Massachusetts Institute of Technology

A spectral condition for spectral gap: fast mixing in high-temperature Ising models

A Spectral Condition for Spectral Gap: Fast Mixing in High-Temperature Ising Models

Ronen Eldan[*]   Frederic Koehler[†]   Ofer Zeitouni[‡]

Abstract

We prove that Ising models on the hypercube with general quadratic interactions satisfy a Poincaré inequality with respect to the natural Dirichlet form corresponding to Glauber dynamics, as soon as the operator norm of the interaction matrix is smaller than 1. The inequality implies a control on the mixing time of the Glauber dynamics. Our techniques rely on a localization procedure which establishes a structural result, stating that Ising measures may be decomposed into a mixture of measures with quadratic potentials of rank one, and provides a framework for proving concentration bounds for high temperature Ising models.

1   Introduction

In this paper we study the high temperature behavior of the Sherrington-Kirkpatrick model and more general Ising models, especially with regards to mixing of the Glauber dynamics (i.e. Gibbs sampling) chain. More precisely, if μ is the uniform measure over the hypercube $\{\pm 1\}^n$, we consider a general Ising model of the form

$$\frac{d\nu_0}{d\mu}(x) = \frac{1}{Z} \exp\left(\frac{1}{2}\langle x, Jx \rangle + \langle h, x \rangle\right) \tag{1}$$

for an arbitrary symmetric quadratic interaction matrix J and *external field* $h \in \mathbb{R}^n$, where Z, the partition function, is a normalization constant. Because the evaluation of the partition function Z is a difficult computational task, in practice samples from (1) are generally constructed by simulating a Markov chain such as the *Glauber dynamics*, where at each step (in discrete time) a site i is chosen uniformly at random from $[n]$ and the random spin X_i is resampled from its conditional law given $X_{\sim i}$. (Here and throughout, we write $x_{\sim i}$ for the collections $\{x_j\}_{j\neq i}$, with similar notation for $X_{\sim i}$.)

The behavior of Glauber dynamics in the Ising model is a classical and well-studied topic with rich connections to structural properties of the Gibbs measure ν and concentration of measure. As far as sufficient conditions for fast mixing are concerned, one of the most general

and well-known situations where rapid mixing is guaranteed is under *Dobrushin's uniqueness condition* [7], which requires that $\|J\|_{\infty \to \infty} < 1$ or equivalently that $\sum_j |J_{ij}| < 1$ for all rows i. Unfortunately, even though there exist situations where this bound is tight (the mean-field/Curie-Weiss model), in other situations of interest this bound is far from tight.

One notable model where Dobrushin's condition is not satisfied at interesting situations is the celebrated *Sherrington-Kirkpatrick* (SK) model from spin glass theory [21]. In the SK model, J is given by a rescaled matrix from the Gaussian Orthogonal Ensemble so that J is symmetric with off-diagonal entries $J_{ij} \sim \mathcal{N}(0, \beta^2/n)$, where $\beta > 0$ is a parameter specifying the *inverse temperature* of the model. Here the expected ℓ_1 norm of a row of J is on the order of $\beta\sqrt{n}$, so that Dobrushin's uniqueness condition only holds under the restrictive condition $\beta = O(1/\sqrt{n})$. Nevertheless, it is expected that in reality the Glauber dynamics are actually fast mixing for all sufficiently small constant $\beta = O(1)$ (i.e. not shrinking with n). We indeed prove this below, see Theorem 11 and Section 5.

In the classical case of ferromagnetic Ising models on a lattice, it is known that there are close connections between rapid mixing of the Glauber dynamics and functional inequalities such as the log-Sobolev inequality. In a recent breakthrough result, Bauerschmidt and Bodineau [3] proved a form of the log-Sobolev inequality for the SK model at sufficiently high temperature ($\beta < 0.25$). More precisely, they proved that if J is a positive semidefinite matrix of operator norm $\|J\|_{OP}$, then for any probability measure ρ on $\{\pm 1\}^n$,

$$D(\rho || \nu_0) \lesssim \frac{1}{1 - \|J\|_{OP}} \sum_{i=1}^n \mathbb{E}_\nu \left| \partial_i \sqrt{\frac{d\rho}{d\nu_0}}(X) \right|^2 \tag{2}$$

for any model of the form (1), where $D(\rho || \nu_0) = \mathbb{E}_\rho \log \frac{d\rho}{d\nu_0}$ is the *relative entropy* and $\partial_i f(x) = f(x_{\sim i}, x_i = 1) - f(x_{\sim i}, x_i = -1)$ is the discrete gradient on the hypercube. By a standard argument (see e.g. [15, 23]), this implies the following Poincaré-type inequality

$$\mathrm{Var}(\varphi) \lesssim \frac{1}{1 - \|J\|_{OP}} \sum_{i=1}^n \mathbb{E}_{\nu_0} |\partial_i \varphi(X)|^2 \tag{3}$$

as well. Their proof is based upon an explicit decomposition of the measure ν_0 into a mixture of product measures.

However, in the case of the SK model the estimates (2) and (3) are not known to imply polynomial time bounds on the mixing time (or relaxation time) of Glauber dynamics. The reason is a subtle discrepancy between different notions of discrete gradients. A simple example which illustrates this is the uniform measure μ_{even} on the set of vertices with even parity, $\{(x_1, \ldots, x_n \in \{\pm 1\}^n; \prod_i x_i = 1\}$. It is not hard to check that the right hand side of (3) remains unchanged if the measure μ_{even} is replaced by the uniform measure μ, therefore, if we apply the law of total variance, we see that μ_{even} satisfies a Poincaré-type inequality of this form. On the other hand, the Glauber dynamics with respect to this measure is trapped at one vertex.

For the Glauber dynamics, having spectral gap γ (or equivalently, relaxation time $1/\gamma$) is equivalent to the following Poincaré inequality (see e.g. [23]):

$$\mathrm{Var}(\varphi) \leq \frac{1}{\gamma} \mathcal{E}_{\nu_0}(\varphi, \varphi) := \frac{1}{\gamma} \mathbb{E}_{\nu_0} \sum_{i=1}^n (\mathbb{E}_{\nu_0}[\varphi(X)|X_{\sim i}] - \varphi(X))^2 \tag{4}$$

where the rhs $\mathcal{E}_{\nu_0}(\varphi, \varphi)$ here is the *Dirichlet form* corresponding to the continuous time Glauber dynamics. For distributions with full support on the hypercube $\{\pm 1\}^n$, the right hand side of (3) can be realized as the Dirichlet form for a certain dynamics in continuous time, see [17,

2

Equation (3.6)], but these dynamics can have a rate that is super-polynomial (see discussion below). Similarly, the canonical log-Sobolev inequality for the Glauber semigroup in the sense of Gross [10], which implies (4) as well as rapid mixing, replaces the sum on right hand side of (2) by the Dirichlet form $\mathcal{E}_{\nu_0}\left(\sqrt{\frac{d\rho}{d\nu_0}}, \sqrt{\frac{d\rho}{d\nu_0}}\right)$. For some models, the discrepancy between the rhs of (3) and (4) is at most a constant factor and so the difference between the Dirichlet form of the Glauber dynamics and the ℓ_2-norm of the usual discrete gradient can be disregarded.

Unfortunately, for the SK model it turns out the right hand side of (3) can be size $e^{\Theta(\beta\sqrt{n})}$ larger than the right hand side of (4) in some simple examples. (We give such an example in Appendix A.) On the other hand, it is not hard to show that in the reverse direction, the rhs of (4) is never bigger than (3) by more than a constant factor, so that (4) is a stronger estimate. This inequivalence of the two Dirichlet forms reflects the fact that the rate of the continuous-time dynamics corresponding to (3) can be as large as $e^{\Theta(\beta\sqrt{n})}$.

The main result of this paper is a proof of the Poincaré inequality (4) with $\gamma = 1 - \|J\|_{OP}$ (for J psd, as before), from which we obtain polynomial bounds on the mixing time of the Glauber dynamics. It is unclear how to obtain such a result from the product measure decomposition used to prove (2), so a key technical idea in our work is the construction of a new decomposition of the measure ν as a mixture of *rank-one* Ising models (i.e. where J has rank one). This can be thought of as a natural analogue of the *needle decomposition* used in convex geometry [12]. A structural theorem of this form, however not used directly in our result, is formulated in Section 4 below. The needle decomposition itself is generated by a natural stochastic process (a version of stochastic localization [8]) and the smooth nature of the decomposition allows us to explicitly analyze the evolution of the Dirichlet form along this process, allowing us to prove the result.

In the next section, we formulate and prove our basic Poincaré inequality, Theorem 1. The related Appendix A discusses the inequivalence between the Dirichlet forms $\mathcal{E}_\nu(\varphi, \varphi)$ and $\mathbb{E}|\nabla\varphi(X)|^2$. Section 3 is devoted to the estimate on mixing time, Theorem 11. In Section 4 we outline a structural theorem in the spirit of the needle-decompositions mentioned above. Finally, Section 5 is devoted to examples.

2   Poincaré Inequality

Recall that μ denotes the uniform measure on $\{\pm 1\}^n$, and that for a matrix J and a vector h, the Ising measure is defined as

$$d\nu_0(x) = \frac{1}{Z} e^{\frac{1}{2}\langle x, Jx\rangle + \langle h, x\rangle} d\mu. \tag{5}$$

We can clearly assume without loss of generality that J is symmetric and positive definite, which we do henceforth. For any measure ν on $\{\pm 1\}^n$, we define the *Dirichlet form*

$$\mathcal{E}_\nu(\varphi, \varphi) = \mathbb{E}_\nu \sum_{i=1}^n (\mathbb{E}_\nu[\varphi(X) \mid X_{\sim i}] - \varphi(X))^2, \tag{6}$$

where the associated generator of the Glauber dynamics is

$$(\mathcal{L}_\nu\varphi)(x) = \sum_{i=1}^n \left(\mathbb{E}_\nu[\varphi(X) \mid X_{\sim i} = x_{\sim i}] - \varphi(x)\right).$$

3

The main result of this section is a dimension-free Poincaré inequality for ν_0 under the Glauber dynamics, provided that $\|J\|_{OP} < 1$.

Theorem 1. *For ν_0 as in* (5) *with $0 \preceq J \prec \mathrm{Id}$ and any test function $\varphi : \{\pm 1\}^n \to \mathbb{R}$, we have the following inequality:*

$$(1 - \|J\|_{OP})\mathrm{Var}_{\nu_0}(\varphi(X)) \leq \mathcal{E}_{\nu_0}(\varphi, \varphi).$$

Proof. The proof proceeds by a dynamical approach. It is clearly enough to consider φ with Lipshitz norm 1 and $\varphi(\mathbf{1}) = \varphi((1, \ldots, 1)) = 0$, which implies that $\|\varphi\|_\infty \leq n$. We will introduce a path of measures

$$d\nu_t(x) = e^{c_t + \frac{1}{2}\langle x, J_t x\rangle + \langle q_t + h, x\rangle} d\mu(x) =: F_t(x) d\nu_0(x), \tag{7}$$

where J_t, q_t are processes, adapted to the filtration $\mathcal{F}_t$ generated by an n-dimensional Brownian motion W_t, and c_t is a normalization constant. For ν_t as in (7), introduce the barycenter

$$a_t := \int x \, d\nu_t(x) \tag{8}$$

and the test-function adjusted barycenter

$$V_t := \int \varphi(x)(x - a_t) \, d\nu_t(x). \tag{9}$$

To define the process $F_t(x)$, we will first need the following technical result. Define by $\mathcal{H}$ the set of all linear subspaces of $\mathbb{R}^n$.

Lemma 2. *For every $\delta > 0$, there exists a function $C : \mathbb{R}^n \times \mathcal{H} \to \mathbb{M}_{n \times n}$ which attains the following properties. For any linear subspace $H \subset \mathbb{R}^n$,*

1. *For any v the matrix $C(v, H)$ is positive semidefinite and $\mathrm{Im}(C(v, H)) \subseteq H$.*

2. *The map $v \mapsto C(v, H)$ is Lipschitz continuous.*

3. *If $dim(H) = d > 1$ then $\mathrm{Tr}(C(v, H)) \geq d - 1$.*

4. *For any v we have*

$$|C(v, H)v| \leq \delta \tag{10}$$

If the continuity assumption is ignored, then one may simple take $C(v, H)$ as the orthogonal projection onto $H \cap v^\perp$, and for the sake on intuition, the reader may think of C this way. Otherwise, the actual construction (and proof of the lemma) is postponed to subsection 2.0.1.

Fix $\delta \ll 1$ (possibly depending on n), and let C be the function provided by the above lemma. For a symmetric matrix J, introduce the subspace H_J spanned by the eigenvectors of J corresponding to positive eigenvalues (when J is positive semidefinite, this is $\mathrm{Im}(J)$).

We are finally ready to introduce the dynamics for F_t, as the solution to the system of equations

$$\frac{dJ_t}{dt} = - C(V_t, H_{J_t})^2, \qquad J_0 = J \tag{11}$$

$$dF_t(x) = F_t(x)\langle C(V_t, H_{J_t})(x - a_t), dW_t\rangle, \qquad F_0(x) = 1, \forall x \in \{\pm 1\}^n, \tag{12}$$

see (8) and (9) for the definitions of a_t, V_t. Note that the system in (11)-(12) is a stochastic differential equation of dimension $2^n + n(n + 1)/2$. The existence and uniqueness of solutions to this system follow from the next lemma, whose proof is also postponed to Subsection 2.0.1 below. In the proof of existence, we also show that J_t remains positive semidefinite, which essentially follows from the fact $\mathrm{Im}(C(V_t, H_{J_t})) \subseteq H_{J_t}$.

4

Lemma 3. *For any positive semidefinite matrix J, the system of stochastic differential equations (11)-(12) admits a unique strong solution. Furthermore, the matrix J_t is positive semidefinite for all times $t \geq 0$, almost surely.*

We continue with the proof of Theorem 1. Define now $H_t = H_{J_t}$ and $C_t = C(V_t, H_t)$. We start by verifying that the measure ν_t with density $d\nu_t(x) = F_t(x)d\nu_0(x)$ is indeed a probability measure on $\{\pm 1\}^n$. The nonnegativity of $F_t(x)$ is easily verified (it follows from (17) below) so it remains to check that the total mass of ν_t is 1. Note that, due to (8) and (12), we have for the total mass

$$z_t := \sum_{x \in \{\pm 1\}^n} F_t(x)\nu_0(x) \tag{13}$$

that

$$dz_t = \sum_{x \in \{\pm 1\}^n} F_t(x)\langle C_t(x - a_t), dW_t\rangle \nu_0(x) = \langle C_t(a_t - z_t a_t), dW_t\rangle$$

and because $z_0 = 1$, by uniqueness of the solution we have $z_t = 1$ for all time, and hence ν_t is a probability measure as claimed.

Define now

$$M_t = \int \varphi(x)F_t(x)\,d\nu.$$

We have from (12) that

$$dM_t = \left\langle \int \varphi(x)C_t(x - a_t)\,d\nu_t, dW_t \right\rangle = \langle C_t V_t, dW_t\rangle, \tag{14}$$

where $V_t := \int \varphi(x)(x - a_t)d\nu_t$. Note that

$$|d[M]_t/dt| = |C_t V_t|^2 \leq \delta^2, \text{ almost surely,} \tag{15}$$

see (10).

Define the stopping time $T = \min\{t : \text{rank}(J_t) \leq 1\}$ and let $Y_t := \text{Var}_{\nu_t}[\varphi]$. Then

$$dY_t = d\left(\int \varphi^2\,d\nu_t - M_t^2\right) = -d[M]_t + \text{martingale}.$$

Consequently, we get from (15) that

$$|\mathbb{E}Y_T - Y_0| = |\mathbb{E}Y_T - \text{Var}_{\nu_0}[\varphi]| \leq \delta^2 \mathbb{E}T. \tag{16}$$

Next, Ito's formula gives $d\log F_t(x) = \langle C_t(x - a_t), dW_t\rangle - (1/2)|C_t(x - a_t)|^2 dt$ so by integrating, we have

$$\log F_t(x) = c_t + \langle q_t, x\rangle - \langle B_t x, x\rangle/2 \tag{17}$$

with c_t, q_t being some Ito processes and with

$$B_t = \int_0^t C_s^2\,ds \tag{18}$$

(here we use that the matrix C_t is symmetric). Note that, with J_t as in (7), we obtain that

$$J_t = J - B_t, \tag{19}$$

where J is the original interaction matrix.

5

We next claim that almost surely, $T \leq \frac{1}{2}\text{Tr}(J_0) =: T_0$. Indeed, $\frac{d}{dt}\text{Tr}(J_t) \leq -2\dim(H_t) + 2 \leq -2$ for all $t < T$, which means that if $T > T_0$ then $\text{Tr}(J_t) < 0$, contradicting the positive semidefinitenessof J_t. We also deduce by monotonicity that

$$0 \preceq J_T \preceq \|J\|_{OP}\text{Id}.$$

Thus, (7) implies that

$$d\nu_T(x) = e^{c_T + \frac{1}{2}\langle U, x \rangle^2 + \langle q_T + h, x \rangle}d\mu(x) \tag{20}$$

where $|U|^2 \leq \|J\|_{OP}$.

To deduce the final result we use two more facts, proved below: Lemma 8, which says the Poincaré inequality holds for the rank one model ν_T, and Lemma 9, which says the Dirichlet form is a supermartingale under the dynamics (11)-(12). Given these facts, it follows from (16) that

$$(1-\|J\|_{OP})\text{Var}_{\nu_0}(\varphi) \leq (1-\|J\|_{OP})\mathbb{E}\text{Var}_{\nu_T}(\varphi)+\delta^2 T_0 \leq \mathbb{E}\mathcal{E}_{\nu_T}(\varphi, \varphi)+\delta^2 T_0 \leq \mathcal{E}_{\nu_0}(\varphi, \varphi)+\delta^2 T_0.$$

Taking $\delta \to 0$ proves the result. $\qquad\square$

2.0.1 Proof of the existence of the process

In this section we prove the technical lemmas 2 and 3.

Proof of Lemma 2. Introduce a smooth function $\phi : \mathbb{R}_+ \to [0, 1]$ satisfying

$$\phi(0) = 1, \qquad \phi'(0) = 0, \qquad \sup_{z\in\mathbb{R}_+} z\phi(z) \leq \delta. \tag{21}$$

For example, the function

$$\phi(z) = e^{-z^2/2\delta^2}$$

will do. Given a vector $v \in \mathbb{R}^n$ and a linear subspace H of $\mathbb{R}^n$, write $v = v_1 + v_2$ where $v_1 \in H$ and $v_2 \in H^\perp$, write $\hat{v}_1 = v_1/|v_1|$, and set

$$C(v, H) = \text{Proj}_{H \cap v^\perp} + \phi(|v_1|)\hat{v}_1 \otimes \hat{v}_1. \tag{22}$$

When $v_1 = 0$, $C(v, H)$ is just Proj_H, the orthogonal projection onto subspace H. The function $C(v, H)$ is a smooth approximation to the function $A(v, H) = \text{Proj}_{H \cap v^\perp}$; the latter is not smooth owing to a discontinuity when $|v_1|$ is small. Indeed, when $v = v_2 \in H^\perp$, we note that $A(v, H)$ is the projection onto H, while if $v = \epsilon\hat{v}_1 + v_2$ for $\epsilon > 0$, then the operator $A(v, H)$ is a projection onto a codimension 1 subspace of H. On the other hand, $C(v, H)$ smoothes this transition, at the cost that it is not a projection.

From the definition of $C(v, H)$ we have

$$|C(v, H)v| = \phi(|v_1|)|\langle \hat{v}_1, v \rangle| \leq |v_1|\phi(|v_1|) \leq \delta.$$

which shows the last claim (10) in the lemma.

We now justify the second claim of the lemma; the remaining claimsfollow directly from the definition. Rewrite $C(v, H) = \text{Proj}_H + (\phi(|v_1|) - 1)\hat{v}_1 \otimes \hat{v}_1$ and observe that the first term is constant and the second term is Lipschitz in v: this is clear away from zero, and in a neighborhood of zero it follows by rewriting the second term as $\frac{\phi(|v_1|)-1}{|v_1|^2}v_1 \otimes v_1$ and using that $\phi(0) = 1, \phi'(0) = 1$, and ϕ is smooth. $\qquad\square$

6

Proof of Lemma 3. Informally, both existence and uniqueness follow from the fact that H_{J_t} will be piecewise constant. First we note that for a fixed subspace H, the equations

$$\frac{dJ_t}{dt} = -C(V_t, H)^2 \tag{23}$$

$$dF_t(x) = F_t(x)\langle C(V_t, H)(x - a_t), dW_t\rangle \qquad \forall x \in \{\pm 1\}^n, \tag{24}$$

have Lipschitz coefficients (recall that a product of bounded Lipschitz functions is Lipschitz). Therefore, for any initial condition, a strong solution exists and is unique [13]. Consider (23)-(24) with $H := H_J = \text{Im}(J)$ and initial conditions $J_0 = J, F_0(x) = 1$ and define the stopping time $\tau_1 = \inf\{t \geq 0 : \dim(H_{J_t}) \leq \dim(H_{J_0}) - 1\}$. We use this system of equations to define J_t, F_t on the interval $[0, \tau_1]$ and observe that this solution satisfies (11)-(12), provided we show that $H_{J_t} = H$ for $t < \tau_1$, which we do next. First, observe that for $v \in \ker J$ that

$$J_t v = \left(J + \int_0^t -C(V_s, H)^2 ds \right) v = 0$$

where the first equality is by (23), and the second equality uses that $Jv = 0$ and $C(V_s, H)v = (v^T C(V_s, H))^T = 0$ using that $C(V_s, H)$ is symmetric, $\text{Im}(C(V_s, H)) \subseteq H$, and v is in $\ker J$ which is the orthogonal complement of H. Thus, $\ker J \subseteq \ker J_t$ for all $t \leq \tau_1$. Next, by the definition of τ_1, we have for all $t < \tau_1$ that $\dim(H_{J_t}) = \dim(H_{J_0})$; by a dimension count, this implies that $\ker J = \ker J_t$, that H_{J_t} has no negative eigenvalues, and finally that $H_{J_t} = H_{J_0}$ since they are both equal to the orthogonal complement of $\ker J$.

More generally, for all $i \leq \text{rank}(J_0)$ we define the stopping times

$$\tau_i = \{t \geq \tau_{i-1} : \dim(H_{J_t}) \leq \dim(H_{J_0}) - i\}$$

and define J_t, F_t on $t \in [\tau_{i-1}, \tau_i]$ by the solution to (23)-(24) with initial condition $J_{\tau_{i-1}}, F_{\tau_{i-1}}$ at time τ_{i-1} and $H = H_{J_{\tau_{i-1}}}$. Finally, define the solution for $t \geq \tau_{\text{rank}(J_0)}$ similarly, with $H = \emptyset$ (i.e. the solution is constant). This shows existence of the solution and positive semidefiniteness of J_t and essentially the same argument proves uniqueness as well. $\square$

2.1 Rank one inequality

In this section we prove the needed Poincaré inequality for rank one models (Lemma 8). We use the result of [24], which establishes a Poincaré inequality under a condition on the influence matrix referred to as the ℓ_2-Dobrushin uniqueness regime (also studied in [11, 18]).

Definition 4. For two probability measures $\mathbb{P}$ and $\mathbb{Q}$ defined over the same measure space, their *Total Variation (TV) Distance* is defined to be

$$\|\mathbb{P} - \mathbb{Q}\|_{TV} := \sup_A |\mathbb{P}(A) - \mathbb{Q}(A)|$$

where A ranges over all measurable events.

Definition 5. Suppose that X is a random vector supported on a finite set $\mathcal{X}^n$ and distributed according to ν. Define the *influence matrix A* to be the matrix with $\text{diag}(A) = 0$ and

$$A_{ij} := \max_{x_{\sim i}, x'_{\sim i}} \|\mathbb{P}_\nu[X_i = \cdot | X_{\sim i} = x_{\sim i}] - \mathbb{P}_\nu[X_i = \cdot | X_{\sim i} = x'_{\sim i}]\|_{TV}$$

where $x_{\sim i}$ and $x'_{\sim i}$ in $\mathcal{X}^{n-1}$ are allowed to differ only in coordinate j, and $\mathbb{P}_\nu[X_i = \cdot | X_{\sim i} = x_{\sim i}]$ denotes the conditional law of X_i under ν given $X_{\sim i} = x_{\sim i}$. We say that (the law of) X satisfies the ℓ_2-*Dobrushin uniqueness condition* if $\|A\|_{OP} < 1$.

7

Note that in contrast to the interaction matrix J, the influence matrix A has nonnegative entries. We specialize the following Theorem to the setting of spins valued in $\{\pm 1\}$, though it holds in more general settings.

Theorem 6 (Theorem 2.1 of [24])**.** *Suppose that $X \sim \nu$ is a random vector valued in the hypercube $\{\pm 1\}^n$ and let A be the corresponding influence matrix (as in Definition 5). For any test function $\varphi : \{\pm 1\}^n \to \mathbb{R}$,*

$$(1 - \|A\|_{OP})\mathrm{Var}(\varphi) \leq \mathcal{E}_\nu(\varphi, \varphi)$$

where $\mathcal{E}_\nu(\varphi, \varphi)$ is the Dirichlet form associated to the Glauber dynamics under ν.

To use this Theorem, we need to upper bound the spectral norm of the influence matrix for rank one models, which we do in the following Lemma.

Lemma 7. *Suppose that*

$$d\nu(x) = \exp\left(\frac{1}{2}\langle x, u\rangle^2 + \langle h, x\rangle - c\right) d\mu(x) \tag{25}$$

where μ is the uniform measure on $\{\pm 1\}^n$. The influence matrix A of ν (from Definition 5) satisfies $\|A\|_{OP} \leq |u|^2$.

Proof. First observe that

$$\mathbb{E}[X_i \mid X_{\sim i}] = \tanh(u_i\langle X_{\sim i}, u\rangle + h_i).$$

Therefore, from the definition of A_{ij} and since $\tanh(\cdot)$ is 1-Lipschitz, we have

$$A_{ij} = \frac{1}{2} \max_{x_{\sim i}, x'_{\sim i}} |\mathbb{E}_\nu[X_i|X_{\sim i} = x_{\sim i}] - \mathbb{E}_\nu[X_i|X_{\sim i} = x'_{\sim i}]| \leq |u_i u_j| \tag{26}$$

where $x_{\sim i}, x'_{\sim i}$ range over vectors in $\{\pm 1\}^n$ differing only in coordinate j. Define v to be the element-wise absolute value of u, i.e. $v_i = |u_i|$. Since A is a matrix with nonnegative entries, it follows from the Perron-Frobenius Theorem and (26) that $\|A\|_{OP} \leq \|vv^T\|_{OP} = |v|^2 = |u|^2$. $\square$

Combining Lemma 7 and Theorem 6 yields the desired Poincaré inequality for rank one models.

Lemma 8. *Suppose that ν, u are as in* (25)*. Then for any test function $\varphi : \{\pm 1\}^n \to \mathbb{R}$,*

$$(1 - |u|^2)\mathrm{Var}(\varphi) \leq \mathcal{E}_\nu(\varphi, \varphi)$$

where $\mathcal{E}_\nu$ is the Dirichlet form associated to the Glauber dynamics under ν.

2.2 The Dirichlet form is a supermartingale

Lemma 9. *Let W_t be a Brownian motion adapted to a filtration $\mathcal{F}_t$. Let C_t be a matrix-valued process adapted to $\mathcal{F}_t$. Let ν_0 be an arbitrary measure on $\{\pm 1\}^n$ and suppose that F_t is a solution to the SDE*

$$dF_t(x) = F_t(x)\langle C_t(x - a_t), dW_t\rangle, \qquad F_0(x) = 1, \forall x \in \{\pm 1\}^n$$

where $d\nu_t(x) = F_t(x)d\nu_0(x)$ and $a_t = \int x d\nu_t(x)$. Let $\varphi : \{\pm 1\}^n \to \mathbb{R}$ be an arbitrary test function. Then the Dirichlet form $\mathcal{E}_{\nu_t}(\varphi, \varphi)$ is a supermartingale.

Proof. We use that the Dirichlet form can be rewritten as

$$\mathcal{E}_{\nu_t}(\varphi, \varphi) = \sum_{x \sim y} \frac{\nu_t(x)\nu_t(y)}{\nu_t(x) + \nu_t(y)}(\varphi(x) - \varphi(y))^2 \tag{27}$$

where $x \sim y$ denotes the adjacency relation on the hypercube, i.e. x and y differ in exactly one coordinate. To see this, consider arbitrary ν, let X and Y be two independent samples from ν, and observe

$$\begin{aligned}
\mathcal{E}_\nu(\varphi, \varphi) &= \mathbb{E}_\nu \sum_{i=1}^n \mathrm{Var}(\varphi(X) \mid X_{\sim i}) \\
&= \frac{1}{2}\mathbb{E}_\nu \sum_{i=1}^n \mathbb{E}_\nu[(\varphi(Y) - \varphi(X))^2 \mid Y_{\sim i} = X_{\sim i}, X_{\sim i}] \\
&= \frac{1}{2}\mathbb{E}_\nu \sum_{i=1}^n \frac{(\varphi(Y) - \varphi(X))^2 \cdot \mathbf{1}[Y_{\sim i} = X_{\sim i}]}{\mathbb{P}_\nu(Y_{\sim i} = X_{\sim i} \mid X_{\sim i})} \\
&= \sum_{x \sim y} \nu(x)\nu(y)(\varphi(x) - \varphi(y))^2 \sum_{i=1}^n \frac{\mathbf{1}[y_{\sim i} = x_{\sim i}]}{\mathbb{P}_\nu(Y_{\sim i} = x_{\sim i})} \\
&= \sum_{x \sim y} \frac{\nu(x)\nu(y)}{\nu(x) + \nu(y)}(\varphi(x) - \varphi(y))^2 \tag{28}
\end{aligned}$$

where in the second equality we used the identity $\mathrm{Var}(X) = \frac{1}{2}\mathbb{E}[(X-Y)^2]$ for Y an independent copy of X.

Given this, it suffices to show that $\frac{\nu_t(x)\nu_t(y)}{\nu_t(x)+\nu_t(y)}$ is a supermartingale for fixed $x \sim y$. Let us calculate the Ito differential $d\frac{\nu_t(x)\nu_t(y)}{\nu_t(x)+\nu_y(y)}$. We have by Ito's Lemma,

$$d\log\nu_t(x) = \langle dW_t, \tilde{x}\rangle - \frac{1}{2}|\tilde{x}|^2 dt.$$

and

$$d\log\nu_t(y) = \langle dW_t, \tilde{y}\rangle - \frac{1}{2}|\tilde{y}|^2 dt.$$

where $\tilde{x} = C_t(x - a_t), \tilde{y} = C_t(y - a_t)$. Moreover,

$$d\log(\nu_t(x) + \nu_t(y)) = \langle dW_t, \alpha\tilde{x} + \beta\tilde{y}\rangle - \frac{1}{2}|\alpha\tilde{x} + \beta\tilde{y}|^2 dt$$

where $\alpha = \frac{\nu_t(x)}{\nu_t(x)+\nu_t(y)}, \beta = \frac{\nu_t(y)}{\nu_t(x)+\nu_t(y)}$. Therefore,

$$\begin{aligned}
&d\left(\frac{\nu_t(x)\nu_t(y)}{\nu_t(x) + \nu_y(y)}\right) \\
&= \left(\frac{\nu_t(x)\nu_t(y)}{\nu_t(x) + \nu_y(y)}\right)\frac{1}{2}\left(|\alpha\tilde{x} + \beta\tilde{y}|^2 + |\beta\tilde{x} + \alpha\tilde{y}|^2 - |\tilde{x}|^2 - |\tilde{y}|^2\right) dt + \text{ martingale.}
\end{aligned}$$

So again by Ito's Lemma, since $de^{S_t} = e^{S_t}dS_t + \frac{1}{2}e^{S_t}d[S]_t$, we have

$$d\left(\frac{\nu_t(x)\nu_t(y)}{\nu_t(x) + \nu_y(y)}\right) = \left(\frac{\nu_t(x)\nu_t(y)}{\nu_t(x) + \nu_y(y)}\right)\frac{1}{2}\left(|\alpha\tilde{x} + \beta\tilde{y}|^2 + |\beta\tilde{x} + \alpha\tilde{y}|^2 - |\tilde{x}|^2 - |\tilde{y}|^2\right) dt + \text{ martingale.}$$

By convexity of $|\cdot|^2$ and since $\alpha + \beta = 1$, the above expression is a supermartingale. $\quad\square$

9

3 Consequences for mixing time

By standard arguments which we now recall, the Poincaré inequality implies mixing time estimates for the Glauber dynamics. For a Markov semigroup $P_t = e^{t\Lambda}$, reversible with respect to ν, a Poincaré inequality

$$\gamma \mathrm{Var}_\nu[\varphi] \leq \mathcal{E}_\nu(\varphi, \varphi)$$

is equivalent to a spectral gap estimate:

$$\gamma \leq \lambda_1 - \lambda_2$$

where λ_1, λ_2 are the top two eigenvalues of the transition rate matrix Λ (see e.g. [23]). The quantity $1/\gamma$ is known as the *relaxation time* of the Markov chain. As usual, we let $P_t(\cdot, \cdot)$ denote the transition kernel of the Markov chain. Linear algebraic arguments establish the following mixing time estimate:

Theorem 10 (Theorem 20.6 of [16])**.** *Let P_t be a reversible Markov semigroup over the finite state space Ω, with stationary measure π and spectral gap γ. Then,*

$$\max_{x \in \Omega} \|P_t(x, \cdot) - \pi\|_{TV} \leq \epsilon$$

as long as

$$t \geq \frac{1}{\gamma} \log \frac{1}{\epsilon \min_{x \in \Omega} \pi(x)}.$$

Applied to our situation, we have $\gamma = 1 - \|J\|_{OP}$ by Theorem 1 and

$$\min_{x \in \{\pm 1\}^n} \nu_0(x) \geq 2^{-n} e^{-2n\|J\|_{OP} - 2|h|_1}$$

from the definition and Hölder's inequality. As a result, we obtain the following mixing time estimate for Glauber dynamics:

Theorem 11. *For P_t the continuous time Glauber dynamics on ν_0 defined in* (5)*,*

$$\max_{x \in \{\pm 1\}^n} \|P_t(x, \cdot) - \nu_0\|_{TV} \leq \epsilon$$

as long as

$$t \geq \frac{1}{1 - \|J\|_{OP}} \left((1 + 2\|J\|_{OP})n + 2|h|_1 + \log \frac{1}{\epsilon} \right).$$

One unit of time for the continuous dynamics corresponds to a Poissonian (with parameter n) number of steps of the discrete-time Glauber dynamics. Correspondingly, Theorem 1 implies an $O\left(\frac{n^2 + \|h\|_1 n + n \log(1/\epsilon)}{1 - \|J\|_{OP}} \right)$ mixing time estimate for the discrete time Glauber dynamics, using Theorem 12.3 of [16] in place of Theorem 10 above.

4 A needle decomposition theorem

In this section we formulate a structural theorem, which follows as a byproduct of our proof. As mentioned above, this theorem may be thought of as an analogue to the technique due to Kannan, Lovász and Simonovitz [12] used in convex geometry (this technique was later generalized to the context of Riemannian manifolds, see [14]). It roughly states that measures with

10

arbitrary quadratic potentials can be decomposed into mixtures of measures whose potentials are quadratic of rank one, in a way that: (i) The integral of some test function φ is preserved, and (ii) the operator norm of the quadratic potential does not increase.

For $v, u \in \mathbb{R}^n$, consider the measure $w_{v,u}$ on $\{\pm 1\}^n$ defined as

$$\frac{dw_{u,v}}{d\mu}(x) = \frac{1}{Z_{u,v}} \exp\left(\frac{1}{2}\langle u, x\rangle^2 + \langle v, x\rangle\right)$$

where $Z_{u,v} = \int_{\{\pm 1\}^n} e^{\frac{1}{2}\langle u,x\rangle^2 + \langle v,x\rangle} d\mu$.

Following roughly the same lines as the proof of Theorem 1 gives rise to the following result.

Theorem 12. *Let ν_0 be a probability measure on $\{\pm 1\}^n$ of the form*

$$\frac{d\nu_0}{d\mu}(x) = \frac{1}{Z} \exp\left(\frac{1}{2}\langle x, Jx\rangle + \langle h, x\rangle\right), \tag{29}$$

where J is positive definite, and let $\varphi : \{\pm 1\}^n \to \mathbb{R}$. There exists a probability measure m on $\mathbb{R}^n \times \mathbb{R}^n$ such that ν_0 admits the decomposition

$$\nu_0(A) = \int w_{u,v}(A) dm(u,v), \ \ \forall A \subset \{\pm 1\}^n,$$

with the properties that:

1. *m-almost surely, (u,v) are such that*

$$\int \varphi \, dw_{u,v} = \int \varphi \, d\nu_0,$$

2. *such that for m-almost every (u,v), we have $|u| \leq \|J\|_{OP}$,*

3. *and such that for all $x \sim y$ on the hypercube, we have the following inequality of conductances:*

$$\int \frac{w_{u,v}(x)w_{u,v}(y)}{w_{u,v}(x) + w_{u,v}(y)} dm(u,v) \leq \frac{\nu_0(x)\nu_0(y)}{\nu_0(x) + \nu_0(y)}.$$

Proof. (sketch) The decomposition follows by considering the evolution defined in (11) and (12) and defining the measure m according to the decomposition implied by equation (20) above. The last property is a consequence of the corresponding supermartingale property shown in the proof of Lemma 9. $\qquad \square$

The theorem can be used to reduce the concentration of a test function φ over an Ising model to concentration over the measures $w_{u,v}$, as demonstrated by the following corollary.

Corollary 13. *Let $K > 0$ and let $\varphi : \{\pm 1\}^n \to \mathbb{R}$ be a function such that for all $u, v \in \mathbb{R}^n$ with $|u| \leq K$ one has $\mathrm{Var}_{w_{u,v}}[\varphi] \leq 1$. Then for every ν_0 of the form (29) with $\|J\|_{OP} \leq K$, one has $\mathrm{Var}_{\nu_0}[\varphi] \leq 1$.*

11

Proof. Applying Theorem 12, the law of total variance implies that

$$\text{Var}_{\nu_0}[\varphi] = \int_{\mathbb{R}^n \times \mathbb{R}^n} \left(\int \varphi dw_{u,v} - \mathbb{E}_{\nu_0}[\varphi] \right)^2 dm(u,v) + \int_{\mathbb{R}^n \times \mathbb{R}^n} \text{Var}_{w_{u,v}}[\varphi] dm(u,v)$$

$$= \int_{\mathbb{R}^n \times \mathbb{R}^n} \text{Var}_{w_{u,v}}[\varphi] dm(u,v) \leq \sup_{|u| \leq K, v \in \mathbb{R}^n} \text{Var}_{w_{u,v}}[\varphi].$$

The result follows readily. □

Similarly, and analogous to the needle decomposition for convex sets, it allows us to establish functional inequalities for Ising models by reducing to the case of the rank one measures $w_{u,v}$; not just the Poincaré inequality, but also related inequalities such as the Log-Sobolev inequality. For the (bounded rate) Glauber dynamics, there is no uniform Log-Sobolev inequality over the class of models we consider, as there is no such inequality for biased product measures [6] which are a special case. In subsequent work the Modified Log-Sobolev Inequality has been shown using this reduction [1].

5 Some examples

Sherrington-Kirkpatrick (SK) Model. This is the Ising model with J symmetric and $J_{ij} \sim \mathcal{N}(0, \beta^2/n)$, i.e. up to rescaling J is drawn from the Gaussian Orthogonal Ensemble. Letting the diagonal of J be 0, the spectrum of a GOE is contained in $[-2 - \epsilon, 2 + \epsilon]$ asymptotically almost surely for any $\epsilon > 0$ (see e.g. [2]). Therefore our result implies the Poincaré inequality and polynomial time mixing for all $\beta < 1/4$.

Remark 14. The Poincaré inequality (Theorem 1) applied to linear functions gives

$$\text{Var}(\langle w, X \rangle) \leq \frac{1}{1 - \|J\|_{OP}} \mathbb{E}_\nu \sum_{i=1}^n (\mathbb{E}_\nu[\langle w, X \rangle | X_{\sim i}] - \langle w, X \rangle)^2 \leq \frac{1}{1 - \|J\|_{OP}} |w|^2$$

using (28), and estimate (3) from [3] gives a similar bound with a different constant. Thus, both results imply that in the SK model for any fixed $\beta < 1/4$, $\|\Sigma\|_{OP} = O(1)$ with high probability, where $\Sigma = \mathbb{E}_\nu X X^T$ is the covariance matrix. This partially verifies Conjecture 11.5.1 of [22] that $\|\Sigma\|_{OP} = O(1)$ for any $\beta < 1$.

Diluted SK Model (d-Regular). A variety of spin glass models on sparse graphs have been studied in the literature; one well-known "diluted" version of the SK model has the interaction matrix J supported on a sparse Erdős-Reyni random graph — see [21]. Along similar lines, we can consider a dilution where J is supported on a random d-regular graph with $d \geq 3$. If we take a Rademacher disorder, i.e. $J_{ij} \sim Uni\{\pm\beta\}$ for i, j neighbors and otherwise $J_{ij} = 0$, then it follows from a version of Friedman's Theorem that $\|J\|_{OP} \leq \beta(2\sqrt{d-1} + \epsilon)$ a.a.s. for any $\epsilon > 0$ — see [4, 5, 19]. It follows from our results that we have the Poincaré inequality and fast mixing for all $\beta < \frac{1}{4\sqrt{d-1}}$, whereas the model is only in Dobrushin's uniqueness regime for $\beta = O\left(\frac{1}{d}\right)$ — note that up to constants the latter bound is tight for general Ising models on arbitrary d-regular graphs [9, 20].

A Inequivalence of Dirichlet Forms

The Dirichlet form $\mathcal{E}_\nu(\varphi, \varphi)$ can be viewed as the expected norm squared for an appropriate notion of gradient of φ. On the other hand, another natural notion of discrete gradient for functions on the hypercube is given by $(\nabla\varphi)_i(x) = \varphi(x_{\sim i}, x_i = 1) - \varphi(x_{\sim i}, x_i = -1)$ which is used in the result of Bauerschmidt and Bodineau [3], and the norm of this discrete gradient squared is the Dirichlet form of a different semigroup with a variable transition rate [17]. In the case of the SK model, the transition rate of the variable-rate chain is sometimes exponentially large in $\sqrt{n}$; in what follows, we give a simple example of a function φ witnessing that the Dirichlet forms similarly can differ in size by an exponentially large factor in $\sqrt{n}$.

To compare these two Dirichlet forms, we have the following estimates which follow immediately from (27):

$$\left(\min_{x \sim y} \frac{\nu(y)}{\nu(x) + \nu(y)}\right) \mathbb{E}_\nu \|\nabla\varphi\|^2 \lesssim \mathcal{E}_\nu(\varphi, \varphi) \lesssim \mathbb{E}_\nu \|\nabla\varphi\|^2.$$

In the context of the SK Model, the parenthesized term is of size $e^{-\Theta(\beta\sqrt{n})}$ and both estimates are tight up to constants. To see this for the lower bound, define $a \in \{\pm 1\}^n$ by $a_1 = -1$ and $a_j = \text{sgn}(J_{1j})$ otherwise; the significance of this choice is that in the SK model, it's exponentially unlikely to see $X_1 = a_1$ given $X_{\sim 1} = a_{\sim 1}$. Let λ be an atomic measure supported on a, so

$$\frac{d\lambda}{d\nu}(x) = \frac{\mathbf{1}[x = a]}{\nu(a)}.$$

If we define $\varphi = \sqrt{\frac{d\lambda}{d\nu}}$ then, see (28),

$$\mathcal{E}(\varphi, \varphi) = \sum_{x \sim y} \frac{\nu(x)\nu(y)}{\nu(x) + \nu(y)}(\varphi(x) - \varphi(y))^2 = \sum_{y: y \sim a} \frac{\nu(y)}{\nu(a) + \nu(y)} \leq n.$$

In comparison, for the discrete gradient $\nabla\varphi$ we have

$$\mathbb{E}|\nabla\varphi(X)|^2 \geq \frac{e^{c\beta\sqrt{n}}\nu(a)}{\nu(a)} = e^{\Theta(\beta\sqrt{n})}$$

where the lower bound follows by considering the a' which equals a but flipped on the first coordinate, and which (from the definition of the SK model) is $e^{\Theta(\beta\sqrt{n})}$ more likely under ν.

References

[1] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence in high-dimensional expanders: Modified log-sobolev inequalities for fractionally log-concave polynomials and the ising model. *arXiv preprint arXiv:2106.04105*, 2021.

[2] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118. Cambridge university press, 2010.

[3] Roland Bauerschmidt and Thierry Bodineau. A very simple proof of the LSI for high temperature spin systems. *Journal of Functional Analysis*, 276(8):2582–2588, 2019.

[4] Charles Bordenave. A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. *arXiv preprint arXiv:1502.04482*, 2015.

[5] Yash Deshpande, Andrea Montanari, Ryan O'Donnell, Tselil Schramm, and Subhabrata Sen. The threshold for SDP-refutation of random regular NAE-3SAT. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2305–2321. SIAM, 2019.

[6] Persi Diaconis and Laurent Saloff-Coste. Logarithmic Sobolev inequalities for finite markov chains. *The Annals of Applied Probability*, 6(3):695–750, 1996.

[7] Roland Lvovich Dobrushin. The description of a random field by means of conditional probabilities and conditions of its regularity. *Theor. Prob. Appl.*, 13:197–224, 1968.

[8] Ronen Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23(2):532–569, 2013.

[9] Andreas Galanis, Daniel Stefankovic, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Combinatorics, Probability and Computing*, 25:500–559, 2016.

[10] Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.

[11] Thomas P Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 39–46. IEEE, 2006.

[12] Ravi Kannan, László Lovász, and Miklós Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13(3-4):541–559, 1995.

[13] Ioannis Karatzas and Steven E Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1998.

[14] Bo'az Klartag. Needle decompositions in Riemannian geometry. *Mem. Amer. Math. Soc.*, 249(1180):v+77, 2017.

[15] Michel Ledoux. Concentration of measure and logarithmic Sobolev inequalities. In *Seminaire de probabilites XXXIII*, pages 120–216. Springer, 1999.

[16] David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

[17] Fabio Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics (Saint-Flour, 1997)*, volume 1717 of *Lecture Notes in Math.*, pages 93–191. Springer, Berlin, 1999.

[18] Katalin Marton. Logarithmic Sobolev inequalities in discrete product spaces. *Combinatorics, Probability and Computing*, 28(6):919–935, 2019.

[19] Sidhanth Mohanty, Ryan O'Donnell, and Pedro Paredes. Explicit near-Ramanujan graphs of every degree. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 510–523, 2020.

14

[20] Allan Sly and Nike Sun. The computational hardness of counting in two-spin models on d-regular graphs. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 361–369. IEEE, 2012.

[21] Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54. Springer Science & Business Media, 2010.

[22] Michel Talagrand. *Mean Field Models for Spin Glasses: Volume II: Advanced replica-symmetry and low temperature*. Springer, 2011.

[23] Ramon van Handel. Probability in high dimension. Technical report, Princeton Univ., NJ. https://web.math.princeton.edu/~rvan/APC550.pdf, 2016.

[24] Liming Wu. Poincaré and transportation inequalities for Gibbs measures under the dobrushin uniqueness condition. *The Annals of Probability*, 34(5):1960–1989, 2006.