Electron. Commun. Probab. **23** (2018), no. 76, 1–12.
https://doi.org/10.1214/18-ECP173
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Concentration inequalities for polynomials of contracting Ising models

Reza Gheissari* Eyal Lubetzky† Yuval Peres‡

Abstract

We study the concentration of a degree-d polynomial of the N spins of a general Ising model, in the regime where single-site Glauber dynamics is contracting. For $d = 1$, Gaussian concentration was shown by Marton (1996) and Samson (2000) as a special case of concentration for convex Lipschitz functions, and extended to a variety of related settings by e.g., Chazottes *et al.* (2007) and Kontorovich and Ramanan (2008). For $d = 2$, exponential concentration was shown by Marton (2003) on lattices. We treat a general fixed degree d with $O(1)$ coefficients, and show that the polynomial has variance $O(N^d)$ and, after rescaling it by $N^{-d/2}$, its tail probabilities decay as $\exp(-c\, r^{2/d})$ for deviations of $r \geq C \log N$.

Keywords: Ising model; concentration of measure; contraction; independence testing.
AMS MSC 2010: 82B20; 60F05; 82C20.
Submitted to ECP on December 4, 2017, final version accepted on October 1, 2018.

1 Introduction

Concentration of measure for functions of random fields has been extensively studied (see, e.g., [8]). A prototypical example for a system where the underlying variables are weakly dependent is the high-temperature Ising model. The model, in its most general form without an external magnetic field, is a probability measure over configurations $\sigma \in \Omega_N := \{\pm 1\}^N$ (assigning spins to the sites $\{1, \ldots, N\}$), defined as follows: for a set of coupling interactions $\{J_{ij}\}_{1 \leq i,j \leq N}$, the corresponding Ising distribution π is given by

$$\pi(\sigma) = \mathcal{Z}^{-1} \exp\left[-H(\sigma)\right] \quad \text{where} \quad H(\sigma) = -\sum_{i,j} J_{ij}\sigma_i\sigma_j\,,$$

in which $\mathcal{Z}$ (the partition function) is a normalizer. For general $\{J_{ij}\}$ this includes ferromagnetic/anti-ferromagnetic models, and spin-glass systems on arbitrary graphs.

The Gaussian concentration of functions $f : \Omega_N \to \mathbb{R}$ in the high temperature regime has been studied both using analytical methods, adapting tools from the analysis of product spaces to the setting of weakly dependent random variables (see, e.g., [7, 12]), and using probabilistic tools such as coupling (cf. [1]). In the presence of arbitrary couplings $\{J_{ij}\}$, our hypothesis for capturing the high-temperature behavior of the model will be be based on contraction, as in the related works on concentration inequalities in [1, 10, 11, 14], and is closely related to the Dobrushin uniqueness condition in [7].

*Courant Institute, New York University, New York, NY 10012, USA. E-mail: reza@cims.nyu.edu
†Courant Institute, New York University, New York, NY 10012, USA. E-mail: eyal@courant.nyu.edu
‡Microsoft Research, Redmond, WA 98052, USA. E-mail: peres@microsoft.com

Definition. We say an Ising spin system π is *θ-contracting* if there exists a single-site discrete-time Markov chain (X_t) which is reversible w.r.t. π and θ-contracting, i.e.,

$$\max_{\sigma,\sigma':d_H(\sigma,\sigma')=1} W_H\Big(\mathbb{P}_\sigma(X_1\in\cdot),\mathbb{P}_{\sigma'}(X_1\in\cdot)\Big)\le\theta<1\,,$$

where $d_H(\cdot,\cdot)$ is Hamming distance, $W_H(\mu,\nu):=\inf\{\mathbb{E}[d_H(X,Y)]:(X,Y)\sim(\mu,\nu)\}$ is the d_H-Wasserstein distance, and $\mathbb{P}_\sigma$ is the probability starting from an initial state σ.

The discrete-time heat-bath Glauber dynamics for the Ising model is the reversible chain that each step updates the spin of a uniform site i via $\mathbb{P}_\pi(\sigma_i\in\cdot\mid\sigma\restriction_{\{1,\dots,N\}\setminus\{i\}})$. It is well-known that, for the Ising model with interactions J_{ij}, if $\max_i\sum_j|J_{ij}|\le 1-\alpha$, then the corresponding single-site heat-bath Glauber dynamics is θ-contracting with $\theta=1-\alpha/N$, a concrete case where our results apply (see, e.g., [4, §8] and [9, §14.2]).

In this case, for linear functions $f(\sigma)=\sum_i a_i\sigma_i$, it is known, as a special case of results of Marton [11] regarding Gaussian concentration for Lipschitz functions (see also [14] as well as [1, 6, 7, 10]) that there exists $c=c(a_1,\dots,a_N,\alpha)>0$ such that,

$$\mathbb{P}(|f(\sigma)-\mathbb{E}_\pi[f(\sigma)]|\ge u\sqrt{N})\le\exp(-cu^2)\,.$$

For bilinear forms, where $f(\sigma)=\sum_{ij}a_{ij}\sigma_i\sigma_j$, Marton [12] showed that on lattices

$$\mathbb{P}(|f(\sigma)-\mathbb{E}_\pi[f(\sigma)]|\ge uN)\le\exp(-cu)\,,$$

whereas Daskalakis *et al.* [3] showed that, for a general Ising model, in a subset of this regime (contraction as above with $\alpha>\frac{3}{4}$ vs. any $\alpha>0$), $\mathrm{Var}_\pi(f)=O(N^2\log^3 N)$.

Our main result recovers the correct variance and, up to a polynomial pre-factor, the tail probabilities for a polynomial of any fixed degree d (for matching lower bounds, one can take, for instance, the d-th power of the magnetization $f(\sigma)=\sum_i\sigma_i$).

Theorem 1. For every $\alpha,d>0$ there exists $C(\alpha,d)>0$ so that the following holds. Let π be the distribution of the Ising model on N spins with couplings $\{J_{ij}\}$ satisfying

$$\sum_{j:j\sim i}|J_{ij}|\le 1-\alpha\quad\text{for all } 1\le i\le N\,. \tag{1.1}$$

For every polynomial $f\in\mathbb{R}[\sigma_1,\dots,\sigma_N]$ of total-degree d with coefficients in $[-K,K]$,

$$\mathrm{Var}_\pi(f)\le CK^2N^d\,, \tag{1.2}$$

and for every $r>0$,

$$\mathbb{P}_\pi\Big(N^{-d/2}|f(\sigma)-\mathbb{E}_\pi[f(\sigma)]|\ge r\Big)\le CN^{d^2}\exp\bigg(-\frac{r^{2/d}}{CK^{2/d}}\bigg)\,. \tag{1.3}$$

Moreover, (1.2)–(1.3) hold for every Ising model with couplings $\{J_{ij}\}$ for which the corresponding ferromagnetic model with interactions $\{|J_{ij}|\}$ is $(1-\frac{\alpha}{N})$-contracting.

Remark 1.1. In [3], the authors used their variance bounds for bilinear forms of Ising models to bound (in terms of N and ε) the number of samples that are required to distinguish, with high probability, between a product measure and an Ising model whose (symmetrized Kullback-Leibler) distance to any product measure is at least ε. In Section 4, Theorems 4.1–4.2, we present a short application of Theorem 1 to improve the upper bounds of [3] by considering fourth-order statistics of the Ising model.

Remark 1.2. In this paper, we always consider polynomials of Ising models with no external field. As the following example shows, in the presence of an external field, such polynomials may not be concentrated. Let $\mu_i=\mathbb{E}[\sigma_i]$ for all i and expand,

$$\sum a_{ij}\sigma_i\sigma_j=\sum a_{ij}(\sigma_i-\mu_i)(\sigma_j-\mu_j)+\sum a_{ij}\sigma_i\mu_j+\sum a_{ij}\sigma_j\mu_i-\sum a_{ij}\mu_i\mu_j\,.$$

The first term on the right-hand side should have $O(N)$ fluctuations while the second and third terms $\sum_i(\sum_j a_{ij}\mu_j)\sigma_i$ can have order $N^{3/2}$ fluctuations (e.g., if $(\mu_j a_{ij})_j$ all have the same sign), implying (1.2)–(1.3) cannot hold in general under external field.

2 Concentration for quadratic functions

In this section, we prove the special and more straightforward case of concentration for quadratic functions of the Ising model.

Theorem 2.1. *For every $\alpha > 0$ there exists $C(\alpha) > 0$ so that the following holds. Let π be the distribution of the Ising model on N spins with interaction couplings $\{J_{ij}\}$ satisfying* (1.1)*. For $A = \{a_{ij}\}_{i,j=1}^N$, the function $f(\sigma) = \sum_{i,j} a_{ij}\sigma_i\sigma_j$ on Ω_N satisfies*

$$\mathrm{Var}_\pi(f) \le C\sum_{i,j}|a_{ij}|^2\,, \tag{2.1}$$

and for every $r > 0$,

$$\mathbb{P}_\pi\Big(N^{-1}\big|f(\sigma) - \mathbb{E}_\pi[f(\sigma)]\big| > r\Big) \le CN^2\exp\left(-\frac{r}{C\|A\|_\infty}\right)\,. \tag{2.2}$$

Furthermore, this holds for any $\{J_{ij}\}$ such that the Ising model is $(1-\frac{\alpha}{N})$-contracting.

Proof of (2.1). Recall that the variational formula for the spectral gap of a reversible Markov chain (X_t) with transition kernel P and stationary distribution π states that

$$\mathsf{gap} = \inf_f \frac{\mathcal{E}(f,f)}{\mathrm{Var}_\pi(f)} \quad \text{where} \quad \mathcal{E}(f,f) = \frac{1}{2}\sum_{\sigma,\sigma'}\pi(\sigma)P(\sigma,\sigma')\left|f(\sigma)-f(\sigma')\right|^2\,. \tag{2.3}$$

For any single-site discrete-time Markov chain for the Ising model, one has that

$$\max_{\sigma,\sigma'} P(\sigma,\sigma') \le \gamma/N \qquad \text{for some} \quad 0 < \gamma \le 1 \tag{2.4}$$

(e.g., under (1.1), for heat-bath Glauber dynamics $\gamma = \frac{1}{2}[1+\tanh(2(1-\alpha))]$). Thus,

$$\mathcal{E}(f,f) \le \frac{\gamma}{2N}\sum_i \mathbb{E}_\pi\left[(\nabla_i f)^2(\sigma)\right]\,, \tag{2.5}$$

where $(\nabla_i f)(\sigma) := f(\sigma) - f(\sigma^i)$ with σ^i the state obtained from σ by flipping σ_i. Moreover, as mentioned, since this chain satisfies (1.1), it is $(1-\frac{\alpha}{N})$-contracting and therefore has $\mathsf{gap} \ge \alpha/N$ by the results of [2] (see also [9, Theorem 13.1]).

Consider a linear function of the form $g = \sum a_i\sigma_i$; since $|\nabla_i g| = 2|a_i|$, one obtains that $\mathcal{E}(g,g) \le 2\gamma N^{-1}\sum_i |a_i|^2$, and therefore (2.3) implies that

$$\mathrm{Var}_\pi(g) \le \mathsf{gap}^{-1}\mathcal{E}(g,g) \le \frac{2\gamma}{\alpha}\sum_i |a_i|^2\,. \tag{2.6}$$

Returning to the function f, assume w.l.o.g. that $a_{ii} = 0$ for all i (as $\sigma_i^2 = 1$) and let $g_i(\sigma) := \sum_j (a_{ij} + a_{ji})\sigma_j$, so $|(\nabla_i f)(\sigma)| = 2|g_i(\sigma)|$. By symmetry, $\mathbb{E}_\pi[g_i(\sigma)] = 0$, and

$$\mathcal{E}(f,f) \le \frac{2\gamma}{N}\sum_i \mathrm{Var}_\pi\left(g_i(\sigma)\right) \le \frac{4\gamma^2}{\alpha N}\sum_{i,j}|a_{ij}|^2\,,$$

which, again applying (2.3), yields (2.1) for $C = \frac{4\gamma^2}{\alpha^2}$. □

We now proceed to proving the exponential tail bounds on f. Throughout the paper, we say a function f is b-Lipschitz on a set S if $|f(\sigma) - f(\sigma')| \le b\, d_H(\sigma,\sigma')$ for every $\sigma,\sigma' \in S$. A function f is b-Lipschitz if it is so on its whole domain, in our case Ω_N. For subsets of Ω_N, endowed with its graph distance, by the triangle inequality, it suffices to consider only σ,σ' that are neighbors; f is b-Lipschitz on a connected set $S \subset \Omega_N$ if

$$\max_{\sigma,\sigma' \in S: d_H(\sigma,\sigma')=1} |f(\sigma) - f(\sigma')| \le b\,.$$

Proof of (2.2). We begin by bounding the Lipschitz constant of $\frac{1}{\sqrt{N}} f$. Observe that

$$N^{-1}|f(\sigma) - f(\sigma')| \le 2N^{-1} d_H(\sigma,\sigma')\Big[\|\sigma^T A\|_\infty + \|A\sigma'\|_\infty\Big]\,,$$

in light of which, if we define

$$S_b = \Big\{\sigma \,:\, \max\big\{\|A\sigma\|_\infty, \|A^T\sigma\|_\infty\big\} \le \tfrac14 b\sqrt{N}\Big\}\,, \tag{2.7}$$

then $\frac{1}{\sqrt{N}} f$ is b-Lipschitz on S_b—note that we only consider $b \le \|A\|_\infty \sqrt{N}$.

In order to upper bound $\mathbb{P}_\pi(S_b^c)$, we will use the following version of concentration inequalities for Lipschitz functions of general contracting Markov chains [10]:

Proposition 2.2 ([10, Corollary 4.4, Eq. (4.13)], cf. [11, 14]). *Let π be the stationary distribution of a θ-contracting Markov chain with state space Ω, and suppose $g : \Omega \to \mathbb{R}$ is b-Lipschitz w.r.t. to the graph metric induced by the chain. Then for all $r > 0$,*

$$\mathbb{P}_\pi\left(|g(\sigma) - \mathbb{E}_\pi[g(\sigma)]| > r\right) \le 2\exp\left(-\frac{(1-\theta^2)r^2}{2\theta^2 b^2}\right)\,.$$

Note that for every i and every $\sigma,\sigma' \in \Omega_N$, we have $|(A\sigma)_i - (A\sigma')_i| \le \|A\|_\infty \|\sigma - \sigma'\|_1$ and so $\sigma \mapsto (A\sigma)_i$ and $\sigma \mapsto (A^T\sigma)_i$ are both $2\|A\|_\infty$-Lipschitz. By a union bound and Proposition 2.2 with $\theta = 1 - \alpha/N$ and $r = \frac14 b\sqrt{N}$, there exists $\kappa(\alpha) > 0$ such that

$$\mathbb{P}_\pi(S_b^c) \le 4N \exp\left(-\frac{(\frac{2\alpha}{N} - \frac{\alpha^2}{N^2})b^2 N}{32(1-\frac{\alpha}{N})^2(2\|A\|_\infty)^2}\right) \le 4N\exp\left(-\frac{b^2}{\kappa\|A\|_\infty^2}\right)\,. \tag{2.8}$$

Next, consider the McShane–Whitney extension ([13, 15]) of $N^{-\frac12} f$ from S_b, defined by

$$N^{-1/2}\tilde f(\eta) = \min_{\sigma \in S_b}\Big[N^{-1/2} f(\sigma) + b\cdot d_H(\eta,\sigma)\Big]\,; \tag{2.9}$$

this is designed so that f and $\tilde f$ agree on all of S_b, and $N^{-1/2}\tilde f$ is b-Lipschitz on all of Ω_N. As a result, by Proposition 2.2, there exists $C(\alpha) > 0$ such that

$$\mathbb{P}_\pi(|\tilde f(\sigma) - \mathbb{E}_\pi[\tilde f(\sigma)]| > rN) \le 2\exp(-r^2/(Cb^2))\,. \tag{2.10}$$

We also need to control the difference between the means of $f, \tilde f$:

$$\begin{aligned}|\mathbb{E}_\pi[\tilde f(\sigma)] - \mathbb{E}_\pi[f(\sigma)]| &\le \mathbb{E}_\pi\Big[|\tilde f(\sigma) - f(\sigma)|\mathbf{1}\{\sigma \in S_b^c\}\Big] \\ &\le 8\|A\|_\infty N^3 e^{-b^2/(\kappa\|A\|_\infty^2)}\,,\end{aligned} \tag{2.11}$$

where in the last line we used (2.8) to bound $\mathbb{P}_\pi(S_b^c)$, as well as the fact that

$$\max_\sigma\{|f(\sigma)|, |\tilde f(\sigma)|\} \le \|A\|_\infty N^2 + bN^{3/2} \le 2\|A\|_\infty N^2\,.$$

Now let $b = \sqrt{\|A\|_\infty r}$ and observe that if b is such that

$$|\mathbb{E}_\pi[\tilde f(\sigma)] - \mathbb{E}_\pi[f(\sigma)]| \le rN/3$$

holds (in particular, this holds for all $b > 2\sqrt{\kappa \|A\|_\infty^2 \log(\|A\|_\infty N)}$), then

$$\mathbb{P}_\pi(|f(\sigma) - \mathbb{E}_\pi[f(\sigma)]| > rN) \le \mathbb{P}_\pi(|\tilde{f}(\sigma) - \mathbb{E}_\pi[\tilde{f}(\sigma)]| > \tfrac{rN}{3}) + \mathbb{P}_\pi(|\tilde{f}(\sigma) - f(\sigma)| > \tfrac{rN}{3})\,.$$

By (2.10), and the choice of b, the first term above has

$$\mathbb{P}_\pi(|\tilde{f}(\sigma) - \mathbb{E}_\pi[\tilde{f}(\sigma)]| > rN/3) \le 2\exp\Big(-\frac{r}{9C\|A\|_\infty}\Big)\,.$$

Because $\tilde{f}(\sigma) = f(\sigma)$ for all $\sigma \in S_b$, by our choice of b,

$$\mathbb{P}_\pi(|\tilde{f}(\sigma) - f(\sigma)| > rN/3) \le \mathbb{P}_\pi(S_b^c) \le 4N\exp\Big(-\frac{r}{\kappa\|A\|_\infty}\Big)\,.$$

Replacing the requirement of $b > 2\sqrt{\kappa\|A\|_\infty^2 \log(\|A\|_\infty N)}$ with a prefactor of N^2, and combining the above two estimates, we obtain (2.2) for some new $C(\alpha) > 0$. ☐

3 Concentration for general polynomials

The proof of (1.3) when $d \ge 3$ is a more involved dynamical proof; this is needed since f might only be b-Lipschitz on connected components of the analog of S_b from (3.4), as opposed to on that entire set. We will need the following lemma to bound $\mathbb{E}_\pi[\nabla_\ell f]$.

Lemma 3.1. *For every $p, \alpha > 0$ there exists $C(p,\alpha) > 0$ such that the following holds. Consider an Ising model π with couplings $\{J_{ij}\}$ and let $\tilde{\pi}$ be the Ising measure corresponding to couplings $\{|J_{ij}|\}$. If $\tilde{\pi}$ is a $(1-\frac{\alpha}{N})$-contracting Ising system and*

$$h(\sigma) = \sum_{i_1,\ldots,i_p} b_{i_1,\ldots,i_p}\sigma_{i_1}\cdots\sigma_{i_p}$$

is a degree-p polynomial in $(\sigma_1,\ldots,\sigma_N)$ for a degree-p tensor B, then

$$|\mathbb{E}_\pi[h(\sigma)]| \le C\|B\|_\infty N^{p/2}\,.$$

Proof. Begin by considering ferromagnetic models with non-negative couplings, $\{J_{ij}\}$. It is well-known that $\mathbb{E}_\pi[\sigma_{i_1}\cdots\sigma_{i_p}] \ge 0$ in the ferromagnetic Ising model with no external field (e.g., by viewing its FK representation that enjoys monotonicity). Thus,

$$|\mathbb{E}_\pi[h(\sigma)]| \le \sum_{i_1,\ldots,i_p} |b_{i_1,\ldots,i_p}|\mathbb{E}_\pi[\sigma_{i_1}\cdots\sigma_{i_p}]\,,$$

and taking $M_p = (\|B\|_\infty)^{1/p}$, we see that

$$\sum_{i_1,\ldots,i_p} |b_{i_1,\ldots,i_p}|\mathbb{E}_\pi[\sigma_{i_1}\cdots\sigma_{i_p}] \le \mathbb{E}_\pi\Big[\Big|\sum_i M_p\sigma_i\Big|^p\Big]\,.$$

However, $\sum_i M_p\sigma_i$ is clearly an M_p-Lipschitz function, and by spin-flip symmetry of the Ising system, has mean 0, so by Proposition 2.2, there exists $\kappa(\alpha) > 0$ such that

$$\mathbb{P}_\pi\Big(\Big|\sum_i M_p\sigma_i\Big|^p > r^p N^{p/2}\Big) = \mathbb{P}_\pi\Big(\Big|\sum_i M_p\sigma_i\Big| > r\sqrt{N}\Big) \le e^{-r^2/\kappa M_p^2}\,,$$

and therefore, by integrating, $\mathbb{E}_\pi[|\sum_i M_p\sigma_i|^p] \le C\|B\|_\infty N^{p/2}$ for some $C(p,\alpha) > 0$.

Now suppose that $\{J_{ij}\}$ are not all non-negative; using the FK representation of Ising spin systems with general couplings (not necessarily ferromagnetic)—see, e.g., [5, §11.5], and in particular Proposition 259 and Eq. (11.44)—for every $i_1,\ldots,i_p$,

$$\big|\mathbb{E}_\pi[\sigma_{i_1}\cdots\sigma_{i_p}]\big| \le \mathbb{E}_{\tilde{\pi}}[\sigma_{i_1}\cdots\sigma_{i_p}]\,. \tag{3.1}$$

Then, proceeding as before, we see that

$$|\mathbb{E}_\pi[h(\sigma)]| \le \sum_{i_1,\dots,i_p} |b_{i_1,\dots,i_p}||\mathbb{E}_\pi[\sigma_{i_1}\cdots\sigma_{i_p}]| \le \mathbb{E}_{\tilde\pi}\Big[|\sum_i M_p\sigma_i|^p\Big]\,.$$

Since $\tilde\pi$ is contracting, we can apply Proposition 2.2 as before to obtain for the same constant, $C(p,\alpha)>0$ that $|\mathbb{E}_\pi[h(\sigma)]| \le \mathbb{E}_{\tilde\pi}[|\sum_i M_p\sigma_i|^p] \le C\|B\|_\infty N^{p/2}$. ☐

Proof of (1.2). Fix d and recall the variational formula for the spectral gap, (2.3). Following (2.5), we see that for γ defined in (2.4)

$$\mathcal{E}(f,f) \le \frac{\gamma}{2N}\sum_\ell \mathbb{E}_\pi\left[(\nabla_\ell f)^2(\sigma)\right]$$

where $(\nabla_\ell f)(\sigma) = f(\sigma) - f(\sigma^\ell)$ as before. Let $f(\sigma) = \sum_{i_1,\dots,i_d} a_{i_1,\dots,i_d}\sigma_{i_1}\cdots\sigma_{i_d}$ with $\|A\|_\infty \le K$, and w.l.o.g. (every polynomial can be rewritten as a sum of monomials) assume $a_{i_1,\dots,i_d}=0$ if $i_k=i_j$ for some $j\ne k$. Then for every ℓ and every σ,

$$|(\nabla_\ell f)(\sigma)| = 2\Bigg|\sum_{i_2,\dots,i_d} a_{\ell,i_2,\dots,i_d}\sigma_{i_2}\cdots\sigma_{i_d} + \cdots + \sum_{i_1,\dots,i_{d-1}} a_{i_1,\dots,i_{d-1},\ell}\sigma_{i_1}\cdots\sigma_{i_{d-1}}\Bigg|\,,$$

so that $g_\ell(\sigma) := (\nabla_\ell f)^2(\sigma)$ is a $2(d-1)$-degree polynomial in σ with coefficients bounded above by $4\binom{2(d-1)}{(d-1)}K^2$. By Lemma 3.1, there exists $C(\alpha,d)>0$ such that for every ℓ,

$$\mathbb{E}_\pi[g_\ell(\sigma)] \le 4\binom{2(d-1)}{d-1}CK^2N^{d-1}\,,$$

so that using (2.3), (2.5), and the fact that $\mathsf{gap} \ge \alpha/N$, for some new $C(\alpha,d)>0$,

$$\mathrm{Var}_\pi(f) \le \mathsf{gap}^{-1}\mathcal{E}(f,f) \le \frac{N\gamma}{2\alpha}\cdot CK^2N^{d-1} = \frac{C\gamma}{2\alpha}K^2N^d\,.$$ ☐

Proof of (1.3). As before, since $\sigma_i^2=1$, every polynomial of degree d can be rewritten as a sum of monomials of degree at most d. The concentration of the lower-degree monomials can be absorbed into the constant in the prefactor in (1.3) of Theorem 1. Moreover, it suffices by rescaling to prove the theorem for the case $K=1$. Consider a $(1-\frac{\alpha}{N})$-contracting Ising model π; for every d, if f is a monomial of degree d (i.e., $f(\sigma) = \sum_{i_1,\dots,i_d} a_{i_1,\dots,i_d}\sigma_{i_1}\cdots\sigma_{i_d}$ for a d-tensor A with $\|A\|_\infty\le 1$ and $a_{i_1\dots i_d}=0$ if $i_j=i_k$ for some $j\ne k$), there exists $\kappa(\alpha,d)>0$ so that for every $r>0$, and every N,

$$\mathbb{P}_\pi\Big(\frac{1}{N^{d/2}}\big|f(\sigma)-\mathbb{E}_\pi[f(\sigma)]\big|>r\Big) \le \kappa[N^{2+d/2}\log^2(N)]^{d-1}\exp\Big(-r^{2/d}/\kappa\Big)\,. \tag{3.2}$$

Since we are considering d fixed, throughout this section, $\lesssim$ will be with respect to constants that may depend on d. We prove (3.2) inductively over $d\ge 2$. The base case $d=1$ is given by Proposition 2.2. Now assume that for every $p\le d-1$, Eq. (3.2) holds and show it holds for d. Fix $1\le\ell\le N$ and let σ^ℓ be the configuration that differs with σ only in coordinate ℓ. For every σ, we can compute the gradient $N^{-d/2}(\nabla_\ell f)(\sigma)$ as

$$\begin{aligned} N^{-d/2}|f(\sigma)-f(\sigma^\ell)| = 2N^{-d/2}\Bigg|&\sum_{i_2,\dots,i_d} a_{\ell,i_2,\dots,i_d}\sigma_{i_2}\cdots\sigma_{i_d}+\cdots \\ &+\sum_{i_1,\dots,i_{d-1}} a_{i_1,\dots,i_{d-1},\ell}\sigma_{i_1}\cdots\sigma_{i_{d-1}}\Bigg|\,. \end{aligned} \tag{3.3}$$

Define the following set of configurations:

$$S_b = \left\{\sigma : \max_{\ell \le N; j \le d} \Big| \sum_{i_1,\dots,i_d : i_j = \ell} a_{i_1,\dots,i_d} \sigma_{i_1} \cdots \sigma_{i_{j-1}} \sigma_{i_{j+1}} \cdots \sigma_{i_d} \Big| \le \frac{1}{2d} b N^{\frac{d-1}{2}} \right\}. \tag{3.4}$$

Since f is no longer quadratic, the bounds offered by S_b on the gradients in (3.3) no longer imply that f is b-Lipschitz on the entire set S_b; instead, one only has that f is b-Lipschitz on every *connected component* of S_b. In view of this, for each $\eta \in S_b$, let $S_{\eta,b}$ to be the connected component of S_b containing η. By this definition, the triangle inequality, and (3.3), for each $\eta \in S_b$, the function $N^{-(d-1)/2} f$ is b-Lipschitz on $S_{\eta,b}$.

For every η, define the McShane–Whitney extension of $N^{-(d-1)/2} f$ from $S_{\eta,b}$ as

$$N^{-(d-1)/2} \tilde f_\eta(\sigma') = \min_{\sigma \in S_{\eta,b}} \Big[N^{-(d-1)/2} f(\sigma) + b \cdot d_H(\sigma, \sigma') \Big],$$

so that $N^{-(d-1)/2} \tilde f_\eta$ is b-Lipschitz on all of Ω_N and $\tilde f_\eta \restriction_{S_{\eta,b}} = f \restriction_{S_{\eta,b}}$.

Now let (X_t) be the single spin-flip Markov chain which was $(1 - \frac{\alpha}{N})$-contracting with stationary distribution π, and, for each η, bound

$$\mathbb{P}_\eta(N^{-d/2} |f(X_t) - \mathbb{E}_\pi[f(X_t)]| > r) \le \Phi_1 + \Phi_2 + \Psi_1 + \Psi_2\,, \tag{3.5}$$

where

$$\begin{aligned}
\Phi_1 &= \Phi_1(\eta, r) = \mathbb{P}_\eta(N^{-d/2} |\tilde f_\eta(X_t) - \mathbb{E}_\eta[\tilde f_\eta(X_t)]| > \tfrac{r}{4})\,, \\
\Phi_2 &= \Phi_2(\eta, r) = \mathbb{P}_\eta(N^{-d/2} |f(X_t) - \tilde f_\eta(X_t)| > \tfrac{r}{4})\,, \\
\Psi_1 &= \Psi_1(\eta, r) = \mathbf{1}\big\{ N^{-d/2} \big| \mathbb{E}_\eta[\tilde f_\eta(X_t)] - \mathbb{E}_\eta[f(X_t)] \big| > \tfrac{r}{4} \big\}\,, \\
\Psi_2 &= \Psi_2(\eta, r) = \mathbf{1}\big\{ N^{-d/2} \big| \mathbb{E}_\eta[f(X_t)] - \mathbb{E}_\pi[f(X_t)] \big| > \tfrac{r}{4} \big\}\,.
\end{aligned}$$

In order to bound Φ_1 we will need the following result of Luczak (see [10, Eq. (4.14)]):

Proposition 3.2 ([10]). *If (Y_t) is a θ-contracting Markov chain on Ω with stationary measure π, and $g : \Omega \to \mathbb{R}$ is b-Lipschitz w.r.t. to its induced graph metric, then*

$$\max_{y_0 \in \Omega} \mathbb{P}_{y_0} \Big(|g(Y_t) - \mathbb{E}_{y_0}[g(Y_t)]| \ge r \Big) \le 2 \exp\left(- \frac{r^2}{b^2 \sum_{i=0}^t \theta^i} \right).$$

First, by Proposition 3.2, there exists $C(\alpha) > 0$ so that for every $\eta \in S_b$ and every t,

$$\Phi_1 = \mathbb{P}_\eta\big(N^{-d/2} |\tilde f_\eta(X_t) - \mathbb{E}_\eta[\tilde f_\eta(X_t)]| > r/4 \big) \le 2 \exp\left(- \frac{r^2}{C b^2} \right). \tag{3.6}$$

Second, the fact that f and $\tilde f_\eta$ agree on $S_{\eta,b}$ implies that

$$\Phi_2 \le \mathbb{P}_\eta(\tau_{S^c_{\eta,b}} \le t) = \mathbb{P}_\eta(\tau_{S^c_b} \le t)\,, \tag{3.7}$$

where the last equality crucially used that (X_t) is a single-site dynamics (whence starting from η, exiting $S_{\eta,b}$ and exiting S_b are equivalent).

By the definition of $\tilde f_\eta$, we have that $\|\tilde f_\eta\|_\infty \le \|f\|_\infty + N \mathrm{Lip}(f \restriction_{S_{\eta,b}})$, implying that

$$\Psi_1 \le \mathbf{1} \left\{ (1 + b) N^{d/2} \mathbb{P}_\eta(\tau_{S^c_b} \le t) > \tfrac{r}{4} \right\}. \tag{3.8}$$

Finally, if we take $t \ge t_0 := t_{\mathrm{mix}}(\varepsilon)$ for $\varepsilon_r := (8 r N^{d/2})^{-1}$, we obtain,

$$\max_{\eta \in \Omega_N} N^{-d/2} \left| \mathbb{E}_\eta[f(X_t)] - \mathbb{E}_\pi[f(X_t)] \right| \le 2 N^{d/2} \varepsilon_r < r/4\,,$$

so that for all such t, for every $\eta \in \Omega_N$, we have $\Psi_2 = 0$. Because (e.g., [9]) a Markov chain that is θ-contracting with $\theta = 1 - \frac{\alpha}{N}$ has $t_{\text{mix}} \gtrsim N \log N$, by sub-multiplicativity of total variation distance to stationarity, this holds for $t_0 \asymp N \log^2(N)$.

Combining (3.5)–(3.8), we see that for all $\eta \in S_b$ and $t \geq t_0$,

$$\mathbb{P}_\eta(N^{-d/2}|f(X_t) - \mathbb{E}_\pi[f(X_t)]| > r) \leq \mathbf{1}\Big\{(1+b)N^{d/2}\mathbb{P}_\eta(\tau_{S_b^c} \leq t) > \tfrac{r}{4}\Big\} + \mathbb{P}_\eta(\tau_{S_b^c} \leq t) + 2\exp\Big(-\frac{r^2}{Cb^2}\Big).$$

If we now average both sides over $\eta \sim \pi$ and set $t = t_0$, we obtain

$$\begin{aligned}\mathbb{P}_\pi\Big(N^{-d/2}|f(X_t) - \mathbb{E}_\pi[f(X_t)] > r\Big) &\leq \mathbb{P}_\pi(\{\eta : \mathbb{P}_\eta(\tau_{S_b^c} \leq t) > r/((4+4b)N^{d/2})\}) \\ &\quad + \mathbb{P}_\pi(\tau_{S_b^c} \leq t) + \mathbb{P}_\pi(S_b^c) + 2\exp\Big(-\frac{r^2}{Cb^2}\Big) \\ &\leq \Big[2t_0 + (4+4b)r^{-1}N^{d/2}t_0\Big]\mathbb{P}_\pi(S_b^c) + 2\exp\Big(-\frac{r^2}{Cb^2}\Big), \qquad (3.9)\end{aligned}$$

where we used using stationarity of the Markov chain and a union bound over all times up to t_0, and Markov's inequality with $\mathbb{E}_\pi[\mathbb{P}_\eta(\tau_{S_b^c} \leq t)] = \mathbb{P}_\pi(\tau_{S_b^c} \leq t)$.

It remains to bound the probability $\mathbb{P}_\pi(S_b^c)$. Let, for every $1 \leq \ell \leq N$, $1 \leq j \leq d$,

$$g_{\ell,j}(\sigma) = \sum_{i_1,\ldots,i_d : i_j = \ell} a_{i_1,\ldots i_d}\sigma_{i_1}\cdots\sigma_{i_{j-1}}\sigma_{i_{j+1}}\cdots\sigma_{i_d}\,;$$

by the inductive hypothesis there exists $\kappa(\alpha, d) > 0$ such that uniformly over ℓ, j,

$$\mathbb{P}_\pi\big(|g_{\ell,j}(\sigma) - \mathbb{E}_\pi[g_{\ell,j}(\sigma)]| > (bN^{\frac{d-1}{2}})/(4d)\big) \lesssim \big[N^{2+\frac{d-1}{2}}\log^2(N)\big]^{d-2}\exp\Big(-b^{\frac{2}{d-1}}/\kappa\Big).$$

To upper bound $\mathbb{P}_\pi(S_b^c)$, by (3.4) it suffices to show that $|\mathbb{E}_\pi[g_{\ell,j}]|$ is at most $(bN^{\frac{d-1}{2}})/(4d)$ and then union bound over ℓ, j. Since for each ℓ, j, the function $g_{\ell,j}$ is a $d-1$ degree polynomial of the form of $h(\sigma)$ in Lemma 3.1 there exists $C'(\alpha, d) > 0$ such that for every $1 \leq \ell \leq N$ and $1 \leq j \leq d$, $|\mathbb{E}_\pi[g_{\ell,j}]| \leq C'N^{(d-1)/2}$. Therefore, for every $b \geq 4C'd$, by a union bound over $1 \leq \ell \leq N$ and $1 \leq j \leq d$,

$$\mathbb{P}_\pi(S_b^c) \lesssim N\big[N^{2+(d-1)/2}\log^2(N)\big]^{d-2}\exp\Big(-b^{\frac{2}{d-1}}/\kappa\Big). \qquad (3.10)$$

Plugging (3.10) into (3.9), by stationarity of π and $t_0 \asymp dN\log^2(N)$, we obtain

$$\begin{aligned}&\mathbb{P}_\pi(N^{-d/2}|f(\sigma) - \mathbb{E}_\pi[f(\sigma)]| > r) \\ &\qquad \lesssim \big[N^{2+d/2}\log^2(N)\big]^{d-1}\Big[\exp\Big(-\frac{r^2}{Cb^2}\Big) + \exp\Big(-\frac{b^{2/(d-1)}}{\kappa}\Big)\Big],\end{aligned}$$

at which point, the choice of b given by $b = r^{(d-1)/d}$ implies (3.2). □

4 An application to testing Ising models

In [3], independence testing of Ising models was extensively studied. Suppose one is given k samples of N bits, either from a product measure $\mathcal{I}$ or from an Ising measure ν satisfying (1.1), whose Kullback–Leibler distance to $\mathcal{I}$ is at least ε. The goal is to decide with high probability, using a minimum number of samples, which distribution the samples came from. Our variance bound in Theorem 1 allows us to use a fourth-order statistic to improve on the results of [3] in the high-temperature regime of (1.1), including obtaining the sharp upper bound in the case of ferromagnetic Ising models.

Consider an Ising model with couplings $\{J_{ij}\}$ satisfying 1.1 and for every $i \sim j$, let

$$\lambda_{ij}^{\pi} = \mathbb{E}_{\pi}[\sigma_x \sigma_y] - \mathbb{E}_{\pi}[\sigma_x]\mathbb{E}_{\pi}[\sigma_y]\,,$$

which in the absence of external field equals $\mathbb{E}_{\pi}[\sigma_x \sigma_y]$.

The Ising model has the special property that for two Ising models π and ν on N vertices, with couplings $\{J_{ij}^{\pi}\}$ and $\{J_{ij}^{\nu}\}$ and edge-magnetizations λ_{ij}^{π} and λ_{ij}^{ν}, the symmetrized Kullback–Leibler divergence $d_{\mathrm{SKL}}(\pi,\nu)$ is given by

$$d_{\mathrm{SKL}}(\pi,\nu) = \mathbb{E}_{\pi}\Big[\log\Big(\frac{\pi}{\nu}\Big)\Big] - \mathbb{E}_{\nu}\Big[\log\Big(\frac{\nu}{\pi}\Big)\Big] = \sum_{1\le i<j\le N} (J_{ij}^{\pi} - J_{ij}^{\nu})(\lambda_{ij}^{\pi} - \lambda_{ij}^{\nu})\,.$$

Let $\mathcal{I}$ be the product measure on N independent, symmetric ± 1 random variables. That is to say that $J_{ij}^{\mathcal{I}} = \lambda_{ij}^{\mathcal{I}} = 0$ for all i,j and $d_{\mathrm{SKL}}(\pi,\mathcal{I}) = \sum_{i,j} J_{ij}^{\pi}\lambda_{ij}^{\pi}$. Finally, for an Ising model π, let m denote the number of edges, i.e., the number of non-zero J_{ij}^{π}.

Theorem 4.1. *There exists a polynomial time algorithm that uses $O(N/\varepsilon)$ samples from a ferromagnetic Ising model π on N vertices satisfying* (1.1)*, and distinguishes with probability better than $\frac{3}{4}$, whether $\pi = \mathcal{I}$ or $d_{\mathrm{SKL}}(\pi,\mathcal{I}) \ge \varepsilon$. In the specific case where the edge set $\{(ij) : J_{ij}^{\pi} \neq 0\}$ is known, this is improved to $O(\sqrt{m}/\varepsilon)$ samples.*

Theorem 4.2. *There exists a polynomial time algorithm that uses $O(N^2/\varepsilon^2)$ samples from an Ising model π on N vertices satisfying* (1.1)*, and distinguishes with probability better than $\frac{3}{4}$ whether $\pi = \mathcal{I}$ or $d_{\mathrm{SKL}}(\pi,\mathcal{I}) \ge \varepsilon$. In the specific case where the edge set $\{(ij) : J_{ij}^{\pi} \neq 0\}$ is known a priori, this is improved to $O(N\sqrt{m}/\varepsilon^2)$ samples.*

(The previous results of [3] gave a bound of $O(m/\varepsilon)$ in the setting of Theorem 4.1, and a bound of $O(N^{10/3}/\varepsilon^2)$ in the setting of Theorem 4.2.)

The algorithms will take k i.i.d. samples $(\sigma_i^{(1)})_{i\le N}, ..., (\sigma_i^{(k)})_{i\le N}$ from π and compute

$$Z_k = Z_k(\sigma^{(1)}, ..., \sigma^{(k)}) = \sum_{i,j} \Big(\frac{1}{k}\sum_{1\le \ell\le k} \sigma_i^{(\ell)}\sigma_j^{(\ell)}\Big)^2\,; \tag{4.1}$$

in the case where we do know the edge set of the underlying graph a priori, we sum only over $i \sim j$. Let $\mathbb{P}$ be the measure given by $\bigotimes_{i=1}^{k}\pi$. Observe first that

$$\mathbb{E}[Z_k] = \sum_{i,j} (\lambda_{ij}^{\pi})^2 + \frac{1}{k}\sum_{i,j}(1 - \lambda_{ij}^{\pi}) \ge \sum_{i,j} (\lambda_{ij}^{\pi})^2\,. \tag{4.2}$$

At the same time,

$$\mathrm{Var}(Z_k(\sigma)) = \frac{1}{k^4}\mathrm{Var}\Big(\sum_{i,j}\sum_{1\le \ell,\ell'\le k} \sigma_i^{(\ell)}\sigma_j^{(\ell)}\sigma_i^{(\ell')}\sigma_j^{(\ell')}\Big)\,.$$

For every fixed k, we can view $(\sigma_i^{(\ell)})_{1\le i\le N, 1\le \ell\le k}$ as an Ising model on kN vertices, that satisfies (1.1) since it corresponds to k independent copies of an Ising model each satisfying (1.1). Therefore, by Theorem 1, specifically (1.2), we have $\mathrm{Var}(Z_k) \le CN^2/k^2$.

In the case when the underlying graph is known a priori, we have the following.

Lemma 4.3. *Consider k i.i.d. samples $\sigma^{(1)}, ..., \sigma^{(k)}$ from an Ising model π on N vertices and m edges, satisfying* (1.1)*. There exists $C(\alpha) > 0$ such that $\mathrm{Var}(Z_k) \le Cm/k^2$.*

Proof. Again view $(\sigma_i^{(\ell)})_{i,\ell}$ as an Ising model on kN vertices with measure $\pi^k = \bigotimes_{i=1}^{k}\pi$. Recall that since $\{J_{ij}^{\pi}\}$ satisfy (1.1) for $\alpha > 0$, the Ising model is $1 - \alpha/N$ contracting.

Since the spectral gap tensorizes, and π is $1-\alpha/N$ contracting, π^k has inverse spectral gap satisfying $\mathsf{gap}^{-1} \geq \alpha/N$. By the variational form of the gap (2.4)–(2.5), we have,

$$\mathrm{Var}(Z_k) \leq \mathsf{gap}^{-1}\mathcal{E}(Z_k, Z_k) \leq \frac{2\gamma}{\alpha}\sum_{i,\ell}\mathbb{E}\big[(\nabla_{i,\ell}Z_k)^2(\sigma)\big]\,.$$

Now we compute $(\nabla_{i,\ell}Z_k)^2(\sigma)$ for fixed $(i,\ell)=(i^\star,\ell^\star)$ and every σ. Expanding out,

$$\begin{aligned}(\nabla_{i^\star,\ell^\star}Z_k)^2(\sigma) &= \frac{4}{k^4}\sum_{j\sim i^\star, j'\sim i^\star}\mathbb{E}\big[\sigma_j^{\ell^\star}\sigma_{j'}^{\ell^\star}\big]\mathbb{E}\big[(\sum_{\ell\neq\ell^\star}\sigma_{i^\star}^{\ell}\sigma_j^{\ell})(\sum_{\ell'\neq\ell^\star}\sigma_{i^\star}^{\ell'}\sigma_{j'}^{\ell'})\big] \\ &= \frac{4}{k^4}\sum_{j\sim i^\star, j'\sim i^\star}\mathbb{E}\big[\sigma_j^{\ell^\star}\sigma_{j'}^{\ell^\star}\big]\bigg(\sum_{\ell\neq\ell^\star,\ell'\neq\ell^\star}\mathbb{E}\big[\sigma_{i^\star}^{\ell}\sigma_j^{\ell}\sigma_{i^\star}^{\ell'}\sigma_{j'}^{\ell'}\big]\bigg)\,.\end{aligned}$$

When $\ell=\ell'$, the summands in the second sum are given by $\mathbb{E}_\pi[\sigma_j\sigma_{j'}]$, whereas when $\ell\neq\ell'$, we have $\mathbb{E}[\sigma_{i^\star}^{\ell}\sigma_j^{\ell}\sigma_{i^\star}^{\ell'}\sigma_j^{\ell'}] = \mathbb{E}_\pi[\sigma_{i^\star}\sigma_j]\mathbb{E}_\pi[\sigma_{i^\star}\sigma_{j'}]$. Therefore,

$$\begin{aligned}(\nabla_{i^\star,\ell^\star}Z_k)^2(\sigma) &\leq \frac{4}{k^4}\sum_{j,j'\sim i^\star}|\mathbb{E}_\pi[\sigma_j\sigma_{j'}]|\bigg(k|\mathbb{E}_\pi[\sigma_j\sigma_{j'}]| + (k-1)^2|\mathbb{E}_\pi[\sigma_{i^\star}\sigma_j]||\mathbb{E}_\pi[\sigma_{i^\star}\sigma_{j'}]|\bigg) \\ &\leq \frac{4}{k^2}\sum_{j,j'\sim i^\star}\mathbb{E}_{\tilde\pi}[\sigma_j\sigma_{j'}]\,, \end{aligned}\tag{4.3}$$

where the last inequality follows as in (3.1) and $\tilde\pi$ is the ferromagnetic Ising model with couplings $J_{ij}^{\tilde\pi} = |J_{ij}^{\pi}|$ (also satisfying (1.1) with the same α). But, we can write

$$\sum_{j,j'\sim i^\star}\mathbb{E}_{\tilde\pi}[\sigma_j\sigma_{j'}] = \mathbb{E}_{\tilde\pi}\Big[\Big(\sum_j c_j\sigma_j\Big)^2\Big]\,,$$

where $c_j = \mathbf{1}\{J_{i^\star j}\neq 0\}$. For squares of 1-Lipschitz functions of contracting Ising models, we previously noted in (2.6) that

$$\mathbb{E}_{\tilde\pi}\Big[\Big(\sum_j c_j\sigma_j\Big)^2\Big] = \mathrm{Var}_{\tilde\pi}\Big(\sum_j c_j\sigma_j\Big) \leq \frac{2\gamma}{\alpha}\sum_j|c_j|^2 = \frac{2\gamma d_{i^\star}}{\alpha}\,,$$

with $d_{i^\star}$ being the number of nonzero couplings incident $i^\star$. Summing over $i^\star$, and plugging this bound into (4.3) and then into the variational form of the spectral gap, we obtain the desired bound of $\mathrm{Var}(Z_k)\leq(\frac{32\gamma^2}{\alpha^2})(\frac{m}{k^2})$. ☐

Proof of Theorem 4.1. The algorithm we use computes Z_k as defined in (4.1) for $k\geq CN/\varepsilon$ (when we know the underlying graph, $k\geq C'\sqrt{m}/\varepsilon$), then outputs that $\pi=\mathcal{I}$ if $Z_k\leq\varepsilon/4$ and outputs $d_{\mathrm{SKL}}(\pi,\mathcal{I})\geq\varepsilon$ otherwise. We first show that with probability at lteast $\frac{9}{10}$, if $\pi=\mathcal{I}$, the algorithm outputs that. Notice that $\mathbb{E}_{\mathcal{I}}[Z_k]=0$, and by the above computations of the variance, $\mathrm{Var}(Z_k)\leq CN^2/k^2$ (when we know the underlying edge set, $\mathrm{Var}(Z_k)\leq m/k^2$ by Lemma 4.3). By Chebyshev's inequality,

$$\mathbb{P}(Z_k\geq\varepsilon/4)\leq 16\varepsilon^{-2}\mathrm{Var}(Z_k)\,,$$

which, after plugging in the above two bounds on $\mathrm{Var}(Z_k)$ implies the number of samples we require of k is sufficient for the right-hand side to be at most $\frac{9}{10}$.

When π is such that $d_{\mathrm{SKL}}(\pi,\mathcal{I})\geq\varepsilon$, we again have the same bounds on $\mathrm{Var}(Z_k)$. We now lower bound $\mathbb{E}_\pi[Z_k]$ by (4.2) and the definition of $d_{\mathrm{SKL}}(\pi,\mathcal{I})$. Since π is a *ferromagnetic*, for all $J_{ij}^\pi\leq 1$ by the FKG inequality, $\lambda_{ij}^\pi\geq\tanh(J_{ij}^\pi)\geq J_{ij}^\pi/2$. Thus,

$$\mathbb{E}[Z_k]\geq\sum_{i,j}(\lambda_{ij}^\pi)^2\geq\frac{1}{2}\sum_{i\sim j}J_{ij}^\pi\lambda_{ij}^\pi\geq\frac{\varepsilon}{2}\,.$$

Applying Chebyshev's inequality to $\mathbb{P}(Z_k \le \varepsilon/4)$, we see that the number of samples we require of k is sufficient to identify that $d_{\mathrm{SKL}}(\pi, \mathcal{I}) \ge \varepsilon$ with probability at least $\frac{9}{10}$. A union bound over the cases $\pi = \mathcal{I}$ and $d_{\mathrm{SKL}}(\pi, \mathcal{I}) \ge \varepsilon$ concludes the proof. □

Proof of Theorem 4.2. The algorithm again computes the statistic, Z_k defined in (4.1), and now outputs that $\pi = \mathcal{I}$ if $Z_k \le \varepsilon^2/2N$ and outputs $d_{\mathrm{SKL}}(\pi, \mathcal{I}) \ge \varepsilon$ otherwise.

First, consider the situation $\pi = \mathcal{I}$; by similar reasoning to the proof of Theorem 4.1, after $k \ge CN^2/\varepsilon^2$, (when we know the underlying graph, $k \ge C'N\sqrt{m}/\varepsilon$), with probability at least $\frac{9}{10}$, the algorithm outputs that $\pi = \mathcal{I}$. Now suppose that π is such that $d_{\mathrm{SKL}}(\pi, \mathcal{I}) \ge \varepsilon$; we wish to lower bound $\mathbb{E}[Z_k]$. By Cauchy–Schwarz inequality,

$$\sum_{i,j} (\lambda_{ij}^{\pi})^2 \ge \frac{(\sum_{i,j} J_{ij}^{\pi} \lambda_{ij}^{\pi})^2}{\sum_{i,j} (J_{ij}^{\pi})^2} \ge \varepsilon^2 \left(\sum_{i \sim j} (J_{ij}^{\pi})^2 \right)^{-1} .$$

When (1.1) holds, for every i and some $\alpha > 0$, we have $\sum_{j : j \sim i} |J_{ij}^{\pi}| \le 1 - \alpha$. Therefore,

$$\mathbb{E}[Z_k] \ge \varepsilon^2 \left(\max_{i,j} \{|J_{ij}^{\pi}|\} \cdot \sum_i \sum_{j \sim i} |J_{ij}^{\pi}| \right)^{-1} \ge \varepsilon^2 \left(\sum_i [1 - \alpha] \right)^{-1} \ge \frac{\varepsilon^2}{N} .$$

We can then use Chebyshev's inequality to bound

$$\mathbb{P}(Z_k \le \varepsilon^2/(2N)) \le \mathbb{P}(|Z_k - \mathbb{E}[Z_k]| \ge \varepsilon^2/(2N)) \le 4\varepsilon^{-4} N^2 \mathrm{Var}(Z_k)$$

via the aforementioned bounds on $\mathrm{Var}(Z_k)$. Plugging in those bounds implies that the number of samples k we require is sufficient to identify that in this case $d_{\mathrm{SKL}}(\pi, \mathcal{I}) \ge \varepsilon$ with probability at least $\frac{9}{10}$, at which point a union bound concludes the proof. □

References

[1] J. R. Chazottes, P. Collet, C. Külske, and F. Redig. Concentration inequalities for random fields via coupling. *Probability Theory and Related Fields*, 137(1):201–225, 2007. MR-2278456

[2] M.-F. Chen. Trilogy of couplings and general formulas for lower bound of spectral gap. In *Probability towards 2000 (New York, 1995)*, volume 128 of *Lect. Notes Stat.*, pages 123–136. Springer, New York, 1998. MR-1632643

[3] C. Daskalakis, N. Dikkala, and G. Kamath. Testing Ising models. Extended abstract in Proc. of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2018), 1989–2007. MR-3775918

[4] H.-O. Georgii. *Gibbs measures and phase transitions*, volume 9 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, second edition, 2011. MR-2807681

[5] G. Grimmett. The random-cluster model. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 73–123. Springer, Berlin, 2004. MR-2023651

[6] L. A. Kontorovich and K. Ramanan. Concentration inequalities for dependent random variables via the martingale method. *Ann. Probab.*, 36(6):2126–2158, 11 2008. MR-2478678

[7] C. Külske. Concentration inequalities for functions of gibbs fields with application to diffraction and random gibbs measures. *Communications in Mathematical Physics*, 239(1):29–51, 2003. MR-1997114

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

[9] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. MR-2466937

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

[11] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6(3):556–571, 1996. MR-1392329

[12] K. Marton. Measure concentration and strong mixing. *Studia Sci. Math. Hungar.*, 40(1–2):95–113, 2003. MR-2002993

[13] E. J. McShane. Extension of range of functions. *Bull. Amer. Math. Soc.*, 40(12):837–842, 12 1934. MR-1562984

[14] P.-M. Samson. Concentration of measure inequalities for Markov chains and Φ-mixing processes. *Ann. Probab.*, 28(1):416–461, 2000. MR-1756011

[15] H. Whitney. Analytic extensions of differentiable functions defined in closed sets. *Transactions of the American Mathematical Society*, 36(1):63–89, 1934. MR-1501735

Acknowledgments. The authors thank the anonymous referee for helpful suggestions. R.G. and E.L. thank Microsoft Research for its hospitality. E.L. was supported in part by NSF grant DMS-1513403.