A note on concentration for polynomials in the Ising model[*]

Radosław Adamczak[†‡]   Michał Kotowski[‡]   Bartłomiej Polaczyk[‡]
Michał Strzelecki[‡]

Abstract

We present precise multilevel exponential concentration inequalities for polynomials in Ising models satisfying the Dobrushin condition. The estimates have the same form as two-sided tail estimates for polynomials in Gaussian variables due to Latała. In particular, for quadratic forms we obtain a Hanson–Wright type inequality.

We also prove concentration results for convex functions and estimates for nonnegative definite quadratic forms, analogous as for quadratic forms in i.i.d. Rademacher variables, for more general random vectors satisfying the approximate tensorization property for entropy.

1 Introduction

Since its introduction in [26], the Ising model has been a source of numerous mathematical questions. In addition to its physical importance it is appealing to mathematicians, providing an easy to formulate, yet challenging model of dependent random variables and serving as testing ground for many probabilistic ideas. Recently in the context of finite graphs, the Ising model attracted also attention of statisticians and theoretical computer scientists interested, e.g., in estimating the parameters of the model, learning the underlying graph structure or testing some properties of the model in a computationally efficient way (see, e.g., [15, 37, 38, 16]). In particular in the last two decades several authors studied the Ising model from the point of view of concentration

---

[†]Institute of Mathematics of the Polish Academy of Sciences
[‡]Institute of Mathematics, University of Warsaw, Poland. E-mail: R.Adamczak@mimuw.edu.pl,M.Kotowski@mimuw.edu.pl,B.Polaczyk@mimuw.edu.pl,M.Strzelecki@mimuw.edu.pl

of measure phenomena, see, e.g., [31, 12, 30, 13]. While most effort has been devoted to concentration inequalities for functions satisfying appropriate Lipschitz type conditions, recently several papers appeared related to variance bounds or stronger, exponential concentration inequalities for polynomials, to mention the work by Daskalakis, Dikkala, Kamath [15, 14], Gheissari, Lubetzky and Peres [19], Götze, Sambale and Sinulis [21]. Motivation for these developments ranged from statistical and algorithmic (efficient discrimination between samples drawn from an Ising model and i.i.d. samples) to purely probabilistic ones (searching for counterparts of inequalities known in the i.i.d. case).

In this note we complement the results proved in the aforementioned papers by obtaining exponential inequalities for polynomials of the same form as in bounds for polynomials in independent Gaussian (or more generally subgaussian) random variables, which were introduced originally by Latała [28] and subsequently studied, e.g., by Adamczak and Wolff [5]. Such inequalities are expressed in terms of appropriate injective tensor product norms of averaged derivatives of the polynomials in question and in the Gaussian case are known to be optimal up to constants depending only on the degree of the polynomial. Optimality is understood here in a strong sense – the inequalities can be up to constants reversed. Moreover, they are known to imply other, more classical inequalities for multilinear forms, such as Bonami–Nelson inequalities [10, 34]. In particular, for polynomials of degree d they provide multilevel type concentration, of the form $\exp(-ct^2)$ for small values of t up to $\exp(-c't^{2/d})$ for larger values (as opposed to the inequalities from the aforementioned results for the Ising model, which do not yield precise multilevel concentration but rather give weaker bounds of the form $\exp(-c''t^{2/d})$ for all t). As a consequence our estimates imply the previous ones and provide a more accurate description of the tail behavior.

Our approach is similar to the one by Götze, Sambale and Sinulis in that it builds on general Aida–Stroock type moment estimates those Authors obtained for the Ising model in [21], and uses them as a tool in an inductive argument. However, the details are different, while in [21] one works with moments of Euclidean norms of discrete iterated derivatives of multilinear forms, we adapt an argument from [5], linearizing the Euclidean norms with an auxiliary Gaussian sequence, which allows us to treat general functions (seen by the Fourier–Walsh theory as tetrahedral polynomials) and pass from discrete gradients to classical derivatives.

The argument we present may be seen as a method of reduction of concentration properties for polynomials from the Ising model to the i.i.d. Gaussian case. Since the random variables considered in the Ising model take only values ± 1, one could expect a similar reduction to polynomials in i.i.d. Rademacher variables. We are able to obtain such estimates for positive definite quadratic forms, by passing through concentration properties for convex functions which are of independent interest.

The organization of the article is as follows. First, in Section 2 we present our main result (Theorem 2.2) and discuss its relation to known inequalities for the Ising model as well as to the estimates for the i.i.d. case. Next, in Section 3 we discuss the approximate tensorization of entropy (as studied recently by Marton [32] and Caputo, Menz, Tetali [11]) and Aida–Stroock type moment estimates obtained by Götze, Sambale and Sinulis [21]. Using these tools, in Section 4 we present the proof of the main result. The final Section 5 presents estimates for convex functions and Rademacher-type inequalities for quadratic forms.

2 Gaussian type inequality for polynomials

2.1 Basic definitions and notation

Let us begin by introducing the general form of the Ising model on a finite set.

Definition 2.1 (Ising model). *Let n be a positive integer and let μ be the measure on $\{-1, 1\}^n$, having density with respect to the uniform distribution of the form*

$$\mu(\sigma) = \frac{1}{Z} \exp\left(\frac{1}{2}\sum_{i,j=1}^{n} J_{ij}\sigma_i\sigma_j - \sum_{i=1}^{n} h_i\sigma_i\right), \tag{2.1}$$

for any $\sigma \in \{-1, 1\}^n$, where $J = (J_{ij})_{i,j\leq n}$ is a symmetric matrix with vanishing diagonal, $h = (h_i)_{i\leq n} \in \mathbb{R}^n$ and Z is a normalizing constant.

In physical terms the coupling matrix J corresponds to interactions between particles and the vector h describes an external field. The order of magnitude of the constants J_{ij} reflects the temperature (the higher the temperature the smaller the coefficients, which corresponds to weaker interactions), however, as our results will be expressed solely in terms of the coefficients J_{ij} and h_i, we will not incorporate the temperature into the notation.

To obtain concentration inequalities, one needs some control over the coupling constants and the external field, which will allow for sufficient proximity to the i.i.d. case. The conditions we will impose on the model are classical and in the context of concentration of measure appeared already in [32, 19, 21].

Main assumptions. We will assume that

$$|h|_\infty := \max_{i\leq n} |h_i| \leq \alpha \tag{2.2}$$

and the coupling constants satisfy Dobrushin's condition:

$$\max_{i\leq n} \sum_{j=1}^{n} |J_{ij}| \leq 1 - \rho \tag{2.3}$$

for some $\rho > 0$.

In order to formulate concentration of measure estimates for polynomials of the Ising model, which correspond to inequalities obtained by Latała for polynomials in independent Gaussian random variables, we will need to introduce a family of injective tensor product norms on d-index matrices (d-tensors).

To provide transparent notation for multi-indices we will use the following convention. For a positive integer n we will denote $[n] = \{1, \ldots, n\}$. The cardinality of a set I will be denoted by $|I|$. For $\mathbf{i} = (i_1, \ldots, i_d) \in [n]^d$ and $I \subseteq [d]$ we write $\mathbf{i}_I = (i_k)_{k\in I}$. We will also denote $|\mathbf{i}| = \max_{j\leq d} i_j$ and $|\mathbf{i}_I| = \max_{j\in I} i_j$. We will often deal with homogeneous polynomials, defined in terms of multi-indexed matrices (tensors). We will say that a d-indexed matrix $A = (a_\mathbf{i})_{\mathbf{i}\in[n]^d}$ is symmetric if for every permutation σ of the set $[d]$ and every $\mathbf{i} = (i_1, \ldots, i_d) \in [n]^d$, we have $a_\mathbf{i} = a_{i_{\sigma(1)}, \ldots, i_{\sigma(d)}}$. When d is fixed, we will write simply $A = (a_\mathbf{i})_{|\mathbf{i}|\leq n}$. We will say that a d-indexed matrix $A = (a_\mathbf{i})_{|\mathbf{i}|\leq n}$ has vanishing generalized diagonals if $a_\mathbf{i} = 0$ for all $\mathbf{i} = (i_1, \ldots, i_d)$ such that there exist $k \neq l$ with $i_k = i_l$.

Let now P_d be the set of partitions of $[d]$ into nonempty, pairwise disjoint sets. For a partition $\mathcal{I} = \{I_1, \ldots, I_k\} \in P_d$ and a d-indexed matrix $A = (a_\mathbf{i})_{\mathbf{i}\in[n]^d}$, define

$$\|A\|_\mathcal{I} = \sup\left\{\sum_{\mathbf{i}\in[n]^d} a_\mathbf{i} \prod_{l=1}^{k} x_{\mathbf{i}_{I_l}}^{(l)} : \|(x_{\mathbf{i}_{I_l}}^{(l)})\|_2 \leq 1, 1 \leq l \leq k\right\}, \tag{2.4}$$

where $\|(x_{\mathbf{i}_{I_l}})\|_2 = \sqrt{\sum_{|\mathbf{i}_{I_l}| \le n} x_{\mathbf{i}_{I_l}}^2}$. Thus, e.g.,

$$\|(a_{ij})_{i,j \le n}\|_{\{1,2\}} = \sup \Big\{ \sum_{i,j \le n} a_{ij} x_{ij} \colon \sum_{i,j \le n} x_{ij}^2 \le 1 \Big\} = \sqrt{\sum_{i,j \le n} a_{ij}^2} = \|(a_{ij})_{i,j \le n}\|_{HS},$$

$$\|(a_{ij})_{i,j \le n}\|_{\{1\}\{2\}} = \sup \Big\{ \sum_{i,j \le n} a_{ij} x_i y_j \colon \sum_{i \le n} x_i^2 \le 1, \sum_{j \le n} y_j^2 \le 1 \Big\} = \|(a_{ij})_{i,j \le n}\|_{\ell_2^n \to \ell_2^n},$$

$$\|(a_{ijk})_{i,j,k \le n}\|_{\{1,2\}\{3\}} = \sup \Big\{ \sum_{i,j,k \le n} a_{ijk} x_{ij} y_k \colon \sum_{i,j \le n} x_{ij}^2 \le 1, \sum_{k \le n} y_k^2 \le 1 \Big\}.$$

Note that for simplicity in the notation we skip the outer brackets and commas in the subscript and write, e.g., $\| \cdot \|_{\{1\}\{2\}}$ instead of $\| \cdot \|_{\{\{1\},\{2\}\}}$.

In particular for $d = 2$, $\| \cdot \|_{\{1,2\}}$ and $\| \cdot \|_{\{1\}\{2\}}$ coincide with the Hilbert–Schmidt and operator norm of a matrix respectively. Note that for every d and $\mathcal{I} \in P_d$ we have $\|A\|_{\mathcal{I}} \le \|A\|_{\{[d]\}} = \sqrt{\sum_{|\mathbf{i}| \le n} a_{\mathbf{i}}^2}$. The norm $\|A\|_{\{[d]\}}$ can be considered a counterpart of the Hilbert–Schmidt norm for higher order tensors.

We will use the standard notation $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$ for a random variable X. Sometimes, when dealing with independent random variables X, Y, we will use the notation $\mathbb{E}_X$ for the expectation with respect to the variable X (i.e., the conditional expectation given Y).

In what follows, we will write, e.g., c_a, C_a or $c_a(b)$ to denote constants depending only on the parameters a or a, b respectively. The values of such constants may change between occurrences.

By the Fourier–Walsh expansion (see, e.g., [35]), every function $f \colon \{-1, 1\}^n \to \mathbb{R}$ can be written in a unique way as a tetrahedral polynomial, i.e., a polynomial which is affine with respect to every variable (in particular the degree of the polynomial is at most n). Therefore in what follows we will restrict our attention to this representation. In particular, when we speak about gradients ∇f or higher order derivatives $\nabla^k f$, we always think of the usual derivatives of the polynomial function on $\mathbb{R}^n$ given by the tetrahedral representation of f.

2.2 Main result

The main result of this section is

Theorem 2.2. *Let μ be defined by* (2.1) *and assume that the inequality* (2.2) *and the Dobrushin condition* (2.3) *hold. Let X be a random vector distributed according to μ. Then for $d \ge 1$ there exist constants $c = c(d, \alpha, \rho)$ such that for any tetrahedral polynomial $f \colon \{-1, 1\}^n \to \mathbb{R}$ of degree d, and any $t > 0$,*

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \ge t) \le 2 \exp \Big(-c \min_{1 \le k \le d} \min_{\mathcal{I} \in P_k} \Big(\frac{t}{\|\mathbb{E}\nabla^k f(X)\|_{\mathcal{I}}} \Big)^{2/|\mathcal{I}|} \Big). \qquad (2.5)$$

Remark 2.3. *If $h = 0$ and $f(x) = \sum_{|\mathbf{i}| \le n} a_{\mathbf{i}} \prod_{l=1}^{d} x_{i_l}$ for some symmetric d-indexed matrix with vanishing generalized diagonals, then for any $\mathcal{I} \in P_k$ we can estimate*

$$\|\mathbb{E}\nabla^k f(X)\|_{\mathcal{I}}^2 \le \|\mathbb{E}\nabla^k f(X)\|_{\{[k]\}}^2 = \Big(\frac{d!}{(d-k)!} \Big)^2 \sum_{i_1,\ldots,i_k=1}^{n} \Big(\sum_{i_{k+1},\ldots,i_d=1}^{n} a_{\mathbf{i}} \mathbb{E} \prod_{l=k+1}^{d} X_{i_l} \Big)^2$$

$$\le \Big(\frac{d!}{(d-k)!} \Big)^2 \sum_{i_1,\ldots,i_k=1}^{n} \mathbb{E} \Big(\sum_{i_{k+1},\ldots,i_d=1}^{n} a_{\mathbf{i}} \prod_{l=k+1}^{d} X_{i_l} \Big)^2.$$

By (the proof of) Lemma 3.1. in [19], $\mathbb{E}(\sum_{i_{k+1},\ldots,i_d=1}^n a_{\mathbf{i}} \prod_{l=k+1}^d X_i)^2 \leq C_d(\rho) \max_{\mathbf{i}} |a_{\mathbf{i}}|^2 n^{d-k}$. As a consequence, $\|\mathbb{E}\nabla^k f(X)\|_{\mathcal{I}} \leq \widetilde{C}_d(\rho) n^{d/2} \max_{\mathbf{i}} |a_{\mathbf{i}}|$ and Theorem 2.2 (after adjustment of constants) implies

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq t) \leq 2\exp\left(-\widetilde{c}\,\frac{t^{2/d}}{n \max_{\mathbf{i}} |a_{\mathbf{i}}|^{2/d}}\right) \qquad (2.6)$$

for some new constant $\widetilde{c} = \widetilde{c}_d(\rho)$. This inequality was proved by Götze, Sambale, Sinulis in [21] and earlier, up to some additional logarithmic in n factors in the exponent, by Gheissari, Lubetzky and Peres in [19].

Remark 2.4. An inequality analogous to (2.5) for polynomials in Gaussian variables has been obtained in [5]. In this case the inequality can be reversed up to numerical constants depending only on d (in front of and inside the exponent). The proof relied on a reduction to the special case of tetrahedral multilinear forms in independent standard Gaussian variables obtained by Latała [28]. The proof of Theorem 2.2 presented below is a simple adaptation of this idea. The inequality (2.5) is also known to hold for polynomials in i.i.d. subgaussian random variables [5].

Example 2.5. In this example and the following ones we will let c denote a constant which may depend on the parameters α and ρ. Its value may change between occurrences.

For $f(x) = \sum_{i,j=1}^n a_{ij} x_i x_j$, where $A = (a_{ij})_{i,j \leq n}$ is a symmetric matrix with zeros on the diagonal, we obtain

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq t) \leq 2\exp\left(-c\min\left(\frac{t^2}{\|A\|_{HS}^2 + \sum_{i=1}^n(\sum_{j=1}^n a_{ij}\mathbb{E}X_j)^2}, \frac{t}{\|A\|_{\ell_2^n \to \ell_2^n}}\right)\right).$$

Note that if $h = 0$ then $\mathbb{E}X_i = 0$ (as the distribution of X is symmetric) and the right hand side simplifies to

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq t) \leq 2\exp\left(-c\min\left(\frac{t^2}{\|A\|_{HS}^2}, \frac{t}{\|A\|_{\ell_2^n \to \ell_2^n}}\right)\right), \qquad (2.7)$$

which gives a counterpart of the Hanson–Wright inequality known for quadratic forms in independent sub-Gaussian random variables [24], which turned out to be useful, e.g., in random-matrix theory and statistics (see, e.g., [43]). A version of this inequality for strongly mixing Ising models on a lattice was proved by Marton [31].

If one further estimates the operator norm by the Hilbert–Schmidt norm, one obtains a weaker tail bound of the form

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq t) \leq 2\exp\left(-c\frac{t}{\|A\|_{HS}}\right). \qquad (2.8)$$

For quadratic form in independent Rademacher variables such an inequality (together with counterparts for higher order forms) was for the first time established by Bonami [10], Beckner [7] and Gross [23] in the context of hypercontractivity of semigroups. A counterpart of (2.8) for quadratic forms of the Ising model has been recently obtained in [21].

Example 2.6. Consider now $f(x) = \sum_{1 \leq i,j,k \leq n} a_{ijk} x_i x_j x_k$, where $A = (a_{ijk})_{i,j,k \leq n}$ is symmetric with vanishing generalized diagonals. Assume also that $h = 0$.

In this case Theorem 2.2 gives

$$\mathbb{P}(|f(X)| \geq t) \qquad (2.9)$$

$$\leq 2\exp\left(-c\min\left(\frac{t^2}{\|A\|_{\{1,2,3\}}^2 + \sum_i(\sum_{jk} a_{ijk}\mathbb{E}X_j X_k)^2}, \frac{t}{\|A\|_{\{1,2\}\{3\}}}, \frac{t^{2/3}}{\|A\|_{\{1\}\{2\}\{3\}}^{2/3}}\right)\right),$$

where we again used the equality $\mathbb{E}X_i = 0$.

One can wonder if it is possible to obtain estimates just in terms of the norm $\|A\|_{\{1,2,3\}}$, e.g., of the form $\mathbb{P}(|f(X)| \geq t) \leq 2\exp(-ct^{2/3}/\|A\|_{\{1,2,3\}}^{2/3})$ as in the independent case or in the case of quadratic forms discussed above. Clearly, this is true if one can estimate the quantity $\sum_{i,j\leq n}(\mathbb{E}X_i X_j)^2$ by a constant independent of n. However it turns out that if one assumes only the Dobrushin condition (2.3), then the coefficient $\sum_i(\sum_{jk}a_{ijk}\mathbb{E}X_i X_j)^2$ in general cannot be discarded. To see this consider the Ising model on the one-dimensional interval, e.g., with $J_{i,i+1} = J_{i+1,i} = 1/3$ for $i = 1,\dots,n-1$ and $J_{ij} = 0$ otherwise. In this case (2.3) is clearly satisfied with $\rho = 1/3$ and the Hamiltonian is of the form $-\frac{1}{3}\sum_{i=1}^{n-1}\sigma_i\sigma_{i+1}$. Since under the uniform measure on the discrete cube, σ_1 and the products $\sigma_i\sigma_{i+1}$, $i = 1,\dots,n-1$ are independent Rademacher variables, one can see that under the measure μ given by (2.1), the products $\sigma_i\sigma_{i+1}$, are i.i.d. random variables with distribution

$$\mu(\sigma_i\sigma_{i+1} = 1) = 1 - \mu(\sigma_i\sigma_{i+1} = -1) = \frac{1}{1+e^{-2/3}}.$$

In particular $\mathbb{E}_\mu \sigma_i\sigma_{i+1} = a > 0$ is independent of n. Consider now a symmetric 3-indexed matrix $A = (a_{ijk})_{i,j,k\leq n}$ with vanishing generalized diagonals and coefficients a_{ijk} defined for $i < j < k$ with the formula

$$a_{ijk} = \mathbb{1}_{\{i<j=k-1\}}.$$

One can easily see that $\|A\|_{\{1,2,3\}}$ is of order n (as A has $O(n^2)$ nonzero coefficients). However, if X is distributed according to μ, then for $f(X) = \sum_{i,j,k\leq n}a_{ijk}X_i X_j X_k$, $\sqrt{\text{Var }f(X)} = \|f(X)\|_2$ is of order $n^{3/2}$ (as can be checked by using that $\mathbb{E}X_i X_j X_k = 0$, expanding the product $\mathbb{E}f(X)^2$ and performing some elementary combinatorics). This shows that an estimate of the form $\mathbb{P}(|f(X)| \geq t) \leq 2\exp(-c(t/\|A\|_{\{1,2,3\}})^\kappa)$ cannot hold with c, κ independent of n. Of course, by (2.6) we do have the inequality $\mathbb{P}(|f(X)| \geq t) \leq 2\exp(-ct^{2/3}/n)$. However, (2.9) leads to an improvement of this inequality. As one can easily check $\|A\|_{\{1,2\}\{3\}}$ is of the same order as $\|A\|_{\{1,2,3\}}$, i.e., of order n, $\|A\|_{\{1\}\{2\}\{3\}}$ is of order $\sqrt{n}$ and $\sum_i(\sum_{jk}a_{ijk}\mathbb{E}X_j X_k)^2$ is of order n^3. Together with some elementary calculations, this gives

$$\mathbb{P}(|f(X)| \geq t) \leq 2\exp\left(-c\min\left(\frac{t^2}{n^3}, \frac{t^{2/3}}{n^{1/3}}\right)\right).$$

Remark 2.7. In [21], Götze, Sambale and Sinulis propose an appropriate re-centering of a d-linear form $f(x) = \sum_{|\mathbf{i}|\leq n}a_{\mathbf{i}}\prod_{k=1}^d x_{i_k}$ by a polynomial of lower degree, resulting in a (nonhomogeneous) polynomial $f_{A,d}$ such that $\mathbb{E}\nabla^i f(X) = 0$ for all $i = 1,\dots,d-1$. They prove that then

$$\mathbb{P}(|f_{A,d}(X) - \mathbb{E}f_{A,d}(X)| \geq t) \leq 2\exp\left(-c\frac{t^{2/d}}{\|A\|_{\{1,\dots,d\}}^{2/d}}\right)$$

with some $c = c(\alpha, \rho, d)$. Since Theorem 2.2 applies to general (not necessarily homogeneous) polynomials, it yields a refinement of the above inequality, of the form

$$\mathbb{P}(|f_{A,d}(X) - \mathbb{E}f_{A,d}(X)| \geq t) \leq 2\exp\left(-c\min_{\mathcal{I}\in P_d}\left(\frac{t}{\|A\|_{\mathcal{I}}}\right)^{2/|\mathcal{I}|}\right).$$

Example 2.6 shows that in general one cannot eliminate passing to $f_{A,d}$, i.e., the above inequality may not hold for the original polynomial f.

3 Approximate tensorization of entropy and moment estimates

In this section we will present basic tools (coming mostly from the work by Marton [32] and Götze, Sambale, Sinulis [21]) which we will need for the proof of Theorem 2.2 and also in Section 5 to obtain concentration for convex functions and Rademacher-type bounds for quadratic forms.

Let $\mathcal{X} = \prod_{i=1}^{n} \mathcal{X}_i$, where $\mathcal{X}_i$ are Polish spaces with their Borel σ-fields and let μ be a probability distribution on $\mathcal{X}$.

For each $I \subseteq [n]$, and $x = (x_1, \ldots, x_n) \in \mathcal{X}$ denote $x_I = (x_j)_{j \in I}$, $\bar{x}_I = x_{I^c} = (x_j)_{j \notin I} \in \prod_{j \notin I} \mathcal{X}_i$. Let μ_I be the marginal of μ corresponding to the coordinates indexed by I and $\mu_I(\cdot | \bar{x}_I)$ denote the regular conditional distribution of x_I given $\bar{x}_I$ on the probability space $(\mathcal{X}, \mu)$. Thus for any Borel set $A \subset \prod_{i \in I} \mathcal{X}_i$, we have

$$\mu\left(A \times \prod_{j \notin I} \mathcal{X}_j\right) = \int_{\prod_{j \in I^c} \mathcal{X}_j} \mu_I(A | \bar{x}_I) \mu_{I^c}(d\bar{x}_I) = \int_{\mathcal{X}} \mu_I(A | \bar{x}_I) \mu(dx).$$

If $I = \{i\}$ we will write, e.g., $\bar{x}_i, \mu_i$ instead of $\bar{x}_{\{i\}}, \mu_{\{i\}}$.

Recall that for a probability measure μ and a nonnegative function f, the entropy of f relative to μ is defined as $\operatorname{Ent}_\mu(f) = \mathbb{E}_\mu f \log f - \mathbb{E}_\mu f \log \mathbb{E}_\mu f$ whenever $\mathbb{E}_\mu f \log f < \infty$ and $\operatorname{Ent}_\mu(f) = \infty$ otherwise.

The following definition will play a crucial part in what follows.

Definition 3.1 (Approximate tensorization of entropy). *We will say that μ has the approximate tensorization property with constant C (abbrev. $AT(C)$) if for every function $f \colon \mathcal{X} \to [0, \infty)$,*

$$\operatorname{Ent}_\mu(f) \leq C \mathbb{E}_\mu \sum_{i=1}^{n} \operatorname{Ent}_{\mu_i(\cdot | \bar{x}_i)}(f)$$

It is well known that product measures satisfy $AT(1)$, see, e.g., [29, Proposition 5.6]. Recently Marton [32] (see also [11, 21]) proved the following sufficient condition for tensorization of entropy in discrete product spaces.

Theorem 3.2 (Marton, Götze-Sambale-Sinulis). *Let μ be a measure with full support on $\mathcal{X}$. Set*

$$\beta = \min_{i \leq n} \min_{x \in \mathcal{X}} \mu_i(\{x_i\} | \bar{x}_i) \tag{3.1}$$

Let also $A = (a_{ij})_{i,j \leq n}$ satisfy $a_{ii} = 0$ for all i and for $i \neq j$,

$$\|\mu_i(\cdot | \bar{x}_i) - \mu_i(\cdot | \bar{y}_i)\|_{TV} \leq a_{ij}$$

for every $x, y \in \mathcal{X}$ which differ only at the j-th coordinate. Assume moreover that $\|A\|_{\ell_2^n \to \ell_2^n} < 1$. Then the measure μ has the approximate tensorization property with constant $C = 2 \frac{1}{\beta(1 - \|A\|_{\ell_2^n \to \ell_2^n})^2}$.

In particular in [21] the Authors verify that under our main assumptions the approximate tensorization property is satisfied by the Ising model.

Corollary 3.3 (Götze-Sambale-Sinulis). *If a measure μ on $\{-1, 1\}^n$ is defined by (2.1) then for any distinct $i, j \in [n]$ and any $x, y \in \mathcal{X}^n$ differing only at the j-th coordinate*

$$\sup_{\substack{x, y \in \mathcal{X}^n \\ \bar{x}_j = \bar{y}_j}} \|\mu_i(\cdot | \bar{x}_i) - \mu_i(\cdot | \bar{y}_i)\|_{TV} \leq |J_{ij}|$$

In particular if the condition (2.3) is satisfied, then

$$\|A\|_{\ell_2^n \to \ell_2^n} \leq \|J\|_{\ell_1^n \to \ell_1^n} \leq 1 - \rho.$$

If additionally the vector h satisfies (2.2), then the coefficient β defined in (3.1), satisfies

$$\frac{1}{C_{\alpha,\rho}} \leq \beta \leq C_{\alpha,\rho}$$

where $C_{\alpha,\rho}$ depends only on α and ρ. As a consequence the measure μ satisfies $AT(C)$ with C depending only on ρ and α.

We will also need the definition of the discrete gradient on $\mathcal{X}$, induced by the measure μ. To this end we will slightly abuse the notation and write $(\bar{x}_i, y_i)$ for the sequence z such that $z_i = y_i$ and $\bar{z}_i = \bar{x}_i$.

Following [21] let us introduce

Definition 3.4. *For a measurable function $f \colon \mathcal{X} \to \mathbb{R}$ and $i \in [n]$, define*

$$\mathfrak{d}_i f(x) = \left(\frac{1}{2} \int_{\mathcal{X}_i} (f(x) - f(\bar{x}_i, y))^2 \mu_i(dy|\bar{x}_i) \right)^{1/2}$$

$$= \left(\frac{1}{2} \mathbb{E}\left(\left(f(X) - f(X_1, \ldots, X_{i-1}, \widetilde{X}_i, X_{i+1}, \ldots, X_n) \right)^2 \Big| X = x \right) \right)^{1/2},$$

where $X = (X_1, \ldots, X_n)$ is distributed according to μ and the conditional distribution of $\widetilde{X}_i$ given $X = x$ equals to $\mu_i(\cdot|\bar{x}_i)$.

We will denote $\mathfrak{d}f(x) = (\mathfrak{d}_i f(x))_{i \in [n]}$ and regard this vector as an element of $\mathbb{R}^n$ endowed with the standard Euclidean norm $|\cdot|$.

Definition 3.5. *We will say that μ satisfies the logarithmic Sobolev inequality with constant C (abbrev. $LSI(C)$) if for every $f \colon \mathcal{X} \to \mathbb{R}$,*

$$\mathrm{Ent}_\mu(f^2) \leq 2C\mathbb{E}|\mathfrak{d}f|^2. \tag{3.2}$$

Remark 3.6. The notion of logarithmic Sobolev inequality introduced above can be interpreted as the usual logarithmic Sobolev inequality equivalent to hypercontractivity of the related Glauber dynamics/Gibbs sampler (see, e.g., [11, 32, 21]), however, we will not use this interpretation in the sequel.

Using the approximate tensorization property together with a log-Sobolev inequality for two point distributions [17] and a Herbst-type argument of Aida and Stroock [6] (see also [9, 5, 2]), Götze, Sambale and Sinulis [21] proved

Theorem 3.7. *Let μ be a measure on $\{-1,1\}^n$, defined by (2.1), and assume that the inequality (2.2) and the Dobrushin condition (2.3) are satisfied. Then there exists a constant $C = C(\alpha, \rho)$ such that μ satisfies the $LSI(C)$. As a consequence, if X is a random vector distributed according to μ, then for any $p \geq 2$,*

$$\|f(X)\|_p^2 \leq \|f(X)\|_2^2 + 2C(p-2)\|\mathfrak{d}f(X)\|_p^2.$$

Combining the above result with the well known fact that the logarithmic Sobolev inequality (3.2) implies the Poincaré inequality with constant C, i.e.,

$$\mathrm{Var}_\mu f \leq C\mathbb{E}|\mathfrak{d}f(X)|^2,$$

we immediately obtain

Corollary 3.8. *Under the assumptions and notation of Theorem 3.7, for every $f \colon \{-1,1\}^n \to \mathbb{R}$, and any $p \geq 2$,*

$$\|f(X) - \mathbb{E}f(X)\|_p \leq \sqrt{2Cp}\|\mathfrak{d}f(X)\|_p.$$

This inequality will be the basis of our inductive argument in the proof of Theorem 2.2.

4 Proof of Theorem 2.2

Proof of Theorem 2.2. The proof will be an adaptation of arguments from [5] (see also [2]) to the discrete case. To carry it out it will be convenient to introduce an inner product $\langle \cdot, \cdot \rangle$ on the space of k-tensors with the formula

$$\langle A, B \rangle = \sum_{|\mathbf{i}| \leq n} a_{\mathbf{i}} b_{\mathbf{i}},$$

where $A = (a_{\mathbf{i}})_{|\mathbf{i}| \leq n}$, $B = (b_{\mathbf{i}})_{|\mathbf{i}| \leq n}$. Let us also recall the notation $x^1 \otimes \cdots \otimes x^k = (x^1_{i_1} x^2_{i_2} \cdots x^k_{i_k})_{i_1,\ldots,i_k \leq n}$ for any vectors $x^j = (x^j_1, \ldots, x^j_n)$, $j = 1, \ldots, k$.

We will first prove by induction on d that for any positive integer d, and any function $f \colon \{-1,1\}^n \to \mathbb{R}$,

$$\|f(X) - \mathbb{E}f(X)\|_p \leq K^d \|\langle \nabla^d f(X), G_1 \otimes \cdots \otimes G_d \rangle\|_p + \sum_{i=1}^{d-1} K^i \|\langle \mathbb{E}_X \nabla^i f(X), G_1 \otimes \cdots \otimes G_i \rangle\|_p,$$

(4.1)

where K is a constant depending only on α, ρ and $G_1, \ldots, G_d$ are i.i.d. standard Gaussian vectors in $\mathbb{R}^n$ independent of X. Here by $\mathbb{E}_X$ we denote expectation with respect to the random vector X and (as explained in Section 2) the derivatives $\nabla^i f$ denote the derivatives of the tetrahedral polynomial coming from the Fourier–Walsh expansion of f. We remark that d does not necessarily coincide with the degree of f.

To prove (4.1) we will proceed by induction. Consider thus $f(x) = A_0 + \sum_{k=1}^{D} \langle A^k, x^{\otimes k} \rangle$, where $D \leq n$, $A_0 \in \mathbb{R}$ and for $k = 1, \ldots, D$, $A^k = (a^k_{\mathbf{i}})_{|\mathbf{i}| \leq n}$ are k-indexed symmetric matrices with zeros on generalized diagonals (note that by using the Fourier-Walsh decomposition one can uniquely represent any function on $\{-1,1\}^n$ in this form).

Let $\widetilde{X}_i$, $i = 1, \ldots, n$ be $\{-1,1\}$-valued random variables (possibly defined on some extension of the original probability space) such that the conditional distribution of $\widetilde{X}_i$ given $X = x$ equals $\mu_i(\cdot|\bar{x}_i)$.

Recall that for $i \leq n$,

$$\mathfrak{d}_i f(x) = \Big(\frac{1}{2}\mathbb{E}\Big(\big(f(X) - f(X_1, \ldots, X_{i-1}, \widetilde{X}_i, X_{i+1}, \ldots, X_n)\big)^2 \Big| X = x\Big)\Big)^{1/2},$$

hence by Corollary 3.8, using the notation $\bar{X}_i = (X_j)_{j \neq i}$ we have for any $p \geq 2$,

$$\|f(X) - \mathbb{E}f(X)\|_p \leq \sqrt{Cp}\Big(\mathbb{E}\Big|\sum_{i=1}^{n} \mathbb{E}\big((f(X) - f(\bar{X}_i, \widetilde{X}_i))^2\big| X\big)\Big|^{p/2}\Big)^{1/p}.$$

Using Jensen's inequality for the conditional expectation, we can further write

$$\|f(X) - \mathbb{E}f(X)\|_p \leq \sqrt{Cp}\Big(\mathbb{E}\Big|\sum_{i=1}^{n} (f(X) - f(\bar{X}_i, \widetilde{X}_i))^2\Big|^{p/2}\Big)^{1/p}.$$

Define now for $k = 1, \ldots, d$ and $i = 1, \ldots, n$ the $(k-1)$-indexed matrices $A^{k,i} = (a^{k,i}_{\mathbf{i}})_{|\mathbf{i}| \leq n} = (a^k_{(i,\mathbf{i})})_{|\mathbf{i}| \leq n}$, where for $\mathbf{i} = (i_1, \ldots, i_{k-1})$ we write $(i, \mathbf{i}) = (i, i_1, \ldots, i_{k-1})$. Using the fact that the generalized diagonals of the matrices A^k vanish together with the symmetry of A^k, we get

$$f(X) - f(\bar{X}_i, \widetilde{X}_i) = (X_i - \widetilde{X}_i)\sum_{k=1}^{D} k \langle A^{k,i}, X^{\otimes(k-1)} \rangle = (X_i - \widetilde{X}_i)\frac{\partial}{\partial x_i} f(X).$$

Since $|X_i - \widetilde{X}_i| \leq 2$, by combining this equality with the previous estimate, we obtain

$$\|f(X) - \mathbb{E}f(X)\|_p \leq 2\sqrt{Cp}(\mathbb{E}|\nabla f(X)|^p)^{1/p}.$$

Using the fact that if g is a standard Gaussian variable, then for $p \geq 1$ we have $\sqrt{p}M^{-1} \leq \|g\|_p \leq M\sqrt{p}$, where M is a universal constant, we can write the above inequality as

$$\|f(X) - \mathbb{E}f(X)\|_p \leq K\left\|\left\langle \nabla f(X), G \right\rangle\right\|_p \qquad (4.2)$$

for a standard n-dimensional Gaussian vector G, independent of $X, \{\widetilde{X}_i\}_{i \leq n}$ and $K = 2\sqrt{C}M$. This establishes (4.1) for $d = 1$.

The induction step follows just by the case $d = 1$ and the triangle inequality in L_p. Indeed, assuming that (4.1) holds for d, by the triangle inequality and linearity of expectation we get

$$\|f(X) - \mathbb{E}f(X)\|_p \leq K^d\|\langle\nabla^d f(X), G_1 \otimes \cdots \otimes G_d\rangle - \mathbb{E}_X\langle\nabla^d f(X), G_1 \otimes \cdots \otimes G_d\rangle\|_p$$
$$+ \sum_{i=1}^{d} K^i\|\langle\mathbb{E}_X\nabla^i f(X), G_1 \otimes \cdots \otimes G_i\rangle\|_p,$$

Applying now (conditionally on $G_1, \ldots, G_d$) the inequality (4.2) to the first term on the right hand side and using the Fubini theorem we obtain

$$\|\langle\nabla^d f(X), G_1 \otimes \cdots \otimes G_d\rangle - \mathbb{E}_X\langle\nabla^d f(X), G_1 \otimes \cdots \otimes G_d\rangle\|_p \leq K\|\langle\nabla^{d+1} f(X), G_1 \otimes \cdots \otimes G_{d+1}\rangle\|_p$$

(note that $\langle\nabla^d f(X), G_1 \otimes \cdots \otimes G_d\rangle$ is tetrahedral as a polynomial in X, moreover $\langle\nabla^1\langle\nabla^d f(X), G_1 \otimes \ldots \otimes G_d\rangle, G_{d+1}\rangle = \langle\nabla^{d+1} f(X), G_1 \otimes \ldots \otimes G_{d+1}\rangle$).

This ends the induction step and establishes (4.1).

If f is a polynomial of degree d, then $\nabla^d f(X)$ is deterministic (and thus equal to its expectation) so (4.1) can be written in a more concise way

$$\|f(X) - \mathbb{E}f(X)\|_p \leq \sum_{i=1}^{d} K^i\|\langle\mathbb{E}_X\nabla^i f(X), G_1 \otimes \cdots \otimes G_i\rangle\|_p$$

We will now use a result by Latała [28], which asserts the existence of constants C_k, depending only on k, such that for any k-indexed matrix A, and $p \geq 2$,

$$\|\langle A, G_1 \otimes \cdots \otimes G_k\rangle\|_p \leq C_k \sum_{\mathcal{I} \in P_k} p^{|\mathcal{I}|/2}\|A\|_{\mathcal{I}},$$

which yields

$$\|f(X) - \mathbb{E}f(X)\|_p \leq C \sum_{k=1}^{d} \sum_{\mathcal{I} \in P_k} p^{|\mathcal{I}|/2}\|\mathbb{E}\nabla^k f(X)\|_{\mathcal{I}},$$

where C depends on ρ, α, d.

By Chebyshev's inequality in L_p this gives for $p \geq 0$,

$$\mathbb{P}\left(|f(X) - \mathbb{E}f(X)| \geq Ce \sum_{k=1}^{d} \sum_{\mathcal{I} \in P_k} p^{|\mathcal{I}|/2}\|\mathbb{E}\nabla^k f(X)\|_{\mathcal{I}}\right) \leq e^{2-p}$$

(the additional factor e^2 on the right hand side allows to extend the estimate from $p \geq 2$ to all $p \geq 0$). The theorem follows now by a change of variables and adjustment of constants. $\qquad\square$

5 Convex concentration and improved estimates for positive definite quadratic forms

The estimates of Theorem 2.2 are of Gaussian nature, i.e., they have the same form as two-sided estimates valid for polynomials in independent Gaussian variables. Since the values of the random variables in the Ising model are ± 1, it is natural to look for estimates resembling those known for polynomials in independent Rademacher variables. However, in this case the situation is more complicated, as two-sided bounds are known only for polynomials of degree at most 3 (see [20, 27, 3]).

Below in Theorem 5.11 we present estimates similar in nature to those for Rademacher sequences for quadratic forms $\langle AX, X \rangle$, where A is a non-negative definite matrix and X is a random vector with bounded coefficients satisfying the approximate tensorization property. In some situations they improve on the bounds one can get for the Ising model from Theorem 2.2, however, in general they are not comparable to them, because they involve norms of the matrix A and not just its off-diagonal part (note that in the case of the Ising model, the contribution from the diagonal is deterministic). It is natural to conjecture that (similarly as for the i.i.d. case) the assumption of non-negative definiteness is an artefact of our proof and can be actually dropped, however, at present we are not able to obtain such more general bounds.

5.1 Convex concentration

As a tool for proving estimates for quadratic forms we will derive concentration inequalities for convex Lipschitz functions for measures on products of compact sets, satisfying the approximate tensorization property, which are of independent interest. In particular this will allow us to obtain concentration for linear combinations with vector coefficients (see Proposition 5.5), which generalizes estimates obtained in the Rademacher case by Dilworth and Montgomery-Smith [18] (see also [27]).

Recall that a random vector X in $\mathbb{R}^n$ has the convex concentration property with constant K if for any L-Lipschitz convex function $f \colon \mathbb{R}^n \to \mathbb{R}$, and any $t > 0$,

$$\mathbb{P}(|f(X) - \operatorname{Med} f(X)| \geq t) \leq 2\exp(-t^2/K^2L^2). \tag{5.1}$$

It is well known that the above property is up to constant equivalent to concentration around the mean, i.e.,

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq t) \leq 2\exp(-t^2/\widetilde{K}^2L^2), \tag{5.2}$$

i.e., the inequalities (5.1) and (5.2) imply each other and the constants K and $\widetilde{K}$ depend only on one another. We will say that a random vector Z satisfies the dimension-free convex concentration property with constant K if for any N, the random vector $X = (X_1, \ldots, X_N)$, where X_i are i.i.d. copies of Z, satisfies (5.1).

We will now relate the approximate tensorization property of measures on $[-1, 1]^n$ to the convex concentration property, showing in particular that if the distribution of X is given by (2.1), where $|h|_\infty \leq \alpha$ and J_{ij} satisfy the Dobrushin condition (2.3), then X satisfies the dimension-free convex concentration property with a constant depending only on α and ρ (Proposition 5.4 below). Next we will prove that convex concentration property for measures on products of compact sets can be in fact improved by taking into account the uniform bounds on the components of X. Finally, we will illustrate this phenomenon with applications to linear forms with vector coefficients and quadratic non-negatively definite forms (Proposition 5.5 and Theorem 5.11).

In order to pass from approximate tensorization of entropy to dimension-free convex concentration property, we will use weak transportation inequalities, introduced recently by Gozlan, Roberto, Samson and Tetali [22].

Let us denote by $\mathcal{P}_1(\mathbb{R}^n)$ the set of all probability measures on $\mathbb{R}^n$ with finite first moment.

Definition 5.1. *Let μ and ν be probability measures on $\mathbb{R}^n$. Assume that $\nu \in \mathcal{P}_1(\mathbb{R}^n)$. For a convex, lower semicontinuous function $\theta \colon \mathbb{R}^n \to [0, \infty]$, such that $\theta(0) = 0$ define the weak transport cost between μ and ν as*

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf_\pi \int_{\mathbb{R}^n} \theta\Big(x - \int_{\mathbb{R}^n} y p_x(dy)\Big) \mu(dx),$$

where the infimum is taken over all couplings π between μ and ν (i.e., measures on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals μ, ν) and for $x \in \mathbb{R}^n$, $p_x(\cdot)$ is the conditional measure defined (μ-almost surely) by $\pi(dxdy) = p_x(dy)\mu(dx)$.

Using probabilistic notation one can write

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf_{(X,Y)} \mathbb{E}\theta(X - \mathbb{E}(Y|X)), \tag{5.3}$$

where the infimum is taken over all pairs of random vectors (X, Y) with values in $\mathbb{R}^n \times \mathbb{R}^n$, such that X is distributed according to μ and Y according to ν.

Recall also that if μ, ν are two probability measures then the relative entropy of ν with respect to μ is given by the formula

$$H(\nu|\mu) = \mathbb{E}_\nu \log \Big(\frac{d\nu}{d\mu}\Big)$$

if ν is absolutely continuous with respect to μ and $H(\nu|\mu) = \infty$ otherwise.

Definition 5.2. *Let $\mu \in \mathcal{P}_1(\mathbb{R}^n)$ and $\theta \colon \mathbb{R}^n \to [0, \infty]$ be a convex lower semicontinuous function with $\theta(0) = 0$. We will say that μ satisfies the inequality $\overline{\mathbf{T}}_\theta$ if for every probability measure $\nu \in \mathcal{P}_1(\mathbb{R}^n)$,*

$$\max\Big(\overline{\mathcal{T}}_\theta(\nu|\mu), \overline{\mathcal{T}}_\theta(\mu|\nu)\Big) \le H(\nu|\mu). \tag{5.4}$$

The following theorem established in [22] describes connections between dimension-free convex concentration, weak transportation inequalities and log-Sobolev inequalities for convex and concave functions.

Theorem 5.3. *Let X be a random vector in $\mathbb{R}^n$ with distribution μ. The following conditions are equivalent.*

(i) *There exists K such that X has the dimension-free convex concentration property with constant K.*

(ii) *There exists c such that μ satisfies the inequality $\overline{\mathbf{T}}_\theta$ with $\theta(x) = c|x|^2$.*

(iii) *There exist $D, \lambda > 0$ such that for every $\mathcal{C}^1$-smooth convex Lipschitz function $f \colon \mathbb{R}^n \to \mathbb{R}$ bounded from below and every $\mathcal{C}^1$-smooth concave Lipschitz function $f \colon \mathbb{R}^n \to \mathbb{R}$ bounded from above such that $f(\cdot) + \lambda |\cdot|^2$ is convex,*

$$\operatorname{Ent} e^{f(X)} \le D\mathbb{E}|\nabla f(X)|^2 e^{f(X)}. \tag{5.5}$$

Moreover, for any two assertions above the constants in one of them may be taken to depend only on the constants in the other one.

Using the above result we can easily obtain the following proposition, which may be useful, e.g., in statistical applications, when dealing with i.i.d. samples drawn from the measure μ (see, e.g., [15] for a discussion of applications related to the Ising model).

Proposition 5.4. *If X is a $[-1,1]^n$-valued random vector with law μ, which satisfies the approximate tensorization $AT(C)$, then X satisfies the dimension-free convex concentration inequality with constant K, depending only on C.*

Proof. The celebrated convex distance inequality by Talagrand (see, e.g., [40, 41]) asserts that any random variable with support in $[-1,1]$ satisfies the dimension-free convex concentration property with a universal constant. In particular by Theorem 5.3 it satisfies the log-Sobolev inequality (5.5) with some universal constants D, λ. Consider any function $f\colon \mathbb{R}^n \to \mathbb{R}$, as in point (iii) of Theorem 5.3. It is easy to see that the function $x_i \mapsto f(\bar{x}_i, x_i)$ also satisfies the conditions of point (iii), this time in dimension one. We can thus apply to this function and to the measure $\mu_i(\cdot|\bar{x}_i)$ the one-dimensional version of (5.5), which together with the condition $AT(C)$ gives (5.5) with constants CD and λ. The proof is now concluded by another application of Theorem 5.3. $\square$

It is easy to see that if a measure μ supported on $[-1,1]^n$ satisfies $\overline{\mathbf{T}}_\theta$ with $\theta(x) = c|x|^2$ then it actually satisfies a stronger inequality $\overline{\mathbf{T}}_\gamma$ with $\gamma(x) = |x|^2$ if $|x|_\infty \le 2$ and $\gamma(x) = \infty$ otherwise. Indeed for the right-hand side to be finite ν must be also supported on $[-1,1]^n$, in which case by (5.3) $\overline{\mathcal{T}}_\theta(\mu|\nu) = \overline{\mathcal{T}}_\gamma(\mu|\nu)$ and $\overline{\mathcal{T}}_\theta(\nu|\mu) = \overline{\mathcal{T}}_\gamma(\nu|\mu)$. In fact, weak transportation inequalities with such strengthened cost functions can hold only for compactly supported measures (see [39]). The interest in such strengthening lies in the fact that by taking into account the boundedness of random variables, it implies concentration inequalities stronger than the subgaussian bound given by (5.2) (see, e.g., [4] for concentration results corresponding to various cost functions θ). As shown in the next proposition, such inequalities can be also easily inferred just at the level of convex concentration. To formulate this result let us introduce a family of norms on $\mathbb{R}^n$ given for $p > 0$ by the formula

$$\|x\|_{\{1\},p} = \sup\Big\{ \sum_{i=1}^n x_i y_i \colon |y| \le \sqrt{p}, |y|_\infty \le 1\Big\}. \tag{5.6}$$

It is not difficult to see that

$$\|x\|_{\{1\},p} \le \sum_{i\le p} x_i^\downarrow + \sqrt{p}\sqrt{\sum_{i>p}(x_i^\downarrow)^2} \le 2\|x\|_{\{1\},p}, \tag{5.7}$$

where $(x_i^\downarrow)_{i\le n}$ is the nonincreasing rearrangement of the sequence $(|x_i|)_{i\le n}$. In fact one has

$$\max\Big(\sum_{i\le p} x_i^\downarrow, \sqrt{p}\sqrt{\sum_{i>p}(x_i^\downarrow)^2}\Big) \le \|x\|_{\{1\},p}. \tag{5.8}$$

Such norms are equivalent to interpolation norms between the spaces ℓ_2^n and ℓ_1^n and in a probabilistic context appeared for the first time in [33, 25], where it is shown that if $\varepsilon_1, \ldots, \varepsilon_n$ are independent Rademacher variables, then for $x \in \mathbb{R}^n$ and $p \ge 2$,

$$\frac{1}{C}\|x\|_{\{1\},p} \le \Big\| \sum_{i=1}^n \varepsilon_i x_i \Big\|_p \le C\|x\|_{\{1\},p},$$

where C is a universal constant. The meaning of the subscript $\{1\}$ will become clear when we define counterparts of this norm for matrices. To keep uniform notation, we introduce it already here.

We are now ready to state the strengthened concentration result.

Proposition 5.5. *Let X be a random vector with values in $[-1,1]^n$, satisfying the convex concentration property* (5.1). *Then for any smooth convex Lipschitz function $f\colon \mathbb{R}^n \to \mathbb{R}$ and any $p > 0$,*

$$\mathbb{P}\Big(|f(X) - M| > C \sup_x \|\nabla f(x)\|_{\{1\},p}\Big) \le 4\exp(-p/K^2). \tag{5.9}$$

where M is the mean or the median of $f(X)$ and C is a universal constant.

This improves on what can be obtained from (5.2) since as one can see from (5.6), $\|x\|_{\{1\},p} \le \sqrt{p}|x|$.

Remark 5.6. By standard regularization arguments (see, e.g., [36, p. 429]) one can drop the smoothness assumptions on f by replacing $\sup_x \|\nabla f(x)\|_{\{1\},p}$ with the Lipschitz constant of f with respect to the norm dual to $\|\cdot\|_{\{1\},p}$. One can also assume that f is defined on $[-1,1]^n$ since one can extend it to $\mathbb{R}^n$ with the formula $\tilde{f}(y) = \sup_{x \in (-1,1)^n}(f(x) + \langle \nabla f(x), y - x\rangle)$ without altering the Lipschitz constant (here $\nabla f(x)$ denotes some subgradient of f at x).

Before proving the above proposition, let us illustrate it with examples to show how it improves on the usual subgaussian convex concentration (5.2).

In view of Remark 5.6, Proposition 5.5 yields the following corollary.

Corollary 5.7. *Let $X = (X_1, \ldots, X_n)$ be a $[-1,1]^n$-valued random vector satisfying the convex concentration property* (5.1). *Let $(E, \|\cdot\|)$ be a Banach space with dual $(E^*, \|\cdot\|_*)$, $v_1, \ldots, v_b \in E$ and let*

$$Z = \Big\|\sum_{i=1}^n v_i X_i\Big\|,$$

then for every $p > 0$,

$$\mathbb{P}\Big(|Z - \mathbb{E}Z| \ge \sup_{\varphi \in E^*, \|\varphi\|_* \le 1} \|(\varphi(v_i))_{i \le n}\|_{\{1\},p}\Big) \le 4e^{-p/K^2}. \tag{5.10}$$

Remark 5.8. In [18], Dilworth and Montgomery-Smith proved that if X is a vector of independent Rademacher variables, then $\mathbb{P}(Z \ge 2\mathbb{E}Z + \sup_{\varphi \in E^*, \|\varphi\|_* \le 1} \|(\varphi(v_i))_{i \le n}\|_{\{1\},p}) \le 2e^{-cp}$. The fact that one can improve this statement to concentration around mean seems to have become by then a part of the folklore.

Example 5.9. Let us now provide a simple one-dimensional example. For illustration purposes it will be more convenient to state it in terms of infinite sequences of random variables. Let thus $X_1, X_2, \ldots$ be centered random variables with values in $[-1,1]$, such that for all n, the vector $X = (X_1, \ldots, X_n)$ satisfies (5.1) with $K = 1$ (for simplicity). Define the random variable $Z = \sum_{i=1}^\infty \frac{1}{i} X_i$ (note that thanks to (5.2) this sequence converges in L_2). Then by (5.2), we get $\mathbb{P}(|Z| \ge t) \le 2\exp(-ct^2)$ for some $c > 0$. However, it is easy to see that $\|(1/i)_{i=1}^\infty\|_{\{1\},p} \simeq \log p$ (up to multiplicative constants) for $p \ge 2$ thus by Proposition 5.5 (after adjusting the constants) we obtain

$$\mathbb{P}(|Z| \ge t) \le 4e^{-e^{c't}}.$$

This Gumbel type tail decay is clearly much faster than Gaussian. Note that thanks to the one-dimensional nature this example can be in fact easily recovered directly from (5.2) by combining it with obvious pointwise bounds on the variables X_i. Since (5.10) can be equivalently restated as

$$\mathbb{P}\Big(|Z - \mathbb{E}Z| \ge \sup_{x \in T} \|x\|_{\{1\},p}\Big) \le 4e^{-p/K^2}$$

for any bounded set T of vectors and $Z = \sup_{x \in T} |\sum_i x_i X_i|$, one can easily create more complicated examples with various types of tail decay.

Proof of Proposition 5.5. The idea of the proof goes back to Talagrand and is by now classical. The main additional observation one needs to make is that exploring the boundedness of the support may lead to improved inequalities for general convex Lipschitz functions rather than just for linear functions.

We will start by proving the inequality in question with the median. Let thus $M = \text{Med } f$ and consider first the convex set $A = \{x \in [-1,1]^n \colon f(x) \leq M\}$, so that $\mathbb{P}(X \in A) \geq 1/2$. Define $g(x) = \text{dist}(x, A)$, then $\text{Med } g(X) = 0$ and by convexity of A, g is a convex function. Note that if for $z \in [-1,1]^n$, $f(z) > M + 3 \sup_x \|\nabla f(x)\|_{\{1\},p}$, then by convexity for any $y \in A$,

$$M + 3\|\nabla f(z)\|_{\{1\},p} < f(z) \leq f(y) + \langle z - y, \nabla f(z) \rangle$$

$$\leq M + 2\sum_{i \leq p}(\frac{\partial}{\partial x_i} f(z))^{\downarrow} + \frac{1}{\sqrt{p}}|z - y|\sqrt{p}\sqrt{\sum_{i > p}((\frac{\partial}{\partial x_i} f(z))^{\downarrow})^2}$$

$$\leq M + 2\|\nabla f(z)\|_{\{1\},p} + \frac{1}{\sqrt{p}}|z - y|\|\nabla f(z)\|_{\{1\},p},$$

(where we used (5.8)) and so $|z - y| > \sqrt{p}$. Taking infimum over all $y \in A$, and recalling that $\text{Med } g(X) = 0$, we obtain

$$\mathbb{P}\Big(f(X) > M + 3\sup_x \|\nabla f(x)\|_{\{1\},p}\Big) \leq \mathbb{P}(g(X) \geq \text{Med } g(X) + \sqrt{p}) \leq 2\exp(-p/K^2),$$

(5.11)

where in the last inequality we used (5.1).

As for the lower tail, we can clearly assume that f is not constant. In particular $\sup_x \|\nabla f(x)\|_{\{1\},p} > 0$. Then, denoting

$$A = \Big\{x \in [-1,1]^n \colon f(x) \leq M - 3\sup_x \|\nabla f(x)\|_{\{1\},p}\Big\},$$

by similar estimates as above one obtains that for $z \in [-1,1]^n$, if $\text{dist}(z, A) < \sqrt{p}$, then $f(z) < M$. Thus, denoting $B = \{x \in \mathbb{R}^n \colon f(x) \geq M\}$ we get $A \subseteq \{x \in [-1,1]^n \colon \text{dist}(x, B) \geq \sqrt{p}\}$. Since $A \subseteq B^c$ we can also assume that $B^c \neq \emptyset$. The function $g(x) = \text{dist}(x, B)$ is 1-Lipschitz, concave on the complement of B and can be extended to a function $\widetilde{g}(x) := \inf_{z \in B^c}(g(z) + \langle \nabla g(z), x - z \rangle)$, which is 1-Lipschitz, concave on $\mathbb{R}^n$ and non-positive on B. Moreover $\widetilde{g} = g$ on B^c. Thus $\text{Med } \widetilde{g}(X) \leq 0$ and so

$$A \subseteq \{x \in [-1,1]^n \colon g(x) \geq \sqrt{p}\} \subseteq \{x \in [-1,1]^n \colon \widetilde{g}(x) \geq \text{Med } \widetilde{g}(X) + \sqrt{p}\}.$$

As a consequence $\mathbb{P}(X \in A) \leq 2\exp(-p/K^2)$, which together with (5.11) proves that

$$\mathbb{P}\Big(|f(X) - \text{Med } f(X)| \geq 3\sup_x \|\nabla f(x)\|_{\{1\},p}\Big) \leq 4e^{-p/K^2}.$$

To pass from the median to the mean, we notice that for $t \geq 1$, $\|x\|_{\{1\},tp} \leq \sqrt{t}\|x\|_{\{1\},p}$, so applying the above estimate with $t^2 p$ instead of p, we get

$$\mathbb{P}\Big(|f(X) - \text{Med } f(X)| \geq 3t\sup_x \|\nabla f(x)\|_{\{1\},p}\Big) \leq 4\exp(-t^2 p/K^2).$$

In particular by Jensen's inequality and integration by parts this yields

$$|\mathbb{E}f(X) - \text{Med } f(X)| \leq \mathbb{E}|f(X) - \text{Med } f(X)| \leq (3 + C'Kp^{-1/2})\sup_x \|\nabla f(x)\|_{\{1\},p}$$

for some universal constant C', which gives

$$\mathbb{P}\Big(|f(X) - \mathbb{E}f(X)| \geq (6 + C'Kp^{-1/2})\sup_x \|\nabla f(x)\|_{\{1\},p}\Big) \leq 4\exp(-p/K^2).$$

If $p > K^2$, this gives (5.9) for $M = \mathbb{E}f(X)$ with $C = 6 + C'$, otherwise (5.9) is trivial, as the right hand side exceeds one. $\qquad\square$

5.2 Quadratic forms

We will now pass to quadratic forms. In order to formulate tail estimates in this case, we need to introduce two additional norms of a symmetric matrix. Following [27] we define

$$\|A\|_{\{1,2\},p} = \|(A_i)_{i\leq n}\|_{\{1\},p},$$

where $A_i = \sqrt{\sum_{j=1}^{n} a_{ij}^2}$ and

$$\|A\|_{\{1\}\{2\},p} = \sup\Big\{ \sum_{i,j=1}^{n} a_{ij}x_i y_j \colon |x|,|y| \leq \sqrt{p},\ |x|_\infty, |y|_\infty \leq 1 \Big\}.$$

We note that by (5.7),

$$\|A\|_{\{1,2\},p} \leq \sum_{i\leq p} A_i^\downarrow + \sqrt{p}\Big(\sum_{i>p}(A_i^\downarrow)^2\Big)^{1/2} \leq 2\|A\|_{\{1,2\},p},$$

which gives a simpler expression if one is interested in concentration up to dimension-free constants.

Remark 5.10. It is easy to see that

$$\|A\|_{\{1,2\},p} = \sup\Big\{ \sum_{i,j=1}^{n} a_{ij}x_{ij}\colon \quad \|x\|_{\{1,2\}} \leq \sqrt{p},\ \max_{i\leq n}\sum_{j=1}^{n} x_{ij}^2 \leq 1 \Big\},$$

which justifies the subscript $\{1,2\}$ used in the notation. It is also clear that $\|A\|_{\{1,2\},p} \leq \sqrt{p}\|A\|_{\{1,2\}} = \sqrt{p}\|A\|_{HS}$ and $\|A\|_{\{1\}\{2\},p} \leq p\|A\|_{\{1\}\{2\}} = p\|A\|_{\ell_2^n \to \ell_2^n}$.

In [27] Latała proved that there exists a universal constant C such that if $X = (\varepsilon_1,\ldots,\varepsilon_n)$, where ε_i's are independent Rademacher variables, then for any symmetric matrix with vanishing diagonal and any $p \geq 2$, one has

$$C^{-1}(\|A\|_{\{1,2\},p} + \|A\|_{\{1\}\{2\},p}) \leq \|\langle AX, X\rangle\|_p \leq C(\|A\|_{\{1,2\},p} + \|A\|_{\{1\}\{2\},p});$$

similar bounds were also obtained for cubic forms in [3]. As a consequence, by Chebyshev's and Paley–Zygmund inequalities (see [27]), in this case for any $p > 0$,

$$\mathbb{P}\Big(|\langle AX, X\rangle| \geq Ce(\|A\|_{\{1,2\},p} + \|A\|_{\{1\}\{2\},p}) \Big) \leq e^{2-p}$$

and

$$\mathbb{P}\Big(|\langle AX, X\rangle| \geq c(\|A\|_{\{1,2\},p} + \|A\|_{\{1\}\{2\},p}) \Big) \geq \min(c, e^{-p}).$$

The result we obtain for quadratic forms in dependent random variables is

Theorem 5.11. *Let X be a centered random vector with values in $[-1,1]^n$, satisfying the convex concentration property with constant K and let $A = (a_{ij})_{i,j\leq n}$ be a symmetric nonnegative definite matrix. Then there exists a constant C_K, depending only on K, such that for any $p \geq 0$,*

$$\mathbb{P}\Big(\langle AX, X\rangle - \mathbb{E}\langle AX, X\rangle \geq C_K(\|A\|_{\{1,2\},p} + \|A\|_{\{1\}\{2\},p}) \Big) \leq 4e^{-p} \qquad (5.12)$$

and

$$\mathbb{P}\Big(\langle AX, X\rangle - \mathbb{E}\langle AX, X\rangle \leq -C_K \min(\|A\|_{\{1,2\},p} + \|A\|_{\{1\}\{2\},p}, \sqrt{p}\|A\|_{\{1,2\}}) \Big) \leq 4e^{-p}.$$

$$(5.13)$$

Remark 5.12. The mean zero assumption in the above theorem is introduced only to simplify its formulation. Clearly, in the general case one can recenter the vector and handle the linear correction by Proposition 5.5.

Remark 5.13. For the Ising model the norms $\|A\|_{\{1,2\},p}$ and $\|A\|_{\{1\}\{2\},p}$ introduce unnecessary contribution from the diagonal of A, which does not influence the value of $\langle AX, X\rangle - \mathbb{E}\langle AX, X\rangle$. However, for random variables not supported on $\{-1, 1\}$, in general this contribution has to be taken into account. It is not difficult to see that up to constants it corresponds to the $\|\cdot\|_{\{1\},p}$ norm of the vector consisting of diagonal elements from A, which is consistent with estimates of Proposition 5.5 as well as tail bounds for sums of independent bounded random variables.

As already mentioned at the beginning of the section, one expects that the assumption of nonnegative definiteness of the matrix A is not needed in Theorem 5.11. In [1] it is shown that the convex concentration property (5.2) implies the Hanson–Wright inequality (2.7) (with c depending on K) for arbitrary matrices by splitting the matrix into the sum of its positive and negative definite parts and treating each of them separately (using convexity). This strategy does not work here, since the $\|\cdot\|_{\mathcal{I},p}$ norms are not invariant under conjugation and the norms of positive and negative parts can be of greater order than the corresponding norms of the original matrix. This can be seen, e.g., with a matrix $A = (a_{ij})_{i,j \leq n}$ such that $a_{1i} = a_{i1} = 1$ for $i \neq 1$ and all the other coefficients are zero. In this case for $1 \ll p \ll n$ we get $\|A\|_{\{1,2\},p} + \|A\|_{\{1\}\{2\},p} \simeq \sqrt{p}\sqrt{n}$, whereas if $A_\pm$ is the positive/negative part of A, then $\|A_\pm\|_{\{1,2\},p} + \|A_\pm\|_{\{1\}\{2\},p} \simeq p\sqrt{n}$.

For the Ising model one may hope that the assumption of nonnegative definiteness of the matrix A in Theorem 5.11 could be removed by a repetition of the proof of Theorem 2.2 with auxiliary Rademacher variables, instead of Gaussian ones, i.e., by proving that for every f (seen as a tetrahedral polynomial) and $p \geq 2$,

$$\|f(X) - \mathbb{E}f(X)\|_p \leq C\Big(\mathbb{E}\|\nabla f(X)\|_{\{1\},p}^p\Big)^{1/p}$$

(actually if one is interested only in quadratic forms, it is enough to prove it for polynomials of degree 2). We do not know if such inequality is satisfied under the assumptions of Theorem 2.2.

Example 5.14. Let us now present an example of a matrix A for which Theorem 5.11 gives a substantially better tail estimate that the one given by the Hanson–Wright inequality. One possibility is to tensorize Example 5.9, i.e., to consider the matrix $A_n = (a_{ij})_{i,j=1}^n$ given by $a_{ij} = \frac{1}{ij}$ for large values of n. Noting that $\langle AX, X\rangle = \langle v, X\rangle^2$ for $v = (1, 1/2, \ldots, 1/n)$ one can argue that this example is still rather about linear combinations than quadratic forms. Let us therefore leave the details to the Reader and instead consider the matrix A_n given by $a_{ij} = \frac{1}{(i+j)^2}$. It is easy to see that for any n, A_n is positive definite (e.g., by noting that for a standard exponential variable Y and $t \geq 0$ we have $\mathbb{E}e^{-tY} = \frac{1}{1+t}$, and using basic properties of the Laplace transform). Now both $\|A_n\|_{HS}$ and $\|A_n\|_{\ell_2^n \to \ell_2^n}$ are of order $\Omega(1)$ as $n \to \infty$, and so, if X_n is a sequence of centered random vectors in $[-1, 1]^n$ satisfying (5.2) with K independent of n, then the Hanson–Wright type inequalities (2.7) give

$$\mathbb{P}(|\langle A_n X_n, X_n\rangle - \mathbb{E}\langle A_n X_n, X_n\rangle| \geq t) \leq 2\exp(-ct)$$

for some dimension independent constant c. On the other hand, it is not difficult to check that we have $\|A_n\|_{\{1,2\},p} \leq C$ and $\|A_n\|_{\{1\}\{2\},p} \leq C \log p$ for some dimension-independent constant C. Thus Theorem 5.11 gives

$$\mathbb{P}(|\langle A_n X_n, X_n\rangle - \mathbb{E}\langle AX_n, X_n\rangle| \geq t) \leq 2e^{-e^{c't}}, \tag{5.14}$$

where c' is another dimension independent constant and we again obtain a strengthened Gumbel type behavior.

The above example is primarily an illustration of the difference between the norms $\sqrt{p}\|A\|_{HS} + p\|A\|_{\ell_2^n \to \ell_2^n}$, used in the Hanson–Wright inequality and the norms $\|A\|_{\{1,2\},p} + \|A\|_{\{1\}\{2\},p}$ used in estimates of Rademacher type given in Theorem 5.11, but in fact one can recover (5.14) by splitting appropriately the matrix A into a sum of two matrices, applying the Hanson–Wright inequality to one of them and the trivial pointwise bound to the other one (similarly as in the one-dimensional case of Example 5.9, where one can apply the pointwise bounds together with the Khintchine inequality). However, this strategy is limited, as in general there does not exist a constant C, independent of p and n, such that for all $n \times n$ matrices A and $p \geq 2$,

$$\inf\{\sqrt{p}\|B\|_{HS} + p\|B\|_{\ell_2^n \to \ell_2^n} + \|D\|_{\ell_1^n(\ell_1^n)} \colon B + D = A\} \leq C(\|A\|_{\{1,2\},p} + \|A\|_{\{1\}\{2\},p}),$$

where $\|(d_{ij})_{ij}\|_{\ell_1^n(\ell_1^n)} = \sum_{i,j \leq n} |d_{ij}|$. To see this one can consider, e.g., the matrix given in Remark 5.13 or matrices of the form $A = vv^T$, where v has one coordinate equal to 1 and the remaining ones equal to p/n, and $p \to \infty$ with n at an appropriate speed (we leave the details to the Reader). This shows that estimates of the form (5.12) do improve on the Hanson–Wright inequality.

Proof of Theorem 5.11. Denote $f(x) = \langle Ax, x \rangle$. By our assumptions this is a convex function and therefore, similarly as in the proof of Proposition 5.5, we can write for any $x, y \in [-1, 1]^n$,

$$f(x) - f(y) \leq \langle \nabla f(x), x - y \rangle \leq (2 + \frac{1}{\sqrt{p}}|x - y|)\|\nabla f(x)\|_{\{1\},p}.$$

Therefore (denoting $M = \operatorname{Med} f(X)$), if $f(x) > M + 3\|\nabla f(x)\|_{\{1\},p}$ then

$$g(x) := \operatorname{dist}\left(x, \{y \in [-1, 1]^n \colon f(y) \leq M\}\right) > \sqrt{p}$$

and thus by the convex concentration assumption applied to the function g (note that g is convex and $\operatorname{Med} g(X) = 0$), we get

$$\mathbb{P}(f(X) - M \geq 3\|\nabla f(X)\|_{\{1\},p}) \leq 2e^{-p/K^2}$$

(observe that here we bound $f(X) - M$ by a random quantity).

As $\nabla f(X) = 2AX$, we can apply Corollary 5.7 to the norm $\|\cdot\|_{\{1\},p}$ to obtain

$$\mathbb{P}(\|\nabla f(X)\|_{\{1\},p} \geq 2\mathbb{E}\|AX\|_{\{1\},p} + C\|A\|_{\{1\}\{2\},p}) \leq 4e^{-p/K^2}.$$

for some universal constant C. Combining the two last inequalities we obtain for some (new) constant C,

$$\mathbb{P}\left(\langle AX, X \rangle \geq M + C(\mathbb{E}\|AX\|_{\{1\},p} + \|A\|_{\{1\}\{2\},p})\right) \leq 6e^{-p/K^2}. \tag{5.15}$$

To get a bound on the upper tail (above the median) it thus suffices to estimate $\mathbb{E}\|AX\|_{\{1\},p}$. In [27] Latała proved that in the case when Y is a vector of independent Rademacher variables,

$$\mathbb{E}\|AY\|_{\{1\},p} \leq C(\|A\|_{\{1,2\},p} + \|A\|_{\{1\}\{2\},p}).$$

In [3] it is mentioned (see the remark before Lemma 8.4) that this inequality can be proved by a chaining argument, relying on concentration properties of vector valued

linear combinations of Rademacher variables. Since X is a centered random vector, satisfying analogous concentration properties as Y, one could follow this approach to recover the above inequality for X. However, the formulations and proofs in [27, 3] are given for general independent variables with log-concave tails and translating the arguments of [3], even if straightforward, is quite tedious. Therefore, instead we will use a recent deep result of Bednorz and Latała [8] concerning suprema of Rademacher processes, which will allow to directly reduce estimates for X to the case of random signs. Their Theorem 1.1 (in a finite dimensional formulation suitable for our purposes) asserts that if $T \subseteq \mathbb{R}^n$ then there exist $T_1, T_2 \subseteq \mathbb{R}^n$ such that $T \subseteq T_1 + T_2$,

$$\sup_{t \in T_1} \sum_{i=1}^n |t_i| \le C \mathbb{E} \sup_{t \in T} \sum_{i=1}^n \varepsilon_i t_i$$

and

$$\mathbb{E} \sup_{t \in T_2} \sum_{i=1}^n g_i t_i \le C \mathbb{E} \sup_{t \in T} \sum_{i=1}^n \varepsilon_i t_i,$$

where ε_i, g_i are sequences of i.i.d. resp. Rademacher and Gaussian variables, and C is a universal constant.

Since our X satisfies convex concentration property, it is in particular subgaussian with constant K and by another deep result, Talagrand's Majorizing Measure Theorem (see [42]), we have

$$\mathbb{E} \sup_{t \in T_2} \sum_{i=1}^n X_i t_i \le C K \mathbb{E} \sup_{t \in T_2} \sum_{i=1}^n g_i t_i$$

for any set T_2. Therefore, expressing $\|AX\|_{\{1\},p}$ as a supremum of linear combinations of X_i's and using the above estimates together with the inequality $|X_i| \le 1$, we see that

$$\mathbb{E}\|AX\|_{\{1\},p} \le C(1+K)\mathbb{E}\|AY\|_{\{1\},p} \le C'(1+K)(\|A\|_{\{1,2\},p} + \|A\|_{\{1\}\{2\},p}) \qquad (5.16)$$

for some universal constants C, C'. Going back to (5.15) we obtain

$$\mathbb{P}\Big(\langle AX, X \rangle \ge M + C(1+K)(\|A\|_{\{1,2\},p} + \|A\|_{\{1\}\{2\},p}) \Big) \le 6e^{-p/K^2}. \qquad (5.17)$$

for some (new) universal constant C. Using the fact that $\|A\|_{\{1,2\},tp} \le \sqrt{t}\|A\|_{\{1,2\},p}$ and $\|A\|_{\{1\}\{2\},tp} \le t\|A\|_{\{1\}\{2\},p}$ for $t \ge 1$, we can easily replace the right hand side by $4e^{-p}$ at the cost of changing $C(1+K)$ to some constant C_K (which can be clearly expressed explicitly in terms of C and K).

As for the lower tail, [4, Theorem 6.5 (ii)] implies that under (5.9), for any smooth convex function $f\colon [-1,1]^n \to \mathbb{R}$,

$$\mathbb{P}(f(X) \le \operatorname{Med} f(X) - C_K \mathbb{E}\|\nabla f(X)\|_{\{1\},p}) \le 4e^{-p}$$

(formally this theorem is stated for norms of the form $\|x\| = \sup\{\langle x, y \rangle \colon \theta(y) \le p\}$ with $\theta\colon \mathbb{R}^n \to \mathbb{R}$, whereas by (5.6) $\|\cdot\|_{\{1\},p}$ corresponds to $\theta(y) = |y|^2$ if $|y|_\infty \le 1$ and $\theta(y) = \infty$ otherwise, but the proof given in [4] does not use the finiteness of θ and in fact one can also deduce the result by approximating θ with finite-valued functions). In particular, the above inequality for $f(x) = \langle Ax, x \rangle$ gives

$$\mathbb{P}\Big(\langle AX, X \rangle \le M - C_K \mathbb{E}\|AX\|_{\{1\},p} \Big) \le 4e^{-p}.$$

The inequality (5.13) with the mean replaced by the median follows now by (5.16) and from the observation that

$$\mathbb{E}\|AX\|_{\{1\},p} \le \sqrt{p}\mathbb{E}|AX| \le \sqrt{p}\sqrt{\mathbb{E}|AX|^2} \le CK\sqrt{p}\|A\|_{\{1,2\}},$$

where the last inequality follows easily by (5.2) and integration by parts.

Now, (5.12) and (5.13) with the median instead of the mean yield

$$\mathbb{P}(|\langle AX, X \rangle - M| \geq C_K t \|A\|_{\{1,2\}}) \leq 4e^{-t},$$

which by another integration by parts gives $\|\mathbb{E}\langle AX, X \rangle - M\| \leq C'_K \|A\|_{\{1,2\}}$. Since for $p \geq 2$, $\|A\|_{\{1,2\}} \leq C\|A\|_{\{1,2\},p}$, this easily allows to pass from concentration around median to concentration around mean (at the cost of increasing the values of the constant C_K). $\qquad\square$

References

[1] R. Adamczak, *A note on the Hanson-Wright inequality for random vectors with dependencies*, Electron. Commun. Probab. **20** (2015), no. 72, 13. MR-3407216

[2] R. Adamczak, W. Bednorz, and P. Wolff, *Moment estimates implied by modified log-Sobolev inequalities*, ESAIM Probab. Stat. **21** (2017), 467–494. MR-3743923

[3] R. Adamczak and R. Latała, *Tail and moment estimates for chaoses generated by symmetric random variables with logarithmically concave tails*, Ann. Inst. Henri Poincaré Probab. Stat. **48** (2012), no. 4, 1103–1136. MR-3052405

[4] R. Adamczak and M. Strzelecki, *On the convex Poincaré inequality and weak transportation inequalities*, Bernoulli **25** (2019), no. 1, 341–374. MR-3892322

[5] R. Adamczak and P. 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

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

[7] W. Beckner, *Inequalities in Fourier analysis*, Ann. of Math. (2) **102** (1975), no. 1, 159–182. MR-0385456

[8] W. Bednorz and R. Latała, *On the boundedness of Bernoulli processes*, Ann. of Math. (2) **180** (2014), no. 3, 1167–1203. MR-3245015

[9] S. G. Bobkov, *The growth of L_p-norms in presence of logarithmic Sobolev inequalities*, Vestnik Syktyvkar Univ. 11.2 (2010), 92–111.

[10] A. Bonami, *Étude des coefficients de Fourier des fonctions de $L^p(G)$*, Ann. Inst. Fourier (Grenoble) **20** (1970), no. fasc. 2, 335–402 (1971). MR-0283496

[11] P. Caputo, G. Menz, and P. Tetali, *Approximate tensorization of entropy at high temperature*, Ann. Fac. Sci. Toulouse Math. (6) **24** (2015), no. 4, 691–716. MR-3434252

[12] S. Chatterjee, *Stein's method for concentration inequalities*, Probab. Theory Related Fields **138** (2007), no. 1-2, 305–321. MR-2288072

[13] J.-R. Chazottes, P. Collet, and F. Redig, *On concentration inequalities and their applications for Gibbs measures in lattice systems*, J. Stat. Phys. **169** (2017), no. 3, 504–546. MR-3711609

[14] C. Daskalakis, N. Dikkala, and G. Kamath, *Concentration of multilinear functions of the Ising model with applications to network data*, Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4–9 December 2017, Long Beach, CA, USA, 2017, pp. 12–22.

[15] C. Daskalakis, N. Dikkala, and G. Kamath, *Testing Ising models*, Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, PA, 2018, pp. 1989–2007. MR-3775918

[16] L. Devroye, A. Mehrabian, and T. Reddad, *The Minimax Learning Rate of Normal and Ising Undirected Graphical Models*, ArXiv e-prints (2018).

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

[18] S. J. Dilworth and S. J. Montgomery-Smith, *The distribution of vector-valued Rademacher series*, Ann. Probab. **21** (1993), no. 4, 2046–2052. MR-1245300

[19] R. Gheissari, E. Lubetzky, and Y. Peres, *Concentration inequalities for polynomials of contracting Ising models*, Electron. Commun. Probab. **23** (2018), Paper No. 76, 12 pp. MR-3873783

[20] E. D. Gluskin and S. Kwapień, *Tail and moment estimates for sums of independent random variables with logarithmically concave tails*, Studia Math. **114** (1995), no. 3, 303–309. MR-1338834

[21] F. Götze, H. Sambale, and A. Sinulis, *Higher order concentration for functions of weakly dependent random variables*, ArXiv e-prints (2018).

[22] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali, *Kantorovich duality for general transport costs and applications*, J. Funct. Anal. **273** (2017), no. 11, 3327–3405. MR-3706606

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

[24] D. L. Hanson and F. T. Wright, *A bound on tail probabilities for quadratic forms in independent random variables*, Ann. Math. Statist. **42** (1971), 1079–1083. MR-0279864

[25] P. Hitczenko, S. Kwapień, *On the Rademacher series. Probability in Banach spaces, 9 (Sandjberg, 1993)*, 31–36, Progr. Probab., **35**, Birkhäuser Boston, Boston, MA, 1994. MR-1308508

[26] E. Ising, *Beitrag zur Theorie des Ferromagnetismus*, Zeitschrift fur Physik **31** (1925), 253–258.

[27] R. Latała, *Tail and moment estimates for some types of chaos*, Studia Math. **135** (1999), no. 1, 39–53. MR-1686370

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

[29] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR-1849347

[30] M. J. Łuczak, *Concentration of measure and mixing for Markov chains*, Fifth Colloquium on Mathematics and Computer Science, Discrete Math. Theor. Comput. Sci. Proc., AI, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2008, pp. 95–120. MR-2508781

[31] K. Marton, *Measure concentration and strong mixing*, Studia Sci. Math. Hungar. **40** (2003), no. 1-2, 95–113. MR-2002993

[32] K. Marton, *Logarithmic Sobolev inequalities in discrete product spaces: a proof by a transportation cost distance*, ArXiv e-prints (2015).

[33] S. J. Montgomery-Smith, *The distribution of Rademacher sums*, Proc. Amer. Math. Soc. **109** (1990), no. 2, 517–522. MR-1013975

[34] E. Nelson, *Construction of quantum fields from Markoff fields*, J. Functional Analysis **12** (1973), 97–112. MR-0343815

[35] R. O'Donnell, *Analysis of Boolean functions*, Cambridge University Press, New York, 2014. MR-3443800

[36] P.M. Samson, *Concentration of measure inequalities for Markov chains and Φ-mixing processes*, Ann. Probab. **28** (2000), no. 1, 416–461. MR-1756011

[37] N. P. Santhanam and M. J. Wainwright, *Information-theoretic limits of selecting binary graphical models in high dimensions*, IEEE Trans. Inform. Theory **58** (2012), no. 7, 4117–4134. MR-2943079

[38] K. Shanmugam, R. Tandon, A. G. Dimakis, and P. Ravikumar, *On the Information Theoretic Limits of Learning Ising Models*, Proceedings of the 27th International Conference on Neural Information Processing Systems, volume 2 of NIPS'14. Cambridge, MA, USA, 2014. MIT Press (2014), 2303–2311.

[39] Y. Shu and M. Strzelecki, *A characterization of a class of convex log-Sobolev inequalities on the real line*, Ann. Inst. Henri Poincaré Probab. Stat. **54** (2018), no. 4, 2075–2091. MR-3865667

[40] M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. (1995), no. 81, 73–205. MR-1361756

[41] M. Talagrand, *New concentration inequalities in product spaces*, Invent. Math. **126** (1996), no. 3, 505–563. MR-1419006

[42] M. Talagrand, *Upper and lower bounds for stochastic processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics], vol. 60, Springer, Heidelberg, 2014, Modern methods and classical problems. MR-3184689

[43] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2018. MR-3837109