Electronic Journal of Probability

Electron. J. Probab. **24** (2019), no. 85, 1–19.
ISSN: 1083-6489 https://doi.org/10.1214/19-EJP338

Higher order concentration for functions of weakly dependent random variables*

Friedrich Götze† Holger Sambale† Arthur Sinulis†

Abstract

We extend recent higher order concentration results in the discrete setting to include functions of possibly dependent variables whose distribution (on the product space) satisfies a logarithmic Sobolev inequality with respect to a difference operator that arises from Glauber type dynamics. Examples include the Ising model on a graph with n sites with general, but weak interactions (i.e. in the Dobrushin uniqueness regime), for which we prove concentration results of homogeneous polynomials, as well as random permutations, and slices of the hypercube with dynamics given by either the Bernoulli-Laplace or the symmetric simple exclusion processes.

Keywords: concentration of measure; logarithmic Sobolev inequalities; Ising model.
AMS MSC 2010: 60E15.
Submitted to EJP on May 9, 2018, final version accepted on June 25, 2019.
Supersedes arXiv:1801.06348.

1 Introduction

In this article, we study *higher order* versions of the *concentration of measure phenomenon* for functions of random variables $X_1, \ldots, X_n$ defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in some Polish space $X_i : \Omega \to S_i$ which are not necessarily independent. The term higher order shall emphasize that we prove tail estimates for functions with possibly non-bounded first order differences, or functions for which the L^∞ norm of its differences increases with the size of the system, even after a proper normalization, such as quadratic forms.

To formalize this intuition we consider certain *difference operators*. By a difference operator we mean an operator $\Gamma : L^\infty(\mu) \to L^\infty(\mu)$ for some probability measure μ satisfying $\Gamma(af + b) = |a|\Gamma(f)$ for $b \in \mathbb{R}$ and either $a > 0$ or $a \in \mathbb{R}$ (note that this is usually not a linear operator in the language of functional analysis). The restriction

*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*".

†Faculty of Mathematics, Bielefeld University. E-mail: goetze@math.uni-bielefeld.de,hsambale@math.uni-bielefeld.de,asinulis@math.uni-bielefeld.de

$f \in L^\infty(\mu)$ is merely a minimal requirement, and in the cases that we will consider in the applications (i.e. finite probability spaces) $L^\infty(\mu)$ is the space of all functions. Here, μ is the distribution of $X \coloneqq (X_1, \ldots, X_n)$ on $S \coloneqq \times_{i=1}^n S_i$.

In this work, we will mainly consider two specific difference operators, for which we use the following notation. For a vector $x = (x_1, \ldots, x_n)$ and $I \subset \{1, \ldots, n\}$ we write $x_I \coloneqq (x_i)_{i \in I}$ and $x_{I^c} \coloneqq (x_i)_{i \notin I}$. If $I = \{j\}$ for some $j \in \{1, \ldots, n\}$ we abbreviate it as x_j and x_{j^c}. Now, consider some set of subsets of $\{1, \ldots, n\}$, denoted by $\mathcal{I}$. Given any subset $I \in \mathcal{I}$ let $\mu(\,\cdot \mid x_{I^c})$ be the regular conditional probability (see Proposition 2.1 for an existence statement), and define

$$\mathfrak{d}_I f(x) \coloneqq \left(\frac{1}{2} \int (f(x) - f(x_{I^c}, y_I))^2 d\mu(y_I \mid x_{I^c}) \right)^{1/2}$$
$$\mathfrak{h}_I f(x) \coloneqq \frac{1}{\sqrt{2}} \| f(x_{I^c}, y_I) - f(x_{I^c}, z_I) \|_{L^\infty(\mu(\,\cdot|x_{I^c}) \otimes \mu(\,\cdot|x_{I^c})(y_I, z_I))}.$$

We let $\mathfrak{d}f = (\mathfrak{d}_I f)_{I \in \mathcal{I}}$ and $\mathfrak{h}f = (\mathfrak{h}_I f)_{I \in \mathcal{I}}$ to obtain a vector of "partial derivatives" indexed by $\mathcal{I}$. Now, for either $\mathfrak{d}$ or $\mathfrak{h}$, we define a difference operator by setting $\Gamma(f) = |\mathfrak{d}f|$ or $\Gamma(f) = |\mathfrak{h}f|$ for the Euclidean norm $|\cdot|$ and call it the associated operator to $(\mathfrak{d}, \mathcal{I})$ or $(\mathfrak{h}, \mathcal{I})$ respectively. Using $\mathfrak{h}$, it is possible to define higher order difference operators $\mathfrak{h}^{(d)}$ for any $d \geq 2$ as follows. For any $I_1, \ldots, I_d \in \mathcal{I}$ set

$$\mathfrak{h}_{I_1 \ldots I_d} f = \mathfrak{h}_{I_1}(\mathfrak{h}_{I_2 \ldots I_d} f), \tag{1.1}$$

and define the tensor of d-th order differences $\mathfrak{h}^{(d)} f(x)$ with coordinates $\mathfrak{h}_{I_1 \ldots I_d} f(x)$ (here, a (d-)tensor is simply a vector indexed by $\mathcal{I}^d$). Again, we define $|\mathfrak{h}^{(d)} f(x)|$ as the Euclidean norm. For instance, $|\mathfrak{h} f(x)|$ is just the Euclidean norm of the "gradient" $\mathfrak{h} f(x)$, and $|\mathfrak{h}^{(2)} f(x)|$ is the Hilbert–Schmidt norm of the "Hessian" $\mathfrak{h}^{(2)} f(x)$. Additionally, we will use the notation $\|f\|_p$ for the p-norm of a function f (with respect to a measure μ which is clear from the context) and write

$$\|\mathfrak{h}^{(d)} f\|_p \coloneqq \left(\mathbb{E}_\mu |\mathfrak{h}^{(d)} f|^p \right)^{1/p}$$

for any $p \in (0, \infty]$ (for $p = \infty$ this is the essential supremum with respect to μ).

Next let us recall the notion of a *Poincaré inequality* and a *logarithmic Sobolev inequality* in the framework of difference operators. We say that μ satisfies a *Poincaré inequality* with constant $\sigma^2 > 0$ with respect to some difference operator Γ (in short: $\mathrm{PI}_\Gamma(\sigma^2)$) if for all $f \in L^\infty(\mu)$

$$\mathrm{Var}_\mu(f) \leq \sigma^2 \, \mathbb{E}_\mu (\Gamma f)^2, \tag{1.2}$$

where $\mathrm{Var}_\mu(f) = \mathbb{E}_\mu f^2 - (\mathbb{E}_\mu f)^2$ is the variance functional. μ is said to satisfy a *logarithmic Sobolev inequality* with constant $\sigma^2 > 0$ with respect to some difference operator Γ (in short: $\mathrm{LSI}_\Gamma(\sigma^2)$) if for all $f \in L^\infty(\mu)$

$$\mathrm{Ent}_\mu(f^2) \leq 2\sigma^2 \, \mathbb{E}_\mu (\Gamma f)^2. \tag{1.3}$$

Here, we denote by $\mathrm{Ent}_\mu(f) \coloneqq \mathrm{Ent}(f) \coloneqq \mathbb{E}_\mu f \log f - \mathbb{E}_\mu f \log \mathbb{E}_\mu f \in [0, \infty]$ the entropy functional defined for nonnegative functions.

It is well known that logarithmic Sobolev inequalities are stronger than Poincaré inequalities, i.e. if μ satisfies $\mathrm{LSI}_\Gamma(\sigma^2)$, it also satisfies a $\mathrm{PI}_\Gamma(\sigma^2)$, see for example [3] in the context of Markov semigroups, [14, Lemma 3.1] in the framework of Markov chains, or [8, Proposition 3.6], where also modified logarithmic Sobolev inequalities have been considered. We shall tacitly use this implication.

We formulate a general concentration result in Section 1.2 which may be applied to functions of the spins in Ising models, of random permutations and on slices of the hypercube. We start with an application to the Ising model with general interactions.

1.1 Ising model

In the special case of the Ising model q^n on n sites the difference operator under consideration can be written as

$$|\mathfrak{d} f|(\sigma) = \Big(\frac{1}{2}\sum_{i=1}^n (f(\sigma) - f(T_i\sigma))^2 q^n(-\sigma_i \mid \sigma_{i^c})\Big)^{1/2} \quad \text{for} \quad \sigma \in \{-1,+1\}^n,$$

where $T_i\sigma = (\sigma_{i^c}, -\sigma_i)$ is the switch operator of the i-th spin and $q^n(\cdot \mid \sigma_{i^c})$ is the conditional probability. Additionally, we have

$$|\mathfrak{h} f|(\sigma) = \Big(\frac{1}{2}\sum_{i=1}^n (f(\sigma) - f(T_i\sigma))^2\Big)^{1/2}.$$

The first result establishes a logarithmic Sobolev inequality for q^n and might be of independent interest.

Proposition 1.1. *Let q^n be the probability measure on $\{-1,+1\}^n$ defined by normalizing $\pi(\sigma) = \exp(\frac{1}{2}\sum_{i,j} J_{ij}\sigma_i\sigma_j + \sum_{i=1}^n h_i\sigma_i)$, where $\max_{i=1,\dots,n}|h_i| \le \widetilde{\alpha}$ and $J = (J_{ij})$ is symmetric, satisfies $J_{ii} = 0$ and*

$$\|J\|_{\ell^\infty \to \ell^\infty} = \max_{i=1,\dots,n} \sum_{j=1}^n |J_{ij}| \le 1 - \alpha. \tag{1.4}$$

There is a constant $C = C(\alpha, \widetilde{\alpha})$ such that for any $f : \{-1,+1\}^n \to \mathbb{R}$

$$\mathrm{Ent}_{q^n}(f^2) \le 2C\,\mathbb{E}_{q^n}|\mathfrak{d} f|^2. \tag{1.5}$$

Moreover, for any f and $p \ge 2$ it holds

$$\|f\|_p^2 - \|f\|_2^2 \le 2C(p-2)\|\mathfrak{h} f\|_p^2. \tag{1.6}$$

Equation (1.5) is a generalization of the LSI on $\{-1,+1\}^n$ equipped with the uniform measure, which corresponds to the Ising model with $J = 0$ and $h = 0$. In general, the case $J = 0$ yields n independent random variables $\sigma_1, \dots, \sigma_n$ with $\mathbb{P}(\sigma_i = 1) = \frac{1}{2}(1 + \tanh(h_i))$. Thus a uniform bound on $\|h\|_\infty$ is necessary in order for the logarithmic Sobolev constant to be stable, see e.g. [14, Theorem A.1], and there cannot be an $\widetilde{\alpha}$-independent constant C. Condition (1.4) appears in various contexts, we shall call it *Dobrushin uniqueness condition* (see for example [17, equations (2.1) and (2.2)]). The Dobrushin uniqueness condition implies an approximate tensorization result, which is central for our proof.

In a series of papers [28, 26, 25] B. Zegarlinski and D. Stroock have established the equivalence of the logarithmic Sobolev inequality and the so-called Dobrushin–Shlosman mixing condition on $\{-1,+1\}^{\mathbb{Z}^d}$. Here we prove one implication using an approximate tensorization result by K. Marton [20] for the easier case $\{-1,+1\}^n$.

From an iteration procedure we obtain the following theorem establishing tail estimates for functions of spins in the Ising model with bounded differences of higher order.

Theorem 1.2. *Let $d \in \mathbb{N}$, q^n as in Proposition 1.1 and $f : \{-1,+1\}^n \to \mathbb{R}$. Assuming the conditions*

$$\|\mathfrak{h}^{(k)} f\|_2 \le 1 \quad \textit{for all } k = 1, \dots, d-1 \tag{1.7}$$

$$\|\mathfrak{h}^{(d)} f\|_\infty \le 1, \tag{1.8}$$

there exists some constant $C = C(\alpha, \widetilde{\alpha}, d) > 0$ such that

$$\mathbb{E}_{q^n} \exp\Big(C|f - \mathbb{E}_{q^n} f|^{2/d}\Big) \le 2.$$

Especially we have

$$q^n(|f - \mathbb{E}_{q^n} f| \ge t) \le 2\exp\left(-Ct^{2/d}\right).$$

As an application, one can show concentration results for homogeneous polynomials of spins in the Ising model with bounded coefficients as follows. To begin with, let us consider the case with $h = 0$.

Theorem 1.3. *Let $d \in \mathbb{N}$, q^n be an Ising model as in Proposition 1.1 with $h = 0$. There is a constant $c = c(d, \alpha) > 0$ such that for any d-tensor $A = (a_I)_{|I|=d}$ the d-homogeneous polynomial $f(\sigma) = \sum_{|I|=d} a_I \prod_{i \in I} \sigma_i =: \sum_{|I|=d} a_I \sigma_I$ satisfies for all $t > 0$*

$$q^n(|f - \mathbb{E}_{q^n} f| \ge t) \le 2\exp\Big(-\frac{t^{2/d}}{cn\|A\|_\infty^{2/d}}\Big). \tag{1.9}$$

This result improves upon [15, Theorem 1] as well as on [11, Theorem 5] by removing the logarithmic dependence in the exponential. More precisely, in [15] it is shown that for some weakly dependent Ising models π without external field (for example, under (1.4)), every degree d polynomial f with coefficients in $[-K, K]$ satisfies

$$\pi\big(|f - \mathbb{E}_\pi f| \ge t\big) \le Cn^{d^2} \exp\Big(-\frac{t^{2/d}}{CnK^{2/d}}\Big).$$

Similar concentration inequalities have been proven in [11]; given a degree d multilinear polynomial f in the spin variables of an Ising model π (in an α-high temperature regime without external field, i.e. for models satisfying (1.4)) with coefficients in $[-K, K]$, [11, Theorem 5] states that there are two constants such that for any $t \ge CK(n\log^2 n)^{d/2}$ we have

$$\pi\big(|f - \mathbb{E}_\pi f| \ge t\big) \le 2\exp\Big(-\frac{\alpha t^{2/d}}{CK^{2/d} n\log n}\Big).$$

In contrast, (1.9) is optimal in terms of the dependence on n and with respect to the power of t. To see this, note that the uniform measure $\mu = \otimes_{i=1}^n \frac{1}{2}(\delta_{-1} + \delta_{+1})$ can also be interpreted as an Ising model. In this case, for the tensor $A = (a_{i_1,\ldots,i_d})$ with entries $a_{i_1,\ldots,i_d} = 1$ if $i_1 \neq \ldots \neq i_d$ and 0, else, we have $\mathrm{Var}(f) \sim n^d$, so that f needs to be normalized by $n^{-d/2}$. Regarding the power of t, the invariance principle in [22, Theorem 2.1] shows that for the same multilinear form f as above the distributions of $f(X)$ and $f(G)$ for a Gaussian vector G are close in Kolmogorov distance. On the other hand, the behavior of a Gaussian chaos is known (see e.g. [18]). Consequently, the decay $t^{2/d}$ is the correct one for large values of t.

By subtracting certain lower-order polynomials from the multilinear forms, we can moreover provide some possibly sharper estimates involving Hilbert–Schmidt norms. Here we also consider Ising models with an external field $h \neq 0$. Note that the major difference to the Ising model without external field is the loss of spin symmetry, i.e. the map $\sigma \mapsto -\sigma$ does not preserve the measure q^n (more precisely, the push-forward is an Ising model with external field $-h$), and hence in general all homogeneous polynomials of odd degree are not centered random variables anymore. To overcome this obstruction we can recover concentration results for polynomial functions in $\widetilde{\sigma}_i := \sigma_i - \mathbb{E}_{q^n} \sigma_i$. To this end, define the diagonal as

$$\Delta_d := \{(i_1, \ldots, i_d) \in \{1, \ldots, n\}^d : |\{i_1, \ldots, i_d\}| < d\}.$$

A tensor $A = (a_{i_1,\ldots,i_d})_{i_1,\ldots,i_d=1,\ldots,n}$ is said to have a vanishing diagonal if $a_I = 0$ for all $I \in \Delta_d$, and symmetric if $a_{i_1,\ldots,i_d} = a_{\pi(i_1),\ldots,\pi(i_d)}$ for any permutation $\pi \in S_d$. For notational convenience, let us write for any subset $I \subseteq \{1, \ldots, n\}$ the product $\sigma_I := \prod_{i \in I} \sigma_i$. We shall stick to the following four cases. Let $d \in \{1, \dots, 4\}$ and define for any d-tensor $A = (a_{i_1,\dots,i_d})$ with vanishing diagonal the functions

$$\begin{aligned} f_{1,A}(\sigma) &= \sum_{i=1}^n a_i \widetilde{\sigma}_i, \\ f_{2,A}(\sigma) &= \sum_{i,j=1}^n a_{ij}(\widetilde{\sigma}_{ij} - \mathbb{E}\,\widetilde{\sigma}_{ij}), \\ f_{3,A}(\sigma) &= \sum_{i,j,k=1}^n a_{ijk}\left(\widetilde{\sigma}_{ijk} - \mathbb{E}\,\widetilde{\sigma}_{ijk} - 3\widetilde{\sigma}_i\,\mathbb{E}(\widetilde{\sigma}_{jk})\right), \\ f_{4,A}(\sigma) &= \sum_{i,j,k,l=1}^n a_{ijkl}\left(\widetilde{\sigma}_{ijkl} - \mathbb{E}\,\widetilde{\sigma}_{ijkl} - 4\widetilde{\sigma}_i\,\mathbb{E}\,\widetilde{\sigma}_{jkl} - 6\widetilde{\sigma}_{ij}\,\mathbb{E}\,\widetilde{\sigma}_{kl} + 6\,\mathbb{E}\,\widetilde{\sigma}_{ij}\,\mathbb{E}\,\widetilde{\sigma}_{kl}\right). \end{aligned}$$

Here, the coefficients are merely of combinatorial nature, making it possible to write the polynomial in a slightly more concise form.

Theorem 1.4. *Let q^n be an Ising model as in Proposition 1.1, possibly with an external field h. Let $d \in \{1, 2, 3, 4\}$ be fixed, $A = (a_{i_1,\dots,i_d})$ a symmetric tensor with vanishing diagonal and $f_{d,A}$ as above. For some constant $C = C(\alpha, \widetilde{\alpha}, d) > 0$ we have*

$$q^n\left(|f_{d,A} - \mathbb{E}_{q^n} f_{d,A}| > t\right) \le 2\exp\Big(-\frac{t^{2/d}}{C\|A\|_{\mathrm{HS}}^{2/d}}\Big). \tag{1.10}$$

Theorem 1.4 can be extended to arbitrary $d \in \mathbb{N}$, i.e. there is a sequence of polynomials of degree d with all "partial derivatives" being centered, which allows for the same iteration of the proof as we show for $d = 1, 2, 3, 4$. However, as the formulation of the $f_{d,A}$ is cumbersome, we refer to [24].

Lastly, note that in (1.10) the constant is given by the Hilbert–Schmidt norm of A, whereas in (1.9) we have the bigger constant $n\|A\|_\infty^{2/d}$. However, (1.10) is valid for the function $f_{d,A}$, not for a general multilinear polynomial as in Theorem 1.3.

1.2 General results

The results for the Ising model are an application of our main result. For measures μ satisfying $\mathrm{LSI}_{(\mathfrak{d},\mathcal{I})}(\sigma^2)$ we derive moment inequalities which relate the $L^p(\mu)$-norms of functions f with $L^p(\mu)$ norms of their differences $|\mathfrak{d}f|$. This leads to concentration of measure of higher order for functions with bounded differences of higher order.

Theorem 1.5. *Let $d \in \mathbb{N}$, suppose that μ satisfies $\mathrm{LSI}_{(\mathfrak{d},\mathcal{I})}(\sigma^2)$ and let $f \in L^\infty(\mu)$. Assuming the conditions*

$$\|\mathfrak{h}^{(k)} f\|_2 \le \min(1, \sigma^{d-k}) \quad \textit{for all } k = 1, \dots, d-1 \tag{1.11}$$

$$\|\mathfrak{h}^{(d)} f\|_\infty \le 1, \tag{1.12}$$

we have

$$\mathbb{E}_\mu \exp\Big(\frac{1}{12\sigma^2 e}|f - \mathbb{E}_\mu f|^{2/d}\Big) \le 2.$$

Since we are interested in the asymptotic for large n, the logarithmic Sobolev constant σ^2 might depend on n. However, if σ^2 is independent of n, one may rewrite condition (1.11) as

$$\|\mathfrak{h}^{(k)} f\|_2 \le 1 \quad \text{for all } k = 1, \dots, d-1. \tag{1.13}$$

Moreover, note that here one needs to control the first $d-1$ differences, but since we require bounds for $L^2(\mu)$ norms, various tools like variance decomposition or Poincaré inequality are available to achieve this.

1.3 Outline

In Section 2 we prove elementary properties of the difference operators, and the main result Theorem 1.5 by estimating the growth of moments under an LSI. Section 3 contains examples of measures satisfying an LSI with respect to the Glauber type dynamics. In Section 3.1 we prove Theorems 1.2 and 1.3 as well as Proposition 1.1 and show by way of example that a third-order polynomial in the Ising model is concentrated around a first order polynomial as an interpretation of Theorem 1.4. Sections 3.2 and 3.3 serve to demonstrate how to interpret the LSI with respect to difference operators corresponding to $(\mathfrak{d}, \mathcal{I})$ in the cases of random walks generated by switchings on either the symmetric group and the Bernoulli-Laplace and symmetric simple exclusion process, to indicate possible further applications. Finally, in Section 4 we give a proof of an approximate tensorization result.

2 Higher order difference operators for dependent arguments

Recall that we are working on a product space of the form $S = S_1 \times \ldots S_n$ for some Polish spaces S_i. For any $I \subset \{1, \ldots, n\}$, we write $S_I \coloneqq \otimes_{i \in I} S_i$ and $S_{I^c} \coloneqq \otimes_{i \in I^c} S_i$ and denote by μ_I (and μ_{I^c}) the push-forward measure under the projection onto S_I (and S_{I^c}).

In order to make sense of the difference operators defined in the introduction, we recall the disintegration theorem in a special form for product spaces (although not endowed with product probability measures) required in our context. For the existence we refer to [12, Chapter III] and for a modern formulation to [5, Theorem 5.3.1].

Proposition 2.1 (Disintegration theorem for product spaces)**.** *Let $S_1, \ldots, S_n$ be Polish spaces, $S \coloneqq S_1 \times \ldots \times S_n$ endowed with the Borel σ-algebra and a Borel probability measure μ. For any $I \subset \{1, \ldots, n\}$ there exists a Markov kernel $(\mu(\cdot \mid x_{I^c}))_{x_{I^c} \in S_{I^c}}$ such that for any $A \in \mathcal{B}(S)$*

$$\mu(A) = \int \mu(A \mid x_{I^c}) d\mu_{I^c}(x_{I^c}).$$

Moreover, the Markov kernel can be seen as a family of probability measures on S_I and for any $f \in L^1(\mu)$ we have the decomposition formula

$$\int f d\mu = \int_{S_{I^c}} \int_{S_I} f(x_{I^c}, y_I) d\mu(y_I \mid x_{I^c}) d\mu_{I^c}(x_{I^c}).$$

Remark 2.2. The quantity $\int |\mathfrak{d}f|^2 d\mu$ can be interpreted as a Dirichlet form. Indeed, defining the Markov kernel $m(x, dy) = \frac{1}{|\mathcal{I}|} \sum_{I \in \mathcal{I}} \mu(dy \mid x_{I^c})$, it can be shown by expanding $\frac{1}{2} \iint (f(x) - f(y))^2 m(x, dy) d\mu(x)$ that if we define L as $Lf(x) = \int f(y) - f(x) m(x, dy)$ for any integrable f, this yields

$$\langle f, -Lf \rangle_\mu = \frac{1}{2|\mathcal{I}|} \sum_{I \in \mathcal{I}} \iint (f(x) - f(x_{I^c}, y_I))^2 d\mu(y_I \mid x_{I^c}) d\mu(x) = \frac{1}{|\mathcal{I}|} \int |\mathfrak{d}f|^2 d\mu.$$

Hence there is an intimate connection to a Markov chain viewpoint in the sense that there is a natural dynamics for which $\int |\mathfrak{d}f|^2 d\mu$ is the Dirichlet form.

The special case given by $\mathcal{I} = \mathcal{I}_1 \coloneqq \{\{i\}, i = 1, \ldots, n\}$ translates into the disintegration with respect to $n-1$ variables and is well known, since the dynamics corresponds to the Glauber dynamics. Here, $\mathfrak{d}f$ and $\mathfrak{h}f$ are vectors in $\mathbb{R}^n$. In probabilistic terms the definition of $\mathfrak{h}_i f(x)$ can be interpreted as an upper bound on the difference of f if one updates the coordinate i, conditional on x_{i^c}. Moreover, $\mathfrak{h}$ already appeared in the works of C. McDiarmid on concentration inequalities for functions with bounded differences, see e.g. [21]. In our situation, $\mathfrak{h}_i f(x)$ can still fluctuate and does not need to be bounded, resulting in possibly non-Gaussian concentration bounds.

In some cases, $\mathfrak{h}_I f$ is a function which depends on x_{I^c} only, e.g. if all the measures $\mu(\cdot \mid x_{I^c})$ have full support. However, we would like to stress that in general the supports do not need to agree for different x_{I^c} and thus the supremum might depend on x_I, especially in situations which incorporate some kind of exclusion. A typical example is the disintegration of the measure on $\{1, \ldots, n\}^n$ given by the push-forward of the uniform random permutation under $\sigma \mapsto (\sigma(i))_{i=1,\ldots,n}$, for which any disintegration is a Dirac measure on one point, see also Section 3.2.

In the case of product measures it is unnecessary to use the disintegration theorem. Instead, one can simply define $\mu(\cdot \mid x_{I^c}) = \otimes_{i \in I} \mu_i$. The definitions then coincide with [7].

To prove Theorem 1.5 we shall need two ingredients: a pointwise estimate on consecutive differences as well as control on the growth of moments under a logarithmic Sobolev inequality.

Lemma 2.3. *For any $f \in L^\infty(\mu)$ and any $d \geq 1$ we have the pointwise estimate*

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

Proof. Let $I \in \mathcal{I}$ and $x \in S$ be fixed and write $\|\cdot\|_{I,x}$ for the norm on $L^\infty(\mu(\cdot \mid x_{I^c}) \otimes \mu(\cdot \mid x_{I^c}))$. Using the reverse triangle inequality for $|\cdot|$ and the triangle inequality for $\|\cdot\|_{I,x}$ we obtain

$$\begin{aligned}
(\mathfrak{h}_I |\mathfrak{h}^{(d)} f|)^2 &= \frac{1}{2} \left\| \left| \mathfrak{h}^{(d)} f \right| (x_{I^c}, y_I) - \left| \mathfrak{h}^{(d)} f \right| (x_{I^c}, z_I) \right\|_{I,x}^2 \\
&\leq \frac{1}{2} \left\| \left| \mathfrak{h}^{(d)} f(x_{I^c}, y_I) - \mathfrak{h}^{(d)} f(x_{I^c}, z_I) \right|^2 \right\|_{I,x} \\
&= \frac{1}{2^{d+1}} \left\| \sum_{I_1, \ldots, I_d} (\mathfrak{h}_{I_1 \ldots I_d} f(x_{I^c}, y_I) - \mathfrak{h}_{I_1 \ldots I_d} f(x_{I^c}, z_I))^2 \right\|_{I,x} \\
&\leq \frac{1}{2^{d+1}} \sum_{I_1, \ldots, I_d} \left(\mathfrak{h}_I \mathfrak{h}_{I_1 \ldots I_d} f \right)^2 .
\end{aligned}$$

Summing over $I \in \mathcal{I}$ and taking the square root yields the result. □

By an adaption of an argument by S. G. Bobkov [6, Theorem 2.1], which in turn is based on arguments going back to L. Gross [16] as well as S. Aida and D. Stroock [3], we have the following result.

Proposition 2.4. *Let μ be a measure on a product space of Polish spaces satisfying $\mathrm{LSI}_{(\mathfrak{d},\mathcal{I})}(\sigma^2)$ with constant $\sigma^2 > 0$. Then, for any $f \in L^\infty(\mu)$ and any $p \geq 2$, we have*

$$\|f\|_p^2 - \|f\|_2^2 \leq 2\sigma^2 (p-2) \|\mathfrak{d} f\|_p^2 \tag{2.1}$$

as well as

$$\|f\|_p^2 - \|f\|_2^2 \leq 2\sigma^2 (p-2) \|\mathfrak{h} f\|_p^2. \tag{2.2}$$

Actually, up to a constant, $\mathrm{LSI}_{(\mathfrak{d},\mathcal{I})}(\sigma^2)$ is equivalent to (2.1), which has also been remarked in [6]. In the proof, we write $x_+ \coloneqq \max(x, 0)$ for the positive part of a real number, and $x_+^2 \coloneqq (x_+)^2$.

Proof. Let $p > 0$, and let f be any measurable function on an arbitrary probability space such that $0 < \|f\|_{p+\varepsilon} < \infty$ for some $\varepsilon > 0$. We have the general formula

$$\frac{d}{dp} \|f\|_p = \frac{1}{p^2} \|f\|_p^{1-p} \mathrm{Ent}(|f|^p),$$

which implies

$$\frac{d}{dp}\|f\|_p^2 = \frac{2}{p^2}\|f\|_p^{2-p}\mathrm{Ent}(|f|^p). \tag{2.3}$$

Moreover, note that for any $I \in \mathcal{I}$ we have by Proposition 2.1

$$\begin{aligned}\mathbb{E}_\mu(\mathfrak{d}_I f)^2 &= \frac{1}{2}\iint (f(x) - f(x_{I^c}, y_I))^2 d\mu(y_I \mid x_{I^c}) d\mu(x) \\ &= \iiint (f(x_{I^c}, y_I) - f(x_{I^c}, z_I))_+^2 d\mu(y_I \mid x_{I^c}) d\mu(z_I \mid x_{I^c}) d\mu_{I^c}(x_{I^c}) \\ &= \iint (f(x) - f(x_{I^c}, y_I))_+^2 d\mu(y_I \mid x_{I^c}) d\mu(x).\end{aligned}$$

Therefore, it follows that

$$\mathbb{E}_\mu|\mathfrak{d} f|^2 = \sum_{I \in \mathcal{I}} \iint (f(x) - f(x_{I^c}, y_I))_+^2 \, d\mu(y_I \mid x_{I^c}) d\mu(x) \tag{2.4}$$

Now let $p > 2$ and f be non-constant. (The assumption $\|f\|_{p+\varepsilon} < \infty$ is always true since $f \in L^\infty(\mu)$.) Applying the logarithmic Sobolev inequality (1.3) to the function $g := |f|^{p/2}$ and rewriting this in terms of (2.4) yields

$$\mathrm{Ent}(|f|^p) \le 2\sigma^2 \sum_{I \in \mathcal{I}} \iint (g(x) - g(x_{I^c}, y_I))_+^2 \, d\mu(y_I \mid x_{I^c}) d\mu(x) \tag{2.5}$$

$$= 2\sigma^2 \sum_{I \in \mathcal{I}} \iiint (g(x_{I^c}, y_I) - g(x_{I^c}, z_I))_+^2 d\mu(y_I \mid x_{I^c}) d\mu(z_I \mid x_{I^c}) d\mu_{I^c}(x_{I^c}). \tag{2.6}$$

Using the inequality $(a^{p/2} - b^{p/2})_+^2 \le \frac{p^2}{4} a^{p-2}(a-b)_+^2$ valid for all $a, b \ge 0$ and $p \ge 2$, we obtain

$$(g(x) - g(x_{I^c}, y_I))_+^2 \le \frac{p^2}{4}(|f| - |f|(x_{I^c}, y_I))_+^2 |f|^{p-2} \le \frac{p^2}{4}(f - f(x_{I^c}, y_I))^2 |f|^{p-2},$$

from which it follows in combination with (2.5) that

$$\mathrm{Ent}(|f|^p) \le p^2\sigma^2 \int |f|^{p-2} \sum_{I \in \mathcal{I}} (\mathfrak{d}_I f)^2 d\mu = p^2\sigma^2 \, \mathbb{E}_\mu |f|^{p-2} |\mathfrak{d} f|^2$$

and in combination with (2.6) that

$$\mathrm{Ent}(|f|^p) \le p^2\sigma^2 \mathbb{E}_\mu |f|^{p-2} |\mathfrak{h} f|^2.$$

Hölder's inequality with exponents $\frac{p}{2}$ and $\frac{p}{p-2}$ applied to the last integral yields

$$\mathrm{Ent}(|f|^p) \le p^2\sigma^2 \|\mathfrak{d} f\|_p^2 \|f\|_p^{p-2} \quad \text{and} \quad \mathrm{Ent}(|f|^p) \le p^2\sigma^2 \|\mathfrak{h} f\|_p^2 \|f\|_p^{p-2}.$$

Plugging this into (2.3), we arrive at the differential inequality $\frac{d}{dp}\|f\|_p^2 \le 2\sigma^2 \|\mathfrak{d} f\|_p^2$ or $\frac{d}{dp}\|f\|_p^2 \le 2\sigma^2 \|\mathfrak{h} f\|_p^2$ respectively, which after integration gives (2.1) and (2.2). □

We prove Theorem 1.5 by estimating the growth of moments in the following way. Recall that if a real-valued function f on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ satisfies

$$\|f\|_k \le \gamma k \tag{2.7}$$

for any $k \in \mathbb{N}$ and some constant $\gamma > 0$, it has sub-exponential tails, i.e.

$$\mathbb{E} \exp\Big(\frac{1}{2\gamma e}|f|\Big) \le 2. \tag{2.8}$$

Indeed, for any $c > 0$, we have

$$\mathbb{E}\exp(c|f|) = 1 + \sum_{k=1}^{\infty} c^k \frac{\mathbb{E}|f|^k}{k!} \leq 1 + \sum_{k=1}^{\infty} (c\gamma)^k \frac{k^k}{k!} \leq \sum_{k=0}^{\infty} (c\gamma e)^k,$$

where the last inequality follows from $k! \geq (\frac{k}{e})^k$ for all $k \in \mathbb{N}$. Inserting $c = \frac{1}{2\gamma e}$ we arrive at (2.8).

Proof of Theorem 1.5. We assume $\mathbb{E}_\mu f = 0$, as we can recenter f and use the properties of the difference operator to show that all inequalities hold for $f - \mathbb{E}_\mu f$ as well.

For any $p \geq 2$, applying (2.2) to $|\mathfrak{h}^{(k-1)} f|$ for $k = 1, \ldots, d$ and Lemma 2.3 in the second step gives

$$\begin{aligned}\|\mathfrak{h}^{(k-1)} f\|_p^2 &\leq \|\mathfrak{h}^{(k-1)} f\|_2^2 + 2\sigma^2(p-2)\|\mathfrak{h}|\mathfrak{h}^{(k-1)} f|\|_p^2 \\ &\leq \|\mathfrak{h}^{(k-1)} f\|_2^2 + 2\sigma^2(p-2)\|\mathfrak{h}^{(k)} f\|_p^2.\end{aligned}$$

As pointed out in the introduction, μ in particular satisfies a $\mathrm{PI}_{(\mathfrak{d},\mathcal{I})}(\sigma^2)$, and it is easily seen that this moreover implies a $\mathrm{PI}_{(\mathfrak{h},\mathcal{I})}(\sigma^2)$. Consequently, by an iteration and an application of the Poincaré inequality for $\mathfrak{h}$ we arrive at

$$\begin{aligned}\|f\|_p^2 &\leq \|f\|_2^2 + \sum_{k=1}^{d-1} (2\sigma^2(p-2))^k \|\mathfrak{h}^{(k)} f\|_2^2 + (2\sigma^2(p-2))^d \|\mathfrak{h}^{(d)} f\|_p^2 \\ &\leq \sigma^2 \|\mathfrak{h} f\|_2^2 + \sum_{k=1}^{d-1} (2\sigma^2(p-2))^k \|\mathfrak{h}^{(k)} f\|_2^2 + (2\sigma^2(p-2))^d \|\mathfrak{h}^{(d)} f\|_p^2 \\ &\leq \sum_{k=1}^{d-1} (2\sigma^2 p)^k \|\mathfrak{h}^{(k)} f\|_2^2 + (2\sigma^2 p)^d \|\mathfrak{h}^{(d)} f\|_p^2.\end{aligned}$$

Now, since $\|\mathfrak{h}^{(k)} f\|_2 \leq \min(1, \sigma^{d-k})$ for all $k = 1, \ldots, d-1$ and $\|\mathfrak{h}^{(d)} f\|_\infty \leq 1$ by assumption, we obtain

$$\|f\|_p^2 \leq \sigma^{2d} \sum_{k=1}^{d} (2p)^k \leq \frac{1}{1-(2p)^{-1}} (2\sigma^2 p)^d \leq (3\sigma^2 p)^d$$

and therefore

$$\|f\|_p \leq (3\sigma^2 p)^{d/2}.$$

Moreover, for all $p < 2$, by Hölder's and Jensen's inequality we have

$$\|f\|_p \leq \|f\|_2 \leq (6\sigma^2)^{d/2}.$$

Considering $p = 2k/d$ for $k \in \mathbb{N}$ yields $\||f|^{2/d}\|_k \leq 6\sigma^2 k$. In view of (2.7), this completes the proof. □

3 Applications

3.1 Ising model

The *Ising model* is a probability measure on its configuration space $\{-1,+1\}^n$. Let $J = (J_{ij})$ be a symmetric matrix with vanishing diagonal (the *interaction matrix*), $h \in \mathbb{R}^n$ (the *external field*) and define $\pi : \{-1,+1\}^n \to \mathbb{R}$ via

$$\pi(\sigma) = \exp\Big(\frac{1}{2}\langle\sigma, J\sigma\rangle + \langle h, \sigma\rangle\Big) = \exp\Big(\frac{1}{2}\sum_{i,j} J_{ij}\sigma_i\sigma_j + \sum_{i=1}^{n} h_i\sigma_i\Big).$$

Equip $\{-1,+1\}^n$ with the Gibbs measure $q^n(\sigma) = Z^{-1}\pi(\sigma)$, with Z being the normalization constant. For each $i \in \{1,\dots,n\}$ denote by $T_i : \{-1,+1\}^n \to \{-1,+1\}^n$ the operator which switches the sign of the i-th coordinate. Here, the factor $\frac{1}{2}$ corresponds to the fact that we made the matrix symmetric, i.e. $J = \widetilde{J} + \widetilde{J}^T$, where $\widetilde{J}$ is an upper triangular matrix. This is consistent with the Curie-Weiss model in [10, Example 2.1] or [9], but not with [15].

We want to use an approximate tensorization of entropy result proven in [20] and the results from the last section to obtain concentration inequalities for polynomials in the spin variables of Ising models which are sufficiently close to being product measures. In [20], the author proves an approximate tensorization property of the relative entropy with respect to a fixed measure q^n in the sense that

$$\mathrm{Ent}_{q^n}(f) \le \frac{C}{\beta}\sum_{i=1}^n \int \mathrm{Ent}_{q^n(\,\cdot\,|y_{i^c})}(f(y_{i^c},\cdot))dq^n(y).$$

Here, β is the minimal conditional probability and C is a constant depending on q^n. However in the proof of [20, Theorem 1] there is a small oversight, hence (and for the sake of completeness) we include a full exposition of the proof in Section 4, see Theorem 4.2. Moreover, [20, Theorem 2] replaces one of the conditions of [20, Theorem 1] by another condition, which is easier to check, see Theorem 4.2 (iii). Indeed, this condition holds via bounds on the operator norm of a coupling matrix $A = (A_{ik})_{i\neq k}$ defined as any matrix such that

$$\sup_{x,z\in\{-1,+1\}^n : x_{k^c} = z_{k^c}} d_{TV}\left(q^n(\,\cdot \mid x_{i^c}), q^n(\,\cdot \mid z_{i^c})\right) \le A_{ik}.$$

Provided that $\|A\|_{\ell^2\to\ell^2} < 1$, an approximate tensorization property holds with $C = (1-\|A\|_{\ell^2\to\ell^2})^{-2}$. Our aim is to prove that these properties hold.

The conditional probabilities of the Ising model $q^n(\,\cdot \mid \sigma_{i^c})$ are given by

$$q^n(\sigma_i \mid \sigma_{i^c}) = \frac{1}{2}\Big(1 + \tanh\big(\sigma_i \sum_j J_{ij}\sigma_j + h_i\sigma_i\big)\Big). \tag{3.1}$$

Lemma 3.1. *Let q^n be an Ising model with an interaction matrix J satisfying $J_{ii} = 0$ and $\|J\|_{\ell^\infty\to\ell^\infty} \le 1-\alpha$. Then $|J| = (|J_{ij}|)_{i,j}$ can be used as a coupling matrix and thus*

$$\||J|\|_{\ell^2\to\ell^2} \le \|J\|_{\ell^\infty\to\ell^\infty} \le 1-\alpha. \tag{3.2}$$

Moreover, if $|h| \le \widetilde{\alpha}$, then for some $c(\alpha,\widetilde{\alpha})$ it holds

$$q^n(\,\cdot \mid \sigma_{i^c}) \in \left(c(\alpha,\widetilde{\alpha}), 1 - c(\alpha,\widetilde{\alpha})\right).$$

Proof. Let $i \neq k$ be fixed and $y, z \in \{-1,+1\}^n$ differ in the k-th coordinate only, i.e. $y = T_k z$. Define $\sigma \coloneqq (z_{i^c}, 1)$ and $m_i(\sigma) \coloneqq \sigma_i \sum_j J_{ij}\sigma_j + h_i\sigma_i$. We have by equation (3.1) and the 1-Lipschitz property of the $\tanh$

$$d_{TV}\left(q^n(\cdot \mid y_{i^c}), q^n(\,\cdot \mid z_{i^c})\right) \le \frac{1}{2}|m_i(\sigma) - m_i(T_k\sigma)| = |J_{ki}|.$$

The inequality (3.2) holds since

$$\||J|\|_{\ell^2\to\ell^2} \le \sqrt{\|J\|_{\ell^\infty\to\ell^\infty}\|J^T\|_{\ell^\infty\to\ell^\infty}} \le 1-\alpha,$$

where we used the estimate $|\lambda_i(JJ^T)| \le \|JJ^T\| \le \|J\|\|J^T\|$ for any operator norm, and that J is symmetric. The second statement follows easily by using equation (3.1) and the estimate

$$\max_{\sigma\in\{-1,+1\}^n}\max_{i=1,\dots,n}\|m_i\|_\infty \le \|J\|_{\ell^\infty\to\ell^\infty} + \|h\|_\infty. \qquad \square$$

We are now ready to prove Proposition 1.1, i.e. the logarithmic Sobolev inequality (1.5) and the moment inequality (1.6).

Proof of Proposition 1.1. We can apply Lemma 3.1 to see that by Theorem 4.2(iii) we have for some $\beta = \beta(\alpha, \widetilde{\alpha})$

$$\mathrm{Ent}_{q^n}(f^2) \leq \frac{1}{\alpha^2 \beta} \sum_{i=1}^n \int \mathrm{Ent}_{q^n(\,\cdot\,|y_{i^c})}(f^2(y_{i^c}, \cdot)) dq^n(y), \tag{3.3}$$

so that it remains to find a uniform bound for the entropy given y_{i^c}. To this end, fix $i \in \{1, \ldots, n\}$, $y_{i^c} \in \{-1,+1\}^{n-1}$ and to lighten notation write $q(\cdot) \coloneqq q^n(\cdot \mid y_{i^c})$. q is a measure on $\{-1,+1\}$ and the Markov chain given by $K(x_0, x_1) = q(x_1)$ is reversible w.r.t. q. By [14, Theorem A.1] (see also [8, Example 3.8]) (K, q) satisfies a logarithmic Sobolev inequality with a constant depending on $q_* = \min_{x \in \{-1,+1\}} q(x)$. However, this constant is bounded from below by Lemma 3.1 uniformly in $y_{i^c} \in \{-1,+1\}^{n-1}$ and $n \in \mathbb{N}$. Thus, we have

$$\mathrm{Ent}_q(f^2) \leq C \iint (f(x) - f(y))^2 dq(x) dq(y). \tag{3.4}$$

Inserting (3.4) into (3.3) yields for some constant $C = C(\alpha, \widetilde{\alpha})$

$$\mathrm{Ent}_{q^n}(f^2) \leq C \sum_{i=1}^n \iiint (f(y_{i^c}, x) - f(y_{i^c}, y))^2 dq(x) dq(y) dq^n(y) = 2C\, \mathbb{E}_{q^n} |\mathfrak{d} f|^2,$$

which proves the LSI for q^n. Equation (1.6) is now a consequence of Proposition 2.4. □

Proof of Theorem 1.2. Theorem 1.2 is an application of Theorem 1.5, since q^n satisfies a logarithmic Sobolev inequality with respect to $\mathcal{I} = \{\{1\}, \ldots, \{n\}\}$. □

One can calculate using the reverse triangle inequality and the monotonicity of the square function as in the proof of Lemma 2.3 that for any $i_1 \neq i_2 \neq \ldots \neq i_d$

$$(\mathfrak{h}_{i_1 \ldots i_d} f)^2 \leq 2^{-d} \big| \big(\prod_{j=1}^d (\mathrm{Id} - T_{i_j}) \big) f \big|^2 \tag{3.5}$$

holds, which also implies

$$|\mathfrak{h}^{(d)} f| \leq \Big(2^{-d} \sum_{|I|=d} \Big(\Big(\prod_{i \in I} (\mathrm{Id} - T_i) \Big) f \Big)^2 \Big)^{1/2}. \tag{3.6}$$

Here, we define for any $f : \{-1,+1\}^n \to \mathbb{R}$ the function $T_i f \coloneqq f \circ T_i$, and define $T_{i_1 \ldots i_d}$ via iteration. Note that on the right-hand side we deliberately chose summing over $|I| = d$ instead of $i_1, \ldots, i_d$, since $\mathfrak{h}_{i_1 \ldots i_d} f = 0$ if $i_j = i_k$ for some $j \neq k$.

For the operator appearing on the right-hand side of equation (3.6), it was already shown in [23] (see also [7, Lemma 2.2]) that the chain of pointwise inequalities from Lemma 2.3 holds.

Proof of Theorem 1.3. Let $f = \sum_{|I|=d} a_I \sigma_I = \sum_{|I|=d} a_I \prod_{i \in I} \sigma_i$ be a d-th order homogeneous polynomial and without loss of generality assume $\|A\|_\infty = 1$ for the tensor $A = (a_I)$. As in the proof of Proposition 1.1 we obtain

$$\|f - \mathbb{E}_{q^n} f\|_p^2 \leq \sum_{k=1}^{d-1} p^k (2C(\alpha))^k \|\mathfrak{h}^{(k)} f\|_2^2 + p^d (2C(\alpha))^d \|\mathfrak{h}^{(d)} f\|_p^2. \tag{3.7}$$

Now for any $k \in \{1, \ldots, d-1\}$ we have $(\mathfrak{h}_{i_1,\ldots,i_k} f)^2 \le 2^k (\sum_{\substack{|I|=d-k \\ i_1,\ldots,i_k \notin I}} a_{I \cup i_1,\ldots,i_k} \sigma_I)^2$ by equation (3.5). It follows from [15, Lemma 3.1] that

$$\|\mathfrak{h}^{(k)} f\|_2^2 = \sum_{i_1,\ldots,i_k} \|\mathfrak{h}_{i_1,\ldots,i_k} f\|_2^2 \le c_k n^d,$$

since for each fixed $i_1, \ldots, i_k$ the integrand is a polynomial of degree at most $2(d-k)$ with coefficients bounded by 1. Hence ultimately we obtain for any $p \ge 2$

$$\|f - \mathbb{E}_{q^n} f\|_p^2 \le n^d (2C(\alpha))^d \max(1, c_1, \ldots, c_{d-1}) \sum_{k=1}^d p^k,$$

which can be rewritten as

$$\|n^{-d/2}(f - \mathbb{E}_{q^n} f)\|_p \le C(\alpha, d) p^{d/2}$$

with $C(\alpha, d) = (2C(\alpha))^{d/2} \max(1, c_1, \ldots, c_{d-1})^{1/2} d^{1/2}$, which is equivalent to the exponential integrability of $|n^{-d/2}(f - \mathbb{E}_{q^n} f)|^{2/d}$, and the result readily follows. □

Remark 3.2. Actually equation (3.7) can be used to provide a more accurate estimate of the tail properties of $f - \mathbb{E}_{q^n} f$ in the spirit of [1, Theorem 7] and [2, Theorem 3.3]. It is based on the idea that Chebyshev's inequality yields for any $p \ge 1$

$$q^n(|f - \mathbb{E}_{q^n} f| \ge e\|f - \mathbb{E}_{q^n} f\|_p) \le \exp(-p). \tag{3.8}$$

First, observe that by taking the square root and using the subadditivity in equation (3.7) we obtain

$$e\|f - \mathbb{E}_{q^n} f\|_p \le e\Big(\sum_{k=1}^{d-1} (2C(\alpha)p\|\mathfrak{h}^{(k)} f\|_2^{2/k})^{k/2} + (2C(\alpha)p\|A\|_{\mathrm{HS}}^{2/d})^{d/2}\Big).$$

Now consider the function

$$\eta_f(t) := \min\left(\frac{t^{2/d}}{2C(\alpha)\|A\|_{\mathrm{HS}}^{2/d}}, \min_{k=1,\ldots,d-1} \frac{t^{2/k}}{2C(\alpha)\|\mathfrak{h}^{(k)} f\|_2^{2/k}}\right),$$

where $A = (a_I)$ is the tensor of coefficients. If $\eta_f(t) \ge 2$ holds, a short calculation shows that we have $e\|f - \mathbb{E}_{q^n} f\|_{\eta_f(t)} \le (de)t$. Applying equation (3.8) to $p = \eta_f(t)$ yields

$$q^n(|f - \mathbb{E}_{q^n} f| \ge (de)t) \le q^n(|f - \mathbb{E}_{q^n} f| \ge e\|f - \mathbb{E}_{q^n} f\|_{\eta_f(t)}) \le \exp\left(-\eta_f(t)\right),$$

so that combined with the trivial estimate (for $p \le 2$) we obtain

$$q^n(|f - \mathbb{E}_{q^n} f| \ge (de)t) \le e^2 \exp(-\eta_f(t)).$$

To remove the de factor, rescale f by de and use the estimate $\eta_{(de)f}(t) \ge \frac{\eta_f(t)}{(de)^2}$.

Proof of Theorem 1.4. Let us prove by induction that for $p \ge 2$ we have for $f = f_{d,A}$

$$\|f - \mathbb{E}_{q^n} f\|_p^2 \le c_d p^d \|A\|_{\mathrm{HS}}^2. \tag{3.9}$$

First, for $d = 1$ this is clear since $f := f_{1,A}(X) = \sum_i a_i \widetilde{X}_i$ and by equation (1.6) we have for $p \ge 2$

$$\|f - \mathbb{E}_{q^n} f\|_p^2 \le 2Cp\|\mathfrak{h} f\|_p^2 = 2Cp\|A\|_{\mathrm{HS}}^2.$$

For $d \geq 2$ use (1.6) again to get

$$\|f - \mathbb{E}_{q^n} f\|_p^2 \leq 2Cp\|\mathfrak{h} f\|_p^2 = 2Cp\|\sum_{i=1}^n (\mathfrak{h}_i f)^2\|_{p/2} \leq 2Cp \sum_{i=1}^n \|\mathfrak{h}_i f\|_p^2$$
$$\leq 2Cp \sum_{i=1}^n c_{d-1} p^{d-1} \|A^{(i)}\|_{\mathrm{HS}}^2 = 2C_d p^d \sum_{i=1}^n \|A^{(i)}\|_{\mathrm{HS}}^2 = 2C_d p^d \|A\|_{\mathrm{HS}}^2.$$

Here we have used the fact for any $f_{d,A}$ we have $\mathfrak{h}_i f = c_d |f_{d-1,A^{(i)}} - \mathbb{E} f_{d-1,A^{(i)}}|$, where $(A^{(i)})_{i_1,\ldots,i_{d-1}} = A_{i_1,\ldots,i_{d-1},i}$ is a symmetric $(d-1)$-tensor with vanishing diagonal.

From equation (3.9) the first inequality easily follows as already shown in the proof of Theorem 1.5. □

In the absence of a magnetic field, i.e. for $h = 0$, the $d = 2$ case translates into the concentration inequality

$$q^n(|f - \mathbb{E}_{q^n} f| > t) \leq 2\exp\Big(-\frac{1}{C}\frac{t}{\|A\|_{\mathrm{HS}}}\Big)$$

for $f(\sigma) = \frac{1}{2}\langle \sigma, A\sigma\rangle = \sum_{i<j} a_{ij}(\sigma_i\sigma_j - \mathbb{E}\,\sigma_i\sigma_j)$ and some symmetric matrix $A = (a_{ij})$. It is also possible to prove this using Theorem 1.2 by showing that

$$\|\mathfrak{h} f\|_2 \leq 4C\|A\|_{\mathrm{HS}} \quad \text{and} \quad \|\mathfrak{h}^{(2)} f\|_\infty \leq 4C\|A\|_{\mathrm{HS}}.$$

The $d = 3$ case has an interesting interpretation since it shows that a polynomial of order three is not concentrated around its mean (which in this case would be zero), but around a first order correction. For example, for the 3-tensor $A = (a_{ijk})$ given by $a_{ijk} = n^{-3/2}$ whenever the three indices are distinct, and 0 otherwise, we obtain concentration inequalities for

$$f(\sigma) = n^{-3/2}\sum_{i\neq j\neq k}\sigma_i\sigma_j\sigma_k - 3n^{-3/2}\sum_{i=1}^n \sigma_i \sum_{j\neq k: j\neq i, k\neq i}\mathbb{E}\,\sigma_j\sigma_k =: f_3(\sigma) + f_1(\sigma).$$

Here, the correction term f_1 is sub-Gaussian, as a short calculation shows that we have $|\mathfrak{h} f_1|^2 = 2n^{-3}\sum_{i=1}^n c_i^2$ for $c_i := 3\,\mathbb{E}\sum_{j\neq k: j\neq i, k\neq i}\sigma_j\sigma_k$, and [15, Lemma 3.1] yields $c_i^2 \leq Cn^2$ for any $i \in \{1,\ldots,n\}$.

More generally, we show concentration inequalities for $f_A(\sigma) = \sum_{i\neq j\neq k} a_{ijk}\sigma_i\sigma_j\sigma_k$ by approximating it with a linear term.

Corollary 3.3. *Let $A = (a_{ijk})$ be a symmetric 3-tensor with vanishing diagonal and define $f(\sigma) = \sum_{i,j,k} a_{ijk}\sigma_i\sigma_j\sigma_k$. We have for any $t \geq 0$ for some constant $C = C(\alpha)$*

$$q^n(|f| \geq t) \leq 4\exp\Bigg(-\frac{1}{C}\min\Bigg(\frac{t^2}{\sum_{i=1}^n\big(\sum_{j,k} a_{ijk}\,\mathbb{E}_{q^n}\sigma_j\sigma_k\big)^2}, \frac{t^{2/3}}{\|A\|_{\mathrm{HS}}^{2/3}}\Bigg)\Bigg).$$

Proof. Let $\widetilde{f} := f - f_1$ for $f_1(\sigma) := 3\sum_{i=1}^n \sigma_i \sum_{j,k} a_{ijk}\,\mathbb{E}_{q^n}\sigma_j\sigma_k$. Clearly, $\widetilde{f} = f_{3,A}$ for the 3-tensor A, so that we have for any $t \geq 0$

$$q^n(|\widetilde{f}| \geq t) \leq 2\exp\Big(-\frac{1}{C}\frac{t^{2/3}}{\|A\|_{\mathrm{HS}}}\Big).$$

On the other hand, we have by Theorem 1.2 an estimate for f_1 as

$$q^n(|f_1| \geq t) \leq 2\exp\Big(-\frac{1}{C}\frac{t^2}{\sum_{i=1}^n\big(\sum_{j,k} a_{ijk}\,\mathbb{E}_{q^n}\sigma_j\sigma_k\big)^2}\Big).$$

The proof now follows by standard arguments and the relation $\{|f| \geq t\} \subset \{|\widetilde{f}| \geq t/2\} \cup \{|f_1| \geq t/2\}$. □

Similar arguments also work for $d = 4$, with terms of order t^2 and t by an approximation with $f_{4,A}$.

3.2 Random permutations

Next we consider random permutations which we shall describe as a probability measure on $\{1,\dots,n\}^n$, more precisely as the uniform measure σ_n on $S_n := \{(x_1,\dots,x_n) : x_i \neq x_j \text{ for all } i \neq j\}$. With this definition it fits into our framework.

Since conditioning on $n-1$ variables is useless (as the disintegrated measure will be a Dirac measure on the remaining element x_i and thus an LSI cannot hold for either difference operator), we shall work with $\mathcal{I}_2 := \{I \subset \{1,\dots,n\}, |I| = 2\}$. In this case, it is easy to see that for any $I = \{i,j\} \in \mathcal{I}_2$ the conditional probability is given by $\mu(\cdot \mid x_{I^c}) = \frac{1}{2}(\delta_{(x_i,x_j)} + \delta_{(x_j,x_i)})$, with x_i, x_j being the two elements distinct from any element in x_{I^c}. So denoting by $\tau_I := \tau_{ij} : S_n \to S_n$ the function which switches the i-th and j-th entry, we can rewrite the difference operator as

$$\mathfrak{d}_I f(x)^2 = \frac{1}{2}\int (f(x) - f(x_{I^c}, y_I))^2 d\sigma_n(y_I \mid x_{I^c}) = \frac{1}{4}(f(x) - f(\tau_I x))^2,$$
$$\mathfrak{h}_I f(x)^2 = \frac{1}{2}|f(x) - f(\tau_I x)|^2.$$

We can rephrase [19, Theorem 1] in the following way.

Lemma 3.4. *Consider (S_n, σ_n) and $\mathcal{I} = \mathcal{I}_2$. Then there exists a constant $c > 0$ independent of n such that*

$$\mathrm{Ent}_{\sigma_n}(f^2) \leq 2c\frac{\log n}{n}\int |\mathfrak{d} f|^2 d\sigma_n,$$

i.e. (S_n, σ_n) satisfies $\mathrm{LSI}_{(\mathfrak{d},\mathcal{I})}(\frac{c \log n}{n})$.

Proof. The proof is rewriting the statement of [19, Theorem 1] in our notation, using the fact that the conditional measures are two-point Dirac measures. More precisely, we have

$$\begin{aligned}
\mathrm{Ent}_{\sigma_n}(f^2) &\leq c\log n \frac{1}{2n}\,\mathbb{E}_{\sigma_n}\sum_I (f(\tau_I x) - f(x))^2 \\
&= 2c\frac{\log n}{n}\sum_I \iiint (f(x_{I^c}, y_I) - f(x_{I^c}, z_I))^2 d\sigma_n(y_I \mid x_{I^c}) d\sigma_n(z_I \mid x_{I^c}) d\sigma_n(x) \\
&= 2c\frac{\log n}{n}\int |\mathfrak{d} f|^2 d\sigma_n,
\end{aligned}$$

where the summation is over all $I = (i,j)$. □

3.3 Bernoulli-Laplace and symmetric simple exclusion process

There are two other Markov chains, whose Dirichlet form can be described in terms of a subset $\mathcal{I}$ and difference operators $\mathfrak{d}_I$, which are the Bernoulli-Laplace model and the symmetric simple exclusion process.

More specifically, let $S_n := \{0,1\}^n$ and define $C_{n,r} = \{x \in \{0,1\}^n : \sum_i x_i = r\}$ (a slice of the hypercube), equipped with the uniform measure $\mu_{n,r}$ on $C_{n,r}$. Furthermore, define the two generators acting on $f : C_{n,r} \to \mathbb{R}$ as

$$K_{n,r} f(\eta) = \sum_{i,j} \eta_i(1-\eta_j)(f(\tau_{ij}\eta) - f(\eta)),$$

which is the generator of the so-called Bernoulli-Laplace model, and

$$L_{n,r} f(\eta) = \sum_{i=1}^n (f(\tau_{i,i+1}\eta) - f(\eta)),$$

called the symmetric simple exclusion process, where $\tau_{ij} : S_n \to S_n$ switches the i-th and the j-th coordinate and $\tau_{n,n+1} \coloneqq \tau_{n,1}$. In [19, Theorem 4, Theorem 5] sharp logarithmic Sobolev constants are derived with respect to the Dirichlet form $D^K_{n,r}(f) = -\mathbb{E}_{\mu_{n,r}} fK_{n,r}f$ and $D^L_{n,r}(f) = -\mathbb{E}_{\mu_{n,r}} f_{n,r}f$ (although with different normalizations), and these correspond to LSIs with respect to $\mathfrak{d}$ in the following way.

Lemma 3.5. *For $\mathcal{I} = \mathcal{I}_{2,<} = \{(i,j) : i < j\}$ we have $\int |\mathfrak{d}f|^2 d\mu_{n,r} = D^K_{n,r}(f)$ and for $\mathcal{I} = \mathcal{I}_1 = \{(i,i+1) : i \in \{1,\dots,n\}\}$ we obtain $\int |\mathfrak{d}f|^2 d\mu_{n,r} = D^L_{n,r}(f)$.*

As a consequence, $\mu_{n,r}$ satisfies a logarithmic Sobolev inequality with respect to $(\mathfrak{d}, \mathcal{I}_{2,<})$ with constant $c\frac{\log \frac{n^2}{r(n-r)}}{n}$ and a logarithmic Sobolev inequality with constant cn^2 with respect to $(\mathfrak{d}, \mathcal{I}_1)$, where c is a constant independent of n and r.

Proof. For brevity's sake, fix n, r and drop the subscripts n, r, i.e. write D^L for $D^L_{n,r}$, D^K for $D^K_{n,r}$ and μ for $\mu_{n,r}$. For $\mathcal{I} = \mathcal{I}_{2,<}$ let $I = (i,j)$ be given and observe that

$$\mu(\,\cdot \mid x_{I^c}) = \begin{cases} \delta_{(1,1)} & \sum_k (x_{I^c})_k = r-2 \\ \frac{1}{2}(\delta_{(0,1)} + \delta_{(1,0)}) & \sum_k (x_{I^c})_k = r-1 \\ \delta_{(0,0)} & \sum_k (x_{I^c})_k = r, \end{cases}$$

and thus

$$\int |\mathfrak{d}f|^2 d\mu = \sum_{I \in \mathcal{I}_{2,<}} \int (\mathfrak{d}_I f)^2 d\mu(\eta) = \frac{1}{2} \sum_I \int (f(\eta) - f(\tau_I \eta))^2 \eta_i (1-\eta_j) d\mu(\eta) = D^K(f).$$

The second case is just a special case of the above one, which yields

$$\int |\mathfrak{d}f|^2 d\mu = \sum_{i=1}^n \int (\mathfrak{d}_i f)^2 d\mu = \frac{1}{2} \sum_{i=1}^n \int (f(\eta) - f(\tau_{i,i+1}\eta))^2 d\mu = D^L(f).$$

Note that we omit $\eta_i(1-\eta_{i+1})$ since otherwise we obtain $\tau_{i,i+1}\eta = \eta$.

The LSI follows from [19, Theorem 4, Theorem 5], taking into account the missing normalization. □

4 Approximate tensorization of the relative entropy in finite product spaces

In this section we reformulate and provide a complete proof of a result by K. Marton [20] and moreover rewrite it in the terms of entropy (of functions) instead of relative entropy of measures. To this end, let $\mathcal{X}$ be a finite set, $\mathcal{X}^n$ its n-fold product and fix a probability measure q^n on $\mathcal{X}^n$, which need not be a product measure. Define the total variation distance

$$d_{TV}(\mu, \nu) \coloneqq \sup_{A \subset \mathcal{X}^n} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in \mathcal{X}^n} |\mu(\{x\}) - \nu(\{x\})|,$$

the relative entropy

$$H(\mu \mid\mid \nu) = \int \frac{d\mu}{d\nu} \log \frac{d\mu}{d\nu} d\nu \quad \text{for} \quad \mu \ll \nu,$$

and the Wasserstein-2-type distance

$$W_2(\mu, \nu) \coloneqq \inf_{\pi \in C(\mu,\nu)} \Big(\sum_{i=1}^n \pi(x_i \neq y_i)^2 \Big)^{1/2}.$$

Here $C(\mu,\nu)$ is the set of all couplings of μ and ν, i.e. probability measures π on $\mathcal{X}^n \times \mathcal{X}^n$ with marginals μ and ν. Note that the infimum in the definition is always attained, since $C(\mu,\nu)$ is a compact subset of $\mathcal{P}(\mathcal{X}^n \times \mathcal{X}^n)$ equipped with the weak topology and the map $\pi \mapsto \left(\sum_{i=1}^n \pi(x_i \neq y_i)^2\right)^{1/2}$ is lower semicontinuous. This fact and the gluing lemma for measures with a common marginal can be used to prove that W_2 is a distance function on $\mathcal{P}(\mathcal{X}^n)$, see for example [27, Chapter 6] for a similar line of reasoning and [4, Theorem 2.1] for the gluing lemma. Denote by μ_i, ν_i the pushforward measure under the projection onto the i-th coordinate of μ and ν respectively. By the subadditivity of the square root (for the upper bound for W_2) as well as the fact that every coupling π on $\mathcal{X}^n \times \mathcal{X}^n$ of μ,ν induces (by the projection onto the coordinates x_i, y_i) a coupling π_i of μ_i, ν_i, we obtain

$$\Big(\sum_{i=1}^n d_{TV}^2(\mu_i,\nu_i)\Big)^{1/2} \leq W_2(\mu,\nu) \leq \sqrt{n} d_{TV}(\mu,\nu). \tag{4.1}$$

We will need the following lemma, which is also found in [20, Lemma 2] with a slightly worse constant.

Lemma 4.1. *Let q be a measure on a finite space $\mathcal{X}$ and let $\beta_q \coloneqq \inf_{x \in \mathcal{X}_+} q(x)$, where $\mathcal{X}_+ \coloneqq \{x \in \mathcal{X} : q(x) > 0\}$. For any measure $p \ll q$ we have*

$$H(p \,||\, q) \leq 2\beta_q^{-1} d_{TV}^2(p,q).$$

Proof. The shifted logarithm $f(x) \coloneqq \log(1+x)$ is a concave function on $(-1,\infty)$, so that for any $x \geq 0$ we have $f(x) \leq f'(0)x = x$. Rewrite $\frac{p}{q} = 1 + \frac{p-q}{q}$ to obtain

$$\begin{aligned} H(p \,||\, q) &= \sum_{\mathcal{X}_+} q(x) \left(1 + \frac{p(x)-q(x)}{q(x)}\right) f\left(\frac{p(x)-q(x)}{q(x)}\right) \\ &\leq \sum_{\mathcal{X}_+} q(x) \left(1 + \frac{p(x)-q(x)}{q(x)}\right) \frac{p(x)-q(x)}{q(x)} = \sum_{\mathcal{X}_+} \frac{(p(x)-q(x))^2}{q(x)} \\ &\leq \beta_q^{-1} \sum_{\mathcal{X}_+} (p(x)-q(x))^2 \leq \beta_q^{-1} d_{TV}(p,q) \sum_{\mathcal{X}} |p(x)-q(x)| = 2\beta_q^{-1} d_{TV}^2(p,q). \quad \square \end{aligned}$$

Unfortunately, the factor $2\beta_q^{-1}$ cannot be removed. To see this, consider $\mathcal{X} = \{0,1\}$ and let $q(0) = 1 - q(1) = \alpha_1$ with $\alpha_1 \in (0,1/2)$ (so that $\beta = \alpha_1$) and consider the family of measures p_ε given by $p_\varepsilon(0) = \alpha_1 + \varepsilon$; an easy calculation yields $d_{TV}^2(p_\varepsilon, q) = \varepsilon^2$ and $H(p_\varepsilon \,||\, q) \sim 2\beta^{-1}\varepsilon^2$ and thus the constant is optimal.

As a consequence of Lemma 4.1, for any measure q on a finite space $\mathcal{X}$ and any p satisfying $p \lll q$ we have

$$d_{TV}^2(p,q) \leq \frac{1}{2} H(p \,||\, q) \leq \beta_q^{-1} d_{TV}^2(p,q).$$

The following theorem is the main result of this section. We use the same notations as before, i.e. for any measure p^n on $\mathcal{X}^n$ and any subset $I \subset \{1,\dots,n\}$ we denote by $p^n(\cdot \mid y_{I^c})$ the conditional probability measure on $\mathcal{X}_I$ given by conditioning on y_{I^c}. For $I = \{1,\dots,n\}$ we set $p^n(\cdot \mid y_{I^c}) = p^n$, and for any $i = 1,\dots,n$ p_i^n denotes the induced measure on the i-th coordinate.

Theorem 4.2. *Let q^n be a measure with full support on $\mathcal{X}^n$ and define the lower bound $\beta \coloneqq \min_{i=1,\dots,n} \min_{x \in \mathcal{X}^n} q^n(x_i \mid x_{i^c})$*

(*i*) *Let p^n a probability measure and assume that for all subsets $I \subseteq \{1,\dots,n\}$ and all $y_{I^c} \in \mathcal{X}_{I^c}$ we have*

$$W_2^2(p^n(\cdot \mid y_{I^c}), q^n(\cdot \mid y_{I^c})) \leq C \sum_{i \in I} \mathbb{E}_{p^n(\cdot \mid y_{I^c})}\, d_{TV}^2(p^n(\cdot \mid y_{i^c}), q^n(\cdot \mid y_{i^c})), \tag{4.2}$$

then

$$H(p^n \mid\mid q^n) \leq \frac{C}{\beta} \sum_{i=1}^n \mathbb{E}_{p^n} H(p^n(\cdot \mid y_{i^c}) \mid\mid q^n(\cdot \mid y_{i^c})) \tag{4.3}$$

(ii) *If f denotes the density of p^n with respect to q^n, this can be rewritten as*

$$\mathrm{Ent}_{q^n}(f) \leq \frac{C}{\beta} \sum_{i=1}^n \int \mathrm{Ent}_{q^n(\cdot \mid y_{i^c})}(f(y_{i^c}, \cdot)) dq^n(y). \tag{4.4}$$

(iii) *Assume that the coupling matrix $A = (a_{ij})_{i \neq j}$ (see Section 3.1) of q^n satisfies the condition $\|A\|_{\ell^2 \to \ell^2} < 1$. Then* (4.2) *holds with $C = (1 - \|A\|_{\ell^2 \to \ell^2})^{-2}$, so that also* (4.3) *and* (4.4) *hold with the same constant.*

Proof. First note that $\beta > 0$ due to the assumption of q^n having full support.

(i): We will prove the theorem by induction. In the case $n = 1$ there is nothing to prove if one interprets $q^1(\cdot \mid y_{1^c}) = q$. For $n \geq 2$, using the decomposition theorem for the relative entropy (see for example [13, Theorem D.13]) gives

$$H(p^n \mid\mid q^n) = \frac{1}{n} \sum_{i=1}^n H(p_i^n \mid\mid q_i^n) + \frac{1}{n} \sum_{i=1}^n \int H(p^n(\cdot \mid y_i) \mid\mid q^n(\cdot \mid y_i)) dp^n(y). \tag{4.5}$$

We will treat the two terms separately. For the first term, using the estimate $H(p_i^n \mid\mid q_i^n) \leq 2\beta^{-1} d_{TV}^2(p_i^n, q_i^n)$ from Lemma 4.1, the inequalities (4.1) and (4.2) as well as Pinsker's inequality gives

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n H(p_i^n \mid\mid q_i^n) &\leq \frac{2}{\beta n} \sum_{i=1}^n d_{TV}^2(p_i^n, q_i^n) \leq \frac{2}{\beta n} W_2^2(p^n, q^n) \\
&\leq \frac{2C}{\beta n} \sum_{i=1}^n \int d_{TV}^2(p^n(\cdot \mid y_{i^c}), q^n(\cdot \mid y_{i^c})) dp^n(y) \\
&\leq \frac{C}{\beta n} \sum_{i=1}^n \int H(p^n(\cdot \mid y_{i^c}) \mid\mid q^n(\cdot \mid y_{i^c})) dp^n(y).
\end{aligned}$$

For the second term we use the induction hypothesis. For each fixed $i \in \{1, \ldots, n\}$ and $y_i \in \mathcal{X}$ we interpret $q^n(\cdot \mid y_i)$ as a measure on $\mathcal{X}_{i^c}$, for which

$$\begin{aligned}
\beta(q^n(\cdot \mid y_i)) &= \min_{j \neq i} \min_{x \in \mathcal{X}_{i^c}} \frac{q^n(x \mid y_i)}{q^n(z \in \mathcal{X}_{i^c} : z_{j^c} = x_{j^c} \mid y_i)} \\
&= \min_{j \neq i} \min_{x \in \in \mathcal{X}_{i^c}} \frac{q^n(x, y_i)}{q^n(z \in \mathcal{X}^n : z_{(ij)^c} = x_{j^c}, z_i = y_i)} \\
&\geq \min_{j=1,\ldots,n} \min_{z \in \mathcal{X}^n} \frac{q^n(z_j, z_{j^c})}{q^n(z_{j^c})} = \beta(q^n)
\end{aligned}$$

and (4.2) hold with the same constant C. To rewrite (4.3) let us denote by $z \in \mathcal{X}_{i^c}$ a generic vector. A short calculation shows that the conditional probability of $p^n(\cdot \mid y_i)$ with respect to the projection $pr_{j^c} : \mathcal{X}_{i^c} \to \mathcal{X}_{(ij)^c}$ for some $j \neq i$ is given by $p^n(\cdot \mid z_{j^c}, y_i)$, and the same holds for $q^n(\cdot \mid y_i)$. Thus we obtain

$$\begin{aligned}
&\int H(p^n(\cdot \mid y_i) \mid\mid q^n(\cdot \mid y_i)) dp^n(y) \\
&\leq \frac{C}{\beta} \sum_{j \neq i} \iint H(p^n(\cdot \mid z_{j^c}, y_i) \mid\mid q^n(\cdot \mid z_{j^c}, y_i)) dp^n(z \mid y_i) dp^n(y) \\
&= \frac{C}{\beta} \sum_{j \neq i} \int H(p^n(\cdot \mid y_{j^c}) \mid\mid q^n(\cdot \mid y_{j^c})) dp^n(y).
\end{aligned}$$

Summation over i gives

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{p^n}\, H(p^n(\,\cdot\mid y_i)\mid\mid q^n(\,\cdot\mid y_i)) \le \frac{C(n-1)}{\beta n}\sum_{i=1}^{n}\mathbb{E}_{p^n}\, H(p^n(\,\cdot\mid y_{i^c})\mid\mid q^n(\,\cdot\mid y_{i^c})),$$

which combined with the first term yields the claim.

(ii): (4.4) is a simple rewriting of (4.3), noting that as a consequence of the disintegration theorem (or in this case Bayes' theorem) we have

$$\frac{dp^n(\,\cdot\mid y_{i^c})}{dq^n(\,\cdot\mid y_{i^c})}(y_i) = \frac{f(y_{i^c}, y_i)}{\int f(y_{i^c}, x_i)dq^n(x_i\mid y_{i^c})} \quad \text{and} \quad \frac{dp^n_{i^c}}{dq^n_{i^c}}(x_{i^c}) = \int f(x_{i^c}, x_i)dq^n(x_i\mid x_{i^c}).$$

(iii): See [20, Theorem 2]. □

As mentioned, in [20, Theorem 1] it is stated that using the quantity

$$\beta \coloneqq \inf_{i=1,\ldots,n}\ \inf_{x\in\mathcal{X}^n : q^n(x)>0} q^n(x_i\mid x_{i^c})$$

one can deduce $q^n(z : z_i = x_i) \ge \beta$ for all x_i such that the left-hand side is nonzero. This is possible only if q^n has full support. A counterexample is given by the push-forward of a random permutation under the map $\sigma \mapsto (\sigma_1, \ldots, \sigma_n)$, which satisfies $\beta = 1$.

Another possibility is to modify the quantity as

$$\widetilde{\beta}(q^n) \coloneqq \inf_{i=1,\ldots,n}\ \inf_{x\in\mathcal{X}^n : q^n(x)>0} q^n(z : z_i = x_i),$$

but this definition does not behave well under conditional probabilities. It is not true that in general that for a fixed $y_i \in \mathcal{X}$ we have $\widetilde{\beta}(q^n(\,\cdot\mid y_i)) \ge \widetilde{\beta}(q^n)$, which can be seen in examples.

References

[1] Radosław Adamczak, *Moment inequalities for U-statistics*, Ann. Probab. **34** (2006), no. 6, 2288–2314. MR-2294982

[2] Radosław Adamczak and Paweł Wolff, *Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order*, Probab. Theory Related Fields **162** (2015), no. 3–4, 531–586. MR-3383337

[3] Shigeki Aida and Daniel W. Stroock, *Moment estimates derived from Poincaré and logarithmic Sobolev inequalities*, Math. Res. Lett. **1** (1994), no. 1, 75–86. MR-1258492

[4] Luigi Ambrosio and Nicola Gigli, *A user's guide to optimal transport*, Modelling and optimisation of flows on networks, Lecture Notes in Math., vol. 2062, Springer, Heidelberg, 2013, pp. 1–155. MR-3050280

[5] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. MR-2401600

[6] Sergey G. Bobkov, *The growth of L^p-norms in presence of logarithmic Sobolev inequalities*, Vestnik Syktyvkar Univ. **11** (2010), no. 2, 92–111.

[7] Sergey G. Bobkov, Friedrich Götze, and Holger Sambale, *Higher order concentration of measure*, Commun. Contemp. Math. **21** (2019), no. 3. MR-3947067

[8] Sergey G. Bobkov and Prasad Tetali, *Modified logarithmic Sobolev inequalities in discrete settings*, J. Theoret. Probab. **19** (2006), no. 2, 289–336. MR-2283379

[9] Guy Bresler and Dheeraj M. Nagaraj, *Stein's method for stationary distributions of Markov chains and application to Ising models*, arXiv preprint (2017). arXiv:1712.05743

[10] Sourav Chatterjee and Partha S. Dey, *Applications of Stein's method for concentration inequalities*, Ann. Probab. **38** (2010), no. 6, 2443–2485. MR-2683635

[11] Constantinos Daskalakis, Nishanth Dikkala, and Gautam Kamath, *Concentration of multilinear functions of the Ising model with applications to network data*, arXiv preprint (2017). arXiv:1710.04170

[12] Claude Dellacherie and Paul-André Meyer, *Probabilities and potential*, North-Holland Mathematics Studies, vol. 29, North-Holland Publishing Co., Amsterdam-New York; North-Holland Publishing Co., Amsterdam-New York, 1978. MR-0521810

[13] Amir Dembo and Ofer Zeitouni, *Large deviations techniques and applications*, Stochastic Modelling and Applied Probability, vol. 38, Springer-Verlag, Berlin, 2010, Corrected reprint of the second (1998) edition. MR-2571413

[14] Persi Diaconis and Laurent Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab. **6** (1996), no. 3, 695–750. MR-1410112

[15] Reza Gheissari, Eyal Lubetzky, and Yuval Peres, *Concentration inequalities for polynomials of contracting Ising models*, Electron. Commun. Probab. **23** (2018), Paper No. 76, 12. MR-3873783

[16] Leonard Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), no. 4, 1061–1083. MR-0420249

[17] Christof Külske, *Concentration inequalities for functions of Gibbs fields with application to diffraction and random Gibbs measures*, Comm. Math. Phys. **239** (2003), no. 1–2, 29–51. MR-1997114

[18] Rafał Latała, *Estimates of moments and tails of Gaussian chaoses*, Ann. Probab. **34** (2006), no. 6, 2315–2331. MR-2294983

[19] Tzong-Yow Lee and Horng-Tzer Yau, *Logarithmic Sobolev inequality for some models of random walks*, Ann. Probab. **26** (1998), no. 4, 1855–1873. MR-1675008

[20] Katalin Marton, *Logarithmic Sobolev inequalities in discrete product spaces: a proof by a transportation cost distance*, arXiv preprint (2015). arXiv:1507.02803

[21] Colin McDiarmid, *On the method of bounded differences*, Surveys in combinatorics, 1989 (Norwich, 1989), London Math. Soc. Lecture Note Ser., vol. 141, Cambridge Univ. Press, Cambridge, 1989, pp. 148–188. MR-1036755

[22] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz, *Noise stability of functions with low influences: invariance and optimality*, Ann. of Math. (2) **171** (2010), no. 1, 295–341. MR-2630040

[23] Holger Sambale, *Second order concentration for functions of independent random variables*, Ph.D. thesis, Bielefeld University, 2016.

[24] Holger Sambale and Arthur Sinulis, *Logarithmic Sobolev inequalities for finite spin systems and applications*, arXiv preprint (2018). arXiv:1807.07765

[25] Daniel W. Stroock and Bogusław Zegarliński, *The equivalence of the logarithmic Sobolev inequality and the Dobrushin–Shlosman mixing condition*, Comm. Math. Phys. **144** (1992), no. 2, 303–323. MR-1152374

[26] Daniel W. Stroock and Bogusław Zegarliński, *The logarithmic Sobolev inequality for discrete spin systems on a lattice*, Comm. Math. Phys. **149** (1992), no. 1, 175–193. MR-1182416

[27] Cédric Villani, *Optimal transport, old and new*, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer-Verlag, Berlin, 2009. MR-2459454

[28] Bogusław Zegarliński, *Dobrushin uniqueness theorem and logarithmic Sobolev inequalities*, J. Funct. Anal. **105** (1992), no. 1, 77–111. MR-1156671

Acknowledgments. We are grateful to two anonymous referees for their valuable remarks, which helped improving the presentation of this paper.