^n \tilde{\alpha}_k \nu(\sigma' : \sigma'_k \neq (\sigma\tau_{ij})_k)$. This leads to

$$\begin{aligned} \Gamma^+(g)(\sigma)^2 &\leq \frac{1}{n} \sum_{i,j=1}^n \left(\sum_{k=1}^n \tilde{\alpha}_k (\hat{\nu}(\sigma'_k \neq \sigma_k) - \hat{\nu}(\sigma'_k \neq (\sigma\tau_{ij})_k)) \right)_+^2 \\ &\leq \frac{2}{n} \sum_{i,j=1}^n (\tilde{\alpha}_i^2 + \tilde{\alpha}_j^2) \leq 4. \end{aligned}$$

Using this and the non-negativity of $d_T(\cdot, A)$, we have

$$\Gamma^+(f)^2 = \frac{1}{256} \Gamma^+(d_T(\cdot, A)^2)^2 \leq \frac{1}{64} d_T(\cdot, A)^2 \Gamma^+(d_T(\cdot, A))^2 \leq f.$$

To show the second property, we proceed similarly to [Boucheron et al. \(2009, Proof of Lemma 1\)](#). By [\(3.7\)](#) and the Cauchy–Schwarz inequality, we have

$$f(\sigma) = \frac{1}{16} \inf_{\nu \in \mathcal{M}(A)} \sum_{k=1}^n \nu(\sigma' : \sigma'_k \neq \sigma_k)^2.$$

Assuming without loss of generality that $f(\sigma) \geq f(\sigma\tau_{ij})$, choose $\hat{\nu} = \hat{\nu}_{ij} \in \mathcal{M}(A)$ such that the value of $f(\sigma\tau_{ij})$ is attained. It follows that

$$f(\sigma) - f(\sigma\tau_{ij}) \leq \frac{1}{16} \sum_{k=1}^n \hat{\nu}(\sigma'_k \neq \sigma_k)^2 - \hat{\nu}(\sigma'_k \neq (\sigma\tau_{ij})_k)^2 \leq \frac{2}{16},$$

which finishes the proof. □

The proof of [Proposition 1.7](#) is now easily completed:

Proof of Proposition 1.7: The difference operator Γ^+ clearly satisfies $\Gamma^+(g^2) \leq 2g\Gamma^+(g)$ for all positive functions g , as well as an mLSI(2). Moreover, as seen in the proof of [Lemma 3.2](#), we have $\Gamma^+(d_T(\cdot, A)) \leq 2$. Thus, by [\(3.6\)](#) it holds for $\lambda \in [0, 1/4)$

$$\pi_n(A) \mathbb{E}_{\pi_n} \exp\left(\frac{\lambda}{4} d_T(\cdot, A)^2\right) \leq \pi_n(A) \exp\left(\frac{\lambda}{4 - 16\lambda} \mathbb{E}_{\pi_n} d_T(\cdot, A)^2\right).$$

Furthermore, [Lemma 3.2](#) shows that

$$\pi_n(A) \exp\left(\frac{\mathbb{E}_{\pi_n} d_T(\cdot, A)^2}{128}\right) \leq 1.$$

So, for $\lambda = 1/36$ we have

$$\pi_n(A) \mathbb{E}_{\pi_n} \exp\left(\frac{d_T(\cdot, A)^2}{144}\right) \leq \pi_n(A) \exp\left(\frac{1}{128} \mathbb{E}_{\pi_n} d_T(\cdot, A)^2\right) \leq 1.$$

□

Proof of Proposition 1.8: Again, the proof mimics the proof given for independent random variables in [Boucheron et al. \(2003\)](#). As stated in [Proposition 1.5](#), the uniform measure π_n on S_n satisfies a Γ^+ -mLSI(2) with respect to

$$\Gamma^+(f)(\sigma)^2 = \frac{1}{n} \sum_{i,j=1}^n (f(\sigma) - f(\sigma\tau_{ij}))_+^2.$$

Writing $f_A(\sigma) := d_T(\sigma, A)$, we have $\Gamma^+(f_A)(\sigma)^2 \leq 4$ as seen in the proof of [Lemma 3.2](#). Hence, by similar arguments as in the proof of [Theorem 1.1](#) we have for any $\lambda \geq 0$

$$\mathbb{E}_{\pi_n} \exp(\lambda(f_A - \mathbb{E}_{\pi_n} f_A)) \leq \exp(4\lambda^2), \tag{3.8}$$

implying the sub-Gaussian estimate $\pi_n(f_A - \mathbb{E}_{\pi_n} f_A \geq t) \leq \exp(-t^2/16)$. Fix a set $A \subseteq S_n$ satisfying $\pi_n(A) \geq 1/2$. As a Γ -mLSI(1) implies a Poincaré inequality (see [Bobkov and Tetali, 2006, Proposition 3.5](#) or [Diaconis and Saloff-Coste, 1996](#)), we also have (by Chebyshev’s inequality)

$$t^2 \pi_n(f_A - \mathbb{E}_{\pi_n} f_A \leq -t) \leq \text{Var}_{\pi_n}(f_A) \leq 2 \mathbb{E}_{\pi_n} \Gamma^+(f_A)^2 \leq 8,$$

which evaluated at $t = \mathbb{E}_{\pi_n} f_A$ yields $(\mathbb{E}_{\pi_n} f_A)^2 \leq 16$. Thus, for any $t \geq 4$ it holds

$$\pi_n(f_A \geq t) \leq \exp\left(-\frac{(t-4)^2}{16}\right) \leq 2 \exp\left(-\frac{t^2}{64}\right), \tag{3.9}$$

where the last inequality follows from $(t-4)^2 \geq t^2/2 - 16$ for any $t \geq 0$ and (3.1). For $t \leq 4$ the inequality (3.9) holds trivially. \square

The proofs of the results for slices of the hypercube work in a very similar way.

Proof of Proposition 1.9: It follows from Gao and Quastel (2003, Theorem 1) that we have for any $f : C_{n,r} \rightarrow \mathbb{R}$

$$\text{Ent}(e^f) \leq \frac{1}{n \binom{n}{r}} \sum_{\eta \in C_{n,r}} \sum_{i < j} (f(\tau_{ij}\eta) - f(\eta))(e^{f(\tau_{ij}\eta)} - e^{f(\eta)}).$$

From here, we may proceed as in the proof of Proposition 1.5. \square

For the proof of Proposition 1.10, we need to establish the following analogue of Lemma 3.2:

Lemma 3.3. *Let $f : C_{n,r} \rightarrow \mathbb{R}$ be a non-negative function such that*

- (1) $\Gamma^+(f)^2 \leq f$,
- (2) $|f(\eta) - f(\tau_{ij}\eta)| \leq 1$ for all η, i, j .

Then for all $t \in [0, \mathbb{E}_{\mu_{n,r}} f]$ we have

$$\mu_{n,r}(\mathbb{E}_{\mu_{n,r}} f - f \geq t) \leq \exp\left(-\frac{t^2}{8 \mathbb{E}_{\mu_{n,r}} f}\right)$$

Epecially we have

$$\mu_{n,r}(f = 0) \exp\left(\frac{\mathbb{E}_{\mu_{n,r}} f}{8}\right) \leq 1.$$

In particular, this holds for $f(\eta) = \frac{1}{32} d_T(\eta, A)^2$, where $A \subset C_{n,r}$ is any set.

Proof of Lemma 3.3: Rewriting Gao and Quastel (2003, Theorem 1), we have that for any positive function g ,

$$\begin{aligned} \text{Ent}_{\mu_{n,r}}(g) &\leq \frac{1}{\binom{n}{r}} \frac{1}{n} \sum_{i < j} \sum_{\eta \in C_{n,r}} (g(\tau_{ij}\eta) - g(\eta))(\log g(\tau_{ij}\eta) - \log g(\eta)) \\ &= \frac{1}{\binom{n}{r}} \frac{2}{n} \sum_{i < j} \sum_{\eta \in C_{n,r}} (g(\tau_{ij}\eta) - g(\eta))(\log g(\tau_{ij}\eta) - \log g(\eta))_+. \end{aligned}$$

From here, we may mimic the proof of Lemma 3.2.

Last, we need to show that $f(\eta) = \frac{1}{32} d_T(\eta, A)^2$ satisfies the two conditions of this lemma. As compared to the proof of Lemma 3.2, some of the constants will change because of the different normalization of the difference operators. However, we may argue similarly and show that $\Gamma^+(d_T(\cdot, A))^2 \leq 8$. Using this and the non-negativity of $d_T(\cdot, A)$ yields

$$\Gamma^+(f)^2 = \frac{1}{1024} \Gamma^+(d_T(\cdot, A)^2)^2 \leq \frac{1}{256} d_T(\cdot, A)^2 \Gamma^+(d_T(\cdot, A))^2 \leq f.$$

Finally, by arguing as above it is easily seen that $|f(\eta) - f(\tau_{ij}\eta)| \leq 2/32$. \square

Proof of Proposition 1.10: As the difference operator Γ^+ satisfies $\Gamma^+(g^2) \leq 2g\Gamma^+(g)$ for all positive functions g , as well as an mLSI(2), it remains to change the proof of Proposition 1.7 in view of the different constants appearing in Lemma 3.3. As noted in the proof of Lemma 3.3, we have $\Gamma^+(d_T(\cdot, A)) \leq \sqrt{8}$. Thus, by (3.6) it holds for $\lambda \in [0, 1/4]$

$$\mu_{n,r}(A) \mathbb{E}_{\mu_{n,r}} \exp\left(\frac{\lambda}{8} d_T(\cdot, A)^2\right) \leq \mu_{n,r}(A) \exp\left(\frac{\lambda}{8 - 32\lambda} \mathbb{E}_{\mu_{n,r}} d_T(\cdot, A)^2\right).$$

Furthermore, Lemma 3.3 shows that

$$\mu_{n,r}(A) \exp\left(\frac{\mathbb{E}_{\mu_{n,r}} d_T(\cdot, A)^2}{256}\right) \leq 1.$$

So, for $\lambda = 1/68$ we have

$$\mu_{n,r}(A) \mathbb{E}_{\mu_{n,r}} \exp\left(\frac{d_T(\cdot, A)^2}{544}\right) \leq \mu_{n,r}(A) \exp\left(\frac{1}{256} \mathbb{E}_{\mu_{n,r}} d_T(\cdot, A)^2\right) \leq 1.$$

□

Finally, we present the proofs of Section 2.

Proof of Proposition 2.8: We show that f is weakly $(k\text{ML}(f), 0)$ -self bounding in the language of Boucheron et al. (2009). To see this, for any $v \in V$ let $f_v(x_{v^c}) := \sum_{e \in E: v \notin e} w_e X_e = f(X_{v^c}, 0)$. Now we have

$$\begin{aligned} \sum_{v \in V} (f(x) - f_v(x_{v^c}))^2 &= \sum_{v \in V} \left(X_v \sum_{e \in E: v \in e} w_e X_{e \setminus v}\right)^2 \leq \sum_{v \in V} X_v \partial_v f(X)^2 \\ &\leq \text{ML}(f) \sum_{v \in V} X_v \partial_v f(X) \leq k\text{ML}(f) f(X). \end{aligned}$$

Here, the first inequality follows from $X_v \in [0, 1]$ and the last one is a consequence of Euler’s homogeneous function theorem and the fact that all quantities involved are positive. Consequently, Boucheron et al. (2009, Theorem 1) yields for any $t \geq 0$

$$\mathbb{P}(f(X) - \mathbb{E} f(X) \geq t) \leq \exp\left(-\frac{t^2}{2k\text{ML}(f)(\mathbb{E} f(X) + t/2)}\right).$$

For the lower bound, apply Boucheron et al. (2009, Theorem 1) to $\tilde{f} = \text{ML}(f)^{-1} f$ which satisfies $0 \leq \tilde{f}(x) - \tilde{f}_v(x_{v^c}) \leq 1$ for all $v \in V$ and $x \in [0, 1]^V$ and is weakly $(k\text{ML}(f)^{-1}, 0)$ -self bounding. □

Proof of Proposition 2.9: The first part follows as above. As for the second part, if we choose $\mathcal{F} = \mathcal{F}_q = \{a \in \mathbb{R}^V : a_v \geq 0, \|a\|_q \leq 1\}$ for some $q \in [1, \infty]$ this leads to

$$f_{\mathcal{F}}(X) = \sup_{a \in \mathcal{F}_q} \sum_{v \in V} a_v X_v = \left(\sum_{v \in V} |X_v|^p\right)^{1/p}$$

for the Hölder conjugate p , which is due to the nonnegativity of the X_i and the dual formulation of the L^p norm in \mathbb{R}^V . □

Proof of Proposition 2.10: Clearly, f_d is d -homogeneous and has positive weights in the sense of (2.1), if we set $V = [n]$ and $E = \{\{j, j + 1, \dots, j + d - 1\}, j = 1, \dots, n\}$, $w_e = 1$. Furthermore, the partial derivatives can be easily bounded: For any fixed $l \in [n]$ there are exactly d terms which depend on X_l , and the product is

bounded by 1. Consequently, $\text{ML}(f_d) = \max_{t \in [n]} \max_{x \in [0,1]^n} \partial_t f(X) = d$. Thus, Proposition 2.8 yields for all $t \geq 0$

$$\mathbb{P}(f_d(X) - \mathbb{E} f_d(X) \geq t) \leq \exp\left(-\frac{t^2}{2d^2(\mathbb{E} f_d(X) + t/2)}\right).$$

The assertion now follows, if we note that $\mathbb{E} f_d(X) = n\eta^d$. □

Let us now prove the results from Section 2.4. To this end, we first need to establish some basic properties of modified logarithmic Sobolev inequalities with respect to the difference operators we use.

Lemma 3.4. *Let μ be a probability measure on a product of Polish spaces $\mathcal{X} = \otimes_{i=1}^n \mathcal{X}_i$ which satisfies a \mathfrak{d} -mLSI(σ^2). Then, μ also satisfies a \mathfrak{d}^+ -mLSI($2\sigma^2$).*

Proof: Let $(\Omega, \mathcal{F}, \nu)$ be a probability space and g a measurable function on it. Then,

$$\begin{aligned} & \iint (g(x) - g(y))^2 d\nu(y) e^{g(x)} d\nu(x) \\ &= \iint ((g(x) - g(y))_+^2 e^{g(x)} + (g(y) - g(x))_+^2 e^{g(x)}) d\nu(y) d\nu(x) \\ &\leq \iint ((g(x) - g(y))_+^2 e^{g(x)} + (g(y) - g(x))_+^2 e^{g(y)}) d\nu(y) d\nu(x) \\ &= 2 \iint (g(x) - g(y))_+^2 d\nu(y) e^{g(x)} d\nu(x). \end{aligned}$$

Applying this to $\nu = \mu(\cdot | x_{i^c})$ and $g = f(x_{i^c}, \cdot)$ for any $i = 1, \dots, n$ yields

$$\int |\mathfrak{d}f|^2 e^f d\mu \leq 2 \int |\mathfrak{d}^+ f|^2 e^f d\mu,$$

which finishes the proof. □

Also note that by monotonicity a \mathfrak{d} -mLSI(σ^2) implies an \mathfrak{h} -mLSI(σ^2), and the same holds for \mathfrak{d}^+ and \mathfrak{h}^+ . Moreover, we recall the duality formula $|x| = \sup_{y \in S^{n-1}} \langle x, y \rangle$.

Proof of Proposition 2.16: First, (2.8) follows by applying Theorem 1.1 to $g = |\mathfrak{d}f|$ and noting that $|\mathfrak{d}(af)| = |a| |\mathfrak{d}f|$ for all $a \in \mathbb{R}$. To see that g is sub-Gaussian with parameter $K = \sqrt{2\sigma^2 b}$ and $C = 1$, note that by Lemma 3.4, μ satisfies a \mathfrak{d}^+ -mLSI($2\sigma^2$), so that we can use (3.5).

The same arguments are valid for \mathfrak{h}^+ and \mathfrak{h} respectively. Here, we additionally use the estimate $|\mathfrak{h}^+ |\mathfrak{h}f|| \leq |\mathfrak{h}^{(2)} f|_{\text{op}}$ (cf. Götze et al., 2020, Lemma 3.2). □

Proof of Proposition 2.17: Let us bound $|\mathfrak{d}^+ h|^2$. Choose the matrix $\tilde{B} \in \mathcal{B}$ maximizing $\sup_{B \in \mathcal{B}} \langle x, Bx \rangle$ and use the monotonicity of $y \mapsto y_+$ to obtain

$$\begin{aligned} |\mathfrak{d}^+ h(x)|^2 &= \sum_{i=1}^n \int (g(x) - g(x_{i^c}, x'_i))_+^2 d\mu(x'_i | x_{i^c}) \\ &\leq \sum_{i=1}^n \sup_{x'_i} \left(2(x_i - x'_i) \sum_{j=1}^n \tilde{B}_{ij} x_j \right)_+^2 \\ &\leq 16 \|\tilde{B}x\|_2^2 \leq 16 \sup_{B \in \mathcal{B}} \|Bx\|_2^2 = 16 f_{\mathcal{B}}^2(x). \end{aligned}$$

Furthermore, we have for some maximizer $\tilde{B} \in \mathcal{B}$ of $\sup_{B \in \mathcal{B}} \|Bx\|$ and $\tilde{v} \in S^{n-1}$ for $\sup_{v \in S^{n-1}} \langle \tilde{B}x, v \rangle$

$$\begin{aligned} |\mathfrak{d}^+ f_{\mathcal{B}}|^2 &\leq \sum_i \sup_{x'_i} \left(\sup_v \langle \tilde{B}x, v \rangle - \sup_v \langle \tilde{B}(x_{i^c}, x'_i), v \rangle \right)_+^2 \\ &\leq \sum_i \sup_{x'_i} \left((x_i - x'_i) \langle \tilde{B}e_i, \tilde{v} \rangle \right)_+^2 \\ &\leq 4 \sum_i \langle \tilde{B}e_i, \tilde{v} \rangle^2 \leq 4 \left(\sup_w \sum_i w_i \langle \tilde{B}e_i, \tilde{v} \rangle \right)^2 \leq 4 \sup_{B \in \mathcal{B}} \|B\|_{\text{op}}^2. \end{aligned}$$

Here, the suprema of v and w are taken over the n -dimensional sphere. We can now apply Corollary 1.2 to $\Gamma = \mathfrak{d}^+$, $\rho = 2\sigma^2$, $g = 4f_{\mathcal{B}}$ and $b = 8\Sigma$ to finish the proof. \square

Proof of Proposition 2.18: The idea of the proof of the mLSIs is already present in Bobkov and Götze (2007). Let $(\Omega, \mathcal{F}, \nu)$ be any probability space. For any function g we have due to the inequality $(a - b)_+(e^a - e^b)_+ \leq \frac{1}{2}(a - b)_+^2(e^a + e^b)$ (for all $a, b \in \mathbb{R}$)

$$\begin{aligned} \text{Cov}_{\nu}(g, e^g) &\leq \frac{1}{2} \iint (g(x) - g(y))_+^2 (e^{g(x)} + e^{g(y)}) d\nu(x) d\nu(y) \\ &= \frac{1}{2} \iint (g(x) - g(y))^2 d\nu(y) e^{g(x)} d\nu(x). \end{aligned}$$

Applying this to $\nu = \mu(\cdot | x_{i^c})$ and $g = f(x_{i^c}, \cdot)$ and using (2.11) yields

$$\begin{aligned} \text{Ent}_{\mu}(e^f) &\leq \frac{\sigma^2}{2} \sum_{i=1}^n \iint (f(x) - f(x_{i^c}, x'_i))^2 d\mu(x'_i | x_{i^c}) e^{f(x)} d\mu(x) \\ &= \frac{\sigma^2}{2} \int |\mathfrak{d}f|^2 e^f d\mu. \end{aligned}$$

To see that μ also satisfies a \mathfrak{d}^+ -mSLI($2\sigma^2$), it remains to apply Lemma 3.4. The exponential inequalities are a consequence of Theorem 3.1. \square

Proof of Theorem 2.19: Write $X = (X_1, \dots, X_n)$. Let us assume that $\mathbb{E} X_i = 0$ for all $i \in \{1, \dots, n\}$, from which the general case follows easily using the inequality

$$\|\max_i |X_i - \mathbb{E} X_i|\|_{\Psi_2} \leq 4 \|\max_i |X_i|\|_{\Psi_2}.$$

Since the X_i are independent, it follows from Proposition 2.18 that their joint distribution \mathbb{P}_X satisfies a \mathfrak{d} -mLSI(1), and we can calculate

$$\begin{aligned} |\mathfrak{d}f|(X) &= \left(\sum_{i=1}^n \int (X_i - y_i)^2 d\mathbb{P}_{X_i}(y_i) \right)^{1/2} = \left(\sum_{i=1}^n X_i^2 + \mathbb{E} X_i^2 \right)^{1/2} \\ &\leq \|X\|_2 - \mathbb{E} \|X\|_2 + \mathbb{E} \|X\|_2 + \left(\sum_{i=1}^n \mathbb{E} X_i^2 \right)^{1/2} =: g(X). \end{aligned}$$

To apply Theorem 1.1, it remains to see that we may set $c = \mathbb{E} \|X\|_2 + \sqrt{\text{Var}(\sum_i X_i^2)}$ and $K = \|\max_i |X_i|\|_{\Psi_2}$. This is seen by noting that

$$\mathbb{P}(g(X) \geq c + t) = \mathbb{P}(\|X\|_2 - \mathbb{E} \|X\|_2 \geq t) \leq 2 \exp \left(-c_2 \frac{t^2}{\|\max_i |X_i|\|_{\Psi_2}^2} \right),$$

where the last step follows from [Klochkov and Zhivotovskiy \(2020, Lemma 1.4\)](#), as $X \mapsto \|X\|_2$ is a convex and 1-Lipschitz function. Note that although [Klochkov and Zhivotovskiy \(2020, Lemma 1.4\)](#) is formulated for $t \geq t_0 > 0$, one can easily find an estimate for all $t \geq 0$, by first multiplying the right hand side by 2, and then adjusting the constant in the exponential. \square

Recall that as discussed above, the application of [Theorem 3.1](#) is only possible for bounded functions, so that an additional truncation step needs to be done. Instead of applying [Theorem 3.1](#) to $f(X) = \sum_i X_i - \mathbb{E} X_i$, it is applied to the sum of the random variables $Y_i := g_R(X_i) - \mathbb{E} g_R(X_i)$ for $g_R(x) = \min(R, \max(x, -R))$ for a suitable truncation level $R > 0$. As the right hand side of [equation \(2.14\)](#) can be chosen to be independent of R , the theorem follows for unbounded random variables by letting $R \rightarrow \infty$.

Acknowledgements

The authors would like to thank the anonymous referees for valuable comments on the preliminary version of this work.

References

- Adamczak, R. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, **13**, no. 34, 1000–1034 (2008). [MR2424985](#).
- Adamczak, R. A note on the Hanson-Wright inequality for random vectors with dependencies. *Electron. Commun. Probab.*, **20**, no. 72, 13 (2015). [MR3407216](#).
- Adamczak, R., Bednorz, W., and Wolff, P. Moment estimates implied by modified log-Sobolev inequalities. *ESAIM Probab. Stat.*, **21**, 467–494 (2017). [MR3743923](#).
- Adamczak, R., Kotowski, M., Polaczyk, B., and Strzelecki, M. A note on concentration for polynomials in the Ising model. *Electron. J. Probab.*, **24**, Paper No. 42, 22 (2019). [MR3949267](#).
- Anderson, G. W., Guionnet, A., and Zeitouni, O. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge (2010). ISBN 978-0-521-19452-5. [MR2760897](#).
- Bobkov, S. G., Chistyakov, G. P., and Götze, F. Second-order concentration on the sphere. *Commun. Contemp. Math.*, **19** (5), 1650058, 20 (2017). [MR3670792](#).
- Bobkov, S. G. and Götze, F. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, **163** (1), 1–28 (1999). [MR1682772](#).
- Bobkov, S. G. and Götze, F. Concentration inequalities and limit theorems for randomized sums. *Probab. Theory Related Fields*, **137** (1-2), 49–81 (2007). [MR2278452](#).
- Bobkov, S. G. and Tetali, P. Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.*, **19** (2), 289–336 (2006). [MR2283379](#).
- Boucheron, S., Bousquet, O., Lugosi, G., and Massart, P. Moment inequalities for functions of independent random variables. *Ann. Probab.*, **33** (2), 514–560 (2005). [MR2123200](#).

- Boucheron, S., Lugosi, G., and Massart, P. A sharp concentration inequality with applications. *Random Structures Algorithms*, **16** (3), 277–292 (2000). [MR1749290](#).
- Boucheron, S., Lugosi, G., and Massart, P. Concentration inequalities using the entropy method. *Ann. Probab.*, **31** (3), 1583–1614 (2003). [MR1989444](#).
- Boucheron, S., Lugosi, G., and Massart, P. On concentration of self-bounding functions. *Electron. J. Probab.*, **14**, no. 64, 1884–1899 (2009). [MR2540852](#).
- Boucheron, S., Lugosi, G., and Massart, P. *Concentration inequalities. A nonasymptotic theory of independence*. Oxford University Press, Oxford (2013). ISBN 978-0-19-953525-5. [MR3185193](#).
- Caputo, P., Menz, G., and Tetali, P. Approximate tensorization of entropy at high temperature. *Ann. Fac. Sci. Toulouse Math. (6)*, **24** (4), 691–716 (2015). [MR3434252](#).
- Carlitz, L., Kurtz, D. C., Scoville, R., and Stackelberg, O. P. Asymptotic properties of Eulerian numbers. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **23**, 47–54 (1972). [MR309856](#).
- Chatterjee, S. *Concentration inequalities with exchangeable pairs*. ProQuest LLC, Ann Arbor, MI (2005). ISBN 978-0542-08643-4. Thesis (Ph.D.)—Stanford University. [MR2707160](#).
- Chatterjee, S. Stein’s method for concentration inequalities. *Probab. Theory Related Fields*, **138** (1-2), 305–321 (2007). [MR2288072](#).
- Chatterjee, S. The missing log in large deviations for triangle counts. *Random Structures Algorithms*, **40** (4), 437–451 (2012). [MR2925306](#).
- Chatterjee, S. and Diaconis, P. A central limit theorem for a new statistic on permutations. *Indian J. Pure Appl. Math.*, **48** (4), 561–573 (2017). [MR3741694](#).
- Clark, L. Limit distribution of ascent, descent or excedance length sums of permutations. *Appl. Anal. Discrete Math.*, **3** (2), 303–309 (2009). [MR2555041](#).
- DeMarco, B. and Kahn, J. Upper tails for triangles. *Random Structures Algorithms*, **40** (4), 452–459 (2012). [MR2925307](#).
- Diaconis, P. *Group representations in probability and statistics*, volume 11 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA (1988). ISBN 0-940600-14-5. [MR964069](#).
- Diaconis, P. and Saloff-Coste, L. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, **6** (3), 695–750 (1996). [MR1410112](#).
- Gao, F. and Quastel, J. Exponential decay of entropy in the random transposition and Bernoulli-Laplace models. *Ann. Appl. Probab.*, **13** (4), 1591–1600 (2003). [MR2023890](#).
- Götze, F. and Sambale, H. Second order concentration via logarithmic Sobolev inequalities. *Bernoulli*, **26** (1), 93–126 (2020). [MR4036029](#).
- Götze, F., Sambale, H., and Sinulis, A. Higher order concentration for functions of weakly dependent random variables. *Electron. J. Probab.*, **24**, Paper No. 85, 19 (2019). [MR4003138](#).
- Götze, F., Sambale, H., and Sinulis, A. Concentration inequalities for bounded functionals via generalized log-Sobolev inequalities. *J. Theoret. Probab.* (2020). [DOI: 10.1007/s10959-020-01016-x](#).
- Holley, R. and Stroock, D. Logarithmic Sobolev inequalities and stochastic Ising models. *J. Statist. Phys.*, **46** (5-6), 1159–1194 (1987). [MR893137](#).

- Hsu, D., Kakade, S. M., and Zhang, T. A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Probab.*, **17**, no. 52, 6 (2012). [MR2994877](#).
- Klochkov, Y. and Zhivotovskiy, N. Uniform Hanson-Wright type concentration inequalities for unbounded entries via the entropy method. *Electron. J. Probab.*, **25**, Paper No. 22, 30 (2020). [MR4073683](#).
- Ledoux, M. On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.*, **1**, 63–87 (1995/97). [MR1399224](#).
- Marton, K. An inequality for relative entropy and logarithmic Sobolev inequalities in Euclidean spaces. *J. Funct. Anal.*, **264** (1), 34–61 (2013). [MR2995699](#).
- Marton, K. Logarithmic Sobolev inequalities in discrete product spaces. *Combin. Probab. Comput.*, **28** (6), 919–935 (2019). [MR4015662](#).
- Massart, P. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.*, **28** (2), 863–884 (2000). [MR1782276](#).
- Maurey, B. Construction de suites symétriques. *C. R. Acad. Sci. Paris Sér. A-B*, **288** (14), A679–A681 (1979). [MR533901](#).
- McDiarmid, C. and Reed, B. Concentration for self-bounding functions and an inequality of Talagrand. *Random Structures Algorithms*, **29** (4), 549–557 (2006). [MR2268235](#).
- Paulin, D. The convex distance inequality for dependent random variables, with applications to the stochastic travelling salesman and other problems. *Electron. J. Probab.*, **19**, no. 68, 34 (2014). [MR3248197](#).
- Reinert, G. and Röllin, A. Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.*, **37** (6), 2150–2173 (2009). [MR2573554](#).
- Rio, E. Inégalités de concentration pour les processus empiriques de classes de parties. *Probab. Theory Related Fields*, **119** (2), 163–175 (2001). [MR1818244](#).
- Royer, G. *An initiation to logarithmic Sobolev inequalities*, volume 14 of *SMF/AMS Texts and Monographs*. American Mathematical Society, Providence, RI; Société Mathématique de France, Paris (2007). ISBN 978-0-8218-4401-4; 0-8218-4401-6. [MR2352327](#).
- Rudelson, M. and Vershynin, R. Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.*, **18**, no. 82, 9 (2013). [MR3125258](#).
- Samson, P.-M. Transport-entropy inequalities on locally acting groups of permutations. *Electron. J. Probab.*, **22**, Paper No. 62, 33 (2017). [MR3690287](#).
- Talagrand, M. An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities. *Proc. Amer. Math. Soc.*, **104** (3), 905–909 (1988). [MR964871](#).
- Talagrand, M. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, (81), 73–205 (1995). [MR1361756](#).
- Talagrand, M. New concentration inequalities in product spaces. *Invent. Math.*, **126** (3), 505–563 (1996). [MR1419006](#).
- van de Geer, S. and Lederer, J. The Bernstein-Orlicz norm and deviation inequalities. *Probab. Theory Related Fields*, **157** (1-2), 225–250 (2013). [MR3101846](#).
- Vershynin, R. *High-dimensional probability. An introduction with applications in data science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge (2018). ISBN 978-1-108-41519-4. [MR3837109](#).

Wolff, P. On some Gaussian concentration inequality for non-Lipschitz functions.
In *High dimensional probability VI*, volume 66 of *Progr. Probab.*, pp. 103–110.
Birkhäuser/Springer, Basel (2013). [MR3443496](#).