# Subgaussian Vector Balancing by an Ising Model*

Heng Guo† Daniel A. Spielman‡ Xinyuan Zhang§

November 3, 2026

**Abstract**

We show that the guaranties of the Gram-Schmidt Walk Algorithm of Bansal, Dadush, Garg, and Lovett can be achieved by an Ising model. Given vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ of norm at most 1, we give a closed form for a centrally symmetric distribution on a vector $\boldsymbol{\sigma} \in \{\pm 1\}^n$ so that the signed sum $\sum_j \boldsymbol{\sigma}(j)\boldsymbol{v}_j$ is 1-subgaussian. We show that the Glauber dynamics for this model mix rapidly provided that the norms of the input vectors are bounded below 1 by an inverse-polynomial.

The proofs employ the theory of Lorentzian polynomials, and we use these techniques to show fast mixing bounds for other spin systems.

*This research was partially supported by ONR Award N00014-24-12611.

†University of Edinburgh. Email: hguo@inf.ed.ac.uk.

‡Yale University. Email: daniel.spielman@yale.edu

§Georgia Institute of Technology. Email: normalvra@gmail.com

# 1 Introduction

Dadush, Garg, Lovett, and Nikolov [DGLN19] proved that Banaszczyks Vector Balancing Theorem [Ban98] is equivalent to the following sampling problem: given vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ in the unit ball, sample a vector $\boldsymbol{\sigma} \in \{\pm 1\}^n$ so that the signed sum $\sum_j \boldsymbol{\sigma}(j)\boldsymbol{v}_j$ has expectation zero and is $C$-subgaussian, for some constant $C$. We call an algorithm that achieves this guarantee a *$C$-subgaussian vector balancing algorithm*. Bansal, Dadush, Garg and Lovett [BDGL19] developed a polynomial-time algorithm, called the Gram-Schmidt Walk, that solves this sampling problem and thereby provides a polynomial-time realization of Banaszczyk's Theorem. Harshaw, Sävje, Spielman and Zhang [HSSZ24] proved that the constant achieved by this algorithm is actually 1, which is the best possible.

We show that one can obtain a 1-subgaussian vector balancing algorithm from an Ising model that encodes a very simple distribution, and that the natural Glauber dynamics sample from this model in polynomial-time. If all the vectors have Euclidean norm at most 1, then the distribution samples $\boldsymbol{\sigma}$ with probability proportional to

$$w(\boldsymbol{\sigma}) \stackrel{\text{def}}{=} \prod_{1 \le i < j \le n} \left(1 - \boldsymbol{\sigma}(i)\boldsymbol{\sigma}(j)\left\langle \boldsymbol{v}_i, \boldsymbol{v}_j \right\rangle\right). \tag{1}$$

This is an Ising model with interaction matrix $\boldsymbol{J}$ given by $\boldsymbol{J}(i,j) = \frac{1}{2}\log\frac{1+\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle}{1-\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle} = \operatorname{atanh}(\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle)$ because $w(\boldsymbol{\sigma})$ is proportional to $\exp(-\boldsymbol{\sigma}^T\boldsymbol{J}\boldsymbol{\sigma}/2)$.

This distribution is well-defined so long as the input does not contain any three parallel unit vectors. If multiple input vectors have the form $\pm\boldsymbol{v}$ where $\boldsymbol{v}$ has norm 1, then we fix their sum to be zero if there are an even number of them, or to $\pm\boldsymbol{v}$ if their number is odd. The mixing time of the Glauber dynamics depends on the reciprocal of 1 minus the maximum norm of a vector. To efficiently sample when some vectors have norm 1 or very close to it, we scale down the vectors and use rejection sampling. See Section 4 for the details.

Our motivation for this paper is three-fold. First, we believe that 1-subgaussian vector balancing is sufficiently fundamental that we should know if it can be done with a simple and cleanly described distribution. To the best of our knowledge, ours is the first closed form $C$-subgaussian sampler for any $C$. Second, if one wants to vary the vector balancing guarantee, it should be much easier to modify the distribution directly than it would be to modify the Gram-Schmidt Walk. Third, we expect that experimenters will be more likely to use 1-subgaussian samplers to design randomized controlled trials, as suggested in [HSSZ24], if they have a better understanding of the distribution.

We now state our main results.

**Theorem 1.** *Assume that the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ all have Euclidean norm at most 1 and that no three are parallel unit vectors. Let $\boldsymbol{\sigma}$ be chosen with probability proportional to $w(\boldsymbol{\sigma})$. Let $\boldsymbol{s} = \sum_i \boldsymbol{\sigma}(i)\boldsymbol{v}_i$. Then, $\boldsymbol{s}$ is a 1-subgaussian random vector. That is, for all vectors $\boldsymbol{\theta}$,*

$$\mathbb{E}\exp(\langle \boldsymbol{\theta}, \boldsymbol{s}\rangle) \le \exp(\|\boldsymbol{\theta}\|^2/2).$$

Theorem 1 is proved as a special case of Theorem 9, which shows that the same holds under arbitrary external fields after re-centering $\boldsymbol{s}$.

We prove rapid mixing of the Glauber dynamics for the Ising distribution (1), when the norms of the input vectors are inverse-polynomially bounded away from 1. This also holds under external fields.

**Theorem 2.** *Let* $\alpha_0 \stackrel{\text{def}}{=} 1 - \max_{i\in[n]} \|\boldsymbol{v}_i\|^2$. *The spectral gap of the Glauber dynamics* $P$ *for the Ising distribution* (1) *satisfies*

$$\operatorname{gap}(P) \geq \frac{\alpha_0}{n}.$$

*Consequently, for every* $0 < \varepsilon < 1$, *its mixing time satisfies*

$$t_{\text{mix}}(\varepsilon) = O\left(\frac{n}{\alpha_0}\left(\log\frac{1}{\varepsilon} + n^2 \log \alpha_0^{-1}\right)\right).$$

This is not quite as fast as the Gram-Schmidt Walk [BDGL19], which can be implemented to run in time $O(n^2 d)$. We conjecture that our distribution satisfies a modified log-Sobolev inequality with constant $O(\alpha_0/n)$, which would get us close to this running time bound: it would imply mixing in time $O((n/\alpha_0)\log(n\log(1/\alpha_0)/\epsilon))$. Those who are willing to relax the subgaussian bound can obtain a $\sqrt{3}$-subgaussian guarantee in time $O(nd)$, although again without a closed form distribution [Ade26].

To sample from the distribution (1), there is a faster way than directly using the mixing time guarantee in Theorem 2, namely the checkpoint technique in [CGVY26, Section 5]. Introduce a temperature parameter, and (non-adaptively and deterministically) find an $\widetilde{O}(\sqrt{n})$-length cooling schedule with bounded $\chi^2$-divergence between two adjacent temperatures. Then, run the Glauber dynamics at each temperature, and once it is mixed, use the last state as a warm start for the next temperature. Overall we only need $\widetilde{O}(n^{1.5}/\alpha_0)$ Markov chain steps. We omit the details here and refer interested readers to read [CGVY26, Section 5], since this is a straightforward and routine application of that technique.

In Section 4 we give a few ways of modifying the distribution and/or the sampling procedure to handle vectors of norm up to 1.

Theorem 2 provides a new spectral criterion for the rapid mixing of Glauber dynamics on general Ising models. Consider the Ising distribution on $\{\pm 1\}^n$ specified by a symmetric interaction matrix $\boldsymbol{J} \in \mathbb{R}^{n\times n}$ with zero diagonal and an arbitrary external field $\boldsymbol{h} \in \mathbb{R}^n$:

$$\forall \boldsymbol{\sigma} \in \{\pm 1\}^n, \qquad \mu(\boldsymbol{\sigma}) \propto \exp\left(\frac{1}{2}\boldsymbol{\sigma}^T \boldsymbol{J}\boldsymbol{\sigma} + \boldsymbol{h}^T\boldsymbol{\sigma}\right). \tag{2}$$

**Theorem 3.** *Let* $\boldsymbol{J} \in \mathbb{R}^{n\times n}$ *be a symmetric matrix with zero diagonal, let* $\boldsymbol{h} \in \mathbb{R}^n$ *be an arbitrary external field, and let* $\boldsymbol{T} = \tanh(\boldsymbol{J}) \stackrel{\text{def}}{=} (\tanh(\boldsymbol{J}(i,j)))_{i,j\in[n]}$ *denote the entrywise hyperbolic-tangent matrix of* $\boldsymbol{J}$. *If* $\lambda_{\max}(\boldsymbol{T}) < 1$, *then the spectral gap of the Glauber dynamics* $P$ *for the Ising distribution* (2) *satisfies*

$$\operatorname{gap}(P) \geq \frac{1 - \lambda_{\max}(\boldsymbol{T})}{n}.$$

*Remark.* For general Ising models, previous work [EKZ22, AJK+22] established the rapid mixing of Glauber dynamics when the interaction matrix satisfies $0 \preceq \boldsymbol{J} \preceq (1-\epsilon)\boldsymbol{I}$ for some $\epsilon > 0$. The conditions of Theorem 3 cover this as a special case. To see this, let $\boldsymbol{T}$ have zero diagonal and off-diagonal entries $\boldsymbol{T}(i,j) = \tanh \boldsymbol{J}(i,j)$. Since diagonal entries do not affect the Ising distribution, it suffices by Theorem 3 to show that $\boldsymbol{H} \stackrel{\text{def}}{=} \boldsymbol{J} - \boldsymbol{T} \succeq 0$. This follows from the fact that $\boldsymbol{H}$ is symmetric, has non-negative diagonals, and is diagonally-dominant: for all $i \in [n]$,

$$\sum_{j\neq i} |\boldsymbol{H}(i,j)| = \sum_{j\neq i} |\boldsymbol{J}(i,j) - \tanh \boldsymbol{J}(i,j)| < \sum_{j\neq i} |\boldsymbol{J}(i,j)|^3 \leq \sum_{j\neq i} |\boldsymbol{J}(i,j)|^2 \leq (\boldsymbol{J}^2)_{i,i} \leq \boldsymbol{J}(i,i) = \boldsymbol{H}(i,i),$$

where we use $|x - \tanh x| \le |x|^3/3$, $|\boldsymbol{J}(i,j)| \le \|\boldsymbol{J}\|_{\mathrm{op}} < 1$, $\boldsymbol{J} \succeq 0$ and $\boldsymbol{J}^2 \preceq \boldsymbol{J}$.

We apply Theorem 3 to the Sherrington–Kirkpatrick model, which corresponds to the Ising distribution (2) with symmetric interaction matrix $\boldsymbol{J} \in \mathbb{R}^{n\times n}$ given by independent entries $\boldsymbol{J}(i,j) \sim \mathcal{N}(0, \beta^2/n)$. The high-probability bound $\lambda_{\max}(\boldsymbol{T}) = 2\beta + o_n(1)$ yields the following guarantee.

**Corollary 4.** *For any $0 \le \beta < 1/2$ and any external field $\boldsymbol{h} \in \mathbb{R}^n$, with probability $1 - o_n(1)$ over the Gaussian couplings $\boldsymbol{J}$, the Glauber dynamics $P$ for the Sherrington–Kirkpatrick model satisfies*

$$\mathrm{gap}(P) \ge \frac{1 - 2\beta - o_n(1)}{n}.$$

The proofs of Theorem 3 and Corollary 4 are given at the end of Section 7. Previously, the spectral radius criterion [EKZ22, AJK+22] established a spectral gap of order $\Omega(1/n)$ for $\beta < 1/4$, which was subsequently improved to $\beta \approx 0.295$ using the trickle-down theorem [AKV24]. Recently, remarkable progress has been achieved in sampling the Sherrington–Kirkpatrick model, including total-variation sampling and warm-start mixing of Glauber dynamics for the zero-field model at $\beta < 1/2$ [DLSS26a, DLSS26b] as well as stronger worst-start mixing guarantees [Wan26, BLO26]. These results were further extended across the entire high-temperature regime $\beta < 1$ [LSS26, HLY26, DHMR26, Ope26b].

This approach also yields spectral gap bounds for general spin systems. Specifically, we establish rapid mixing of Glauber dynamics for anti-ferromagnetic spin systems on complete graphs whose interaction matrices have strictly positive diagonal entries, and derive spectral gap bounds in terms of the interaction matrix and eigenvalues of the graph's adjacency matrix. For the anti-ferromagnetic Potts model, our results extend the parameter range and improve the spectral gap lower bound obtained in recent work of Chen and Wang [CW26]. These extensions are presented in Section 8.

## 2 Preliminaries

Let $[n] = \{1, 2, \ldots, n\}$. A random vector $X \in \mathbb{R}^d$ is $C$-subgaussian if for all vectors $\boldsymbol{\theta} \in \mathbb{R}^d$,

$$\mathbb{E} \exp\left(\langle \boldsymbol{\theta}, X - \mathbb{E}X\rangle\right) \le \exp\left(C^2 \|\boldsymbol{\theta}\|^2/2\right).$$

Let $P$ be a reversible Markov chain on a finite state space $\Omega$ with stationary distribution $\mu$, that is, $\mu(\boldsymbol{\sigma})P(\boldsymbol{\sigma},\boldsymbol{\tau}) = \mu(\boldsymbol{\tau})P(\boldsymbol{\tau},\boldsymbol{\sigma})$ for all $\boldsymbol{\sigma},\boldsymbol{\tau} \in \Omega$. The spectral gap of the Markov chain $P$ is defined as the difference between 1 and the second largest eigenvalue of $P$, and for $0 < \varepsilon < 1$, the mixing time $t_{\mathrm{mix}}(\varepsilon)$ is given by

$$t_{\mathrm{mix}}(\epsilon) \stackrel{\mathrm{def}}{=} \max_{\boldsymbol{\sigma}\in\Omega} \min\left\{t \in \mathbb{N} \mid \mathrm{D_{TV}}\left(P^t(\boldsymbol{\sigma},\cdot), \mu\right) < \epsilon\right\},$$

where $\mathrm{D_{TV}}(\mu,\nu)$ denotes the total variation distance between distributions $\mu$ and $\nu$.

Let $\mu$ be a distribution over a finite product space $\mathcal{Q} = \prod_{i=1}^n Q_i$. The Glauber dynamics for $\mu$ updates a feasible configuration $\boldsymbol{\sigma} \in \mathcal{Q}$ by selecting a coordinate $i \in [n]$ uniformly at random, and replaces $\boldsymbol{\sigma}_i$ by $c \in Q_i$ with probability proportional to $\mu(\boldsymbol{\sigma}^{i\leftarrow c})$. Here, $\boldsymbol{\sigma}^{i\leftarrow c}$ is the configuration obtained from $\boldsymbol{\sigma}$ by replacing its $i$-th coordinate with $c$ and leaving all other coordinates untouched. The Glauber dynamics is reversible with respect to $\mu$.

## 3 Motivating the distribution

We came to the distribution (1) by a semi-rigorous process. We enumerated a few properties that we thought a distribution solving this problem should have, enumerated a couple of distributions that had those properties, and then eliminated those that failed to produce 1-subgaussian distributions in small experiments. This was done with conventional computation before the release of useful LLMs.

If we only have two input vectors and both of them are unit vectors, then the distribution (1) minimizes the operator norm of the covariance matrix of the resulting distribution. As this seems desirable, we sought distributions that agreed with this one when given only two vectors.

When the input is comprised of many identical small vectors in one dimension, say $\boldsymbol{v}_i = \epsilon$ for all $i$, the induced distribution resembles a discrete Gaussian: as $n$ grows, the probability of obtaining the sum $k\epsilon$ becomes approximately proportional to $\exp(-(k\epsilon)^2/2)$, for $k$ equivalent to $n$ modulo 2.

Finally, we wanted to be sure that entirely orthogonal sets of vectors would not interact, and that orthogonal pairs of vectors would not interact directly.

## 4 Handling vectors of norm up to 1

The distribution given by (1) is well-defined so long as there is a $\boldsymbol{\sigma}$ for which $w(\boldsymbol{\sigma})$ is non-zero. And, the only way that $w(\boldsymbol{\sigma})$ can be uniformly zero is if the input contains at least three parallel unit vectors. We can remove such cases from consideration by deterministically setting the signs of some of those vectors. Given an even number of parallel unit vectors, we can set their signs so that their sum is zero, and thus does not contribute to the signed sum. Given an odd number that is larger than three, we can remove all but one by fixing the signs of all but one so that they sum to zero.

While this transformation ensures that the distribution is well-defined and that the 1-subgaussian bound holds, it does not resolve the problem that our bound on the mixing time of Glauber dynamics depends on the reciprocal of 1 minus the maximum norm of an input vector. We suggest two ways of addressing this problem.

A crude solution is to divide all the input vectors by $1 + \epsilon$ for some small $\epsilon$. The Glauber dynamics will mix in time depending on $1/\epsilon$, and the resulting distribution on the *original* vectors is guaranteed to be at most $(1 + \epsilon)$-subgaussian. This will be enough for most practical purposes.

A principled way to address the problem while preserving the 1-subgaussian guarantee begins by observing that division of vectors by $(1 + \epsilon)$ does not substantially change the distribution if no input vectors have inner product close to 1 or $-1$. And, we can remove pairs of vectors with such inner products by replacing them with their sum or difference.

Let $c = \sqrt{3} - 1$. If any pair of input vectors $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ have $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle > c$, then $\boldsymbol{v}_i - \boldsymbol{v}_j$ has norm less than $c$. The same holds for $\boldsymbol{v}_i + \boldsymbol{v}_j$ if $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle < -c$. After performing these replacements, we will be left with an input in which all pairs of vectors have inner products of absolute value at most $c$. After sampling from the distribution on this smaller input, we can replace the new vectors with with the sum or difference that produced them. This will ensure that the 1-subgaussian guarantee is satisfied.

To sample from the resulting distribution, we can now sample from the distribution obtained by dividing all the vectors by $(1 + \epsilon)$, where $\epsilon = 1/(4n^2)$, and using rejection sampling (see [Dev86, Chapter 2, Section 3]). Let $w_\epsilon(\boldsymbol{\sigma})$ the distribution obtained after this division. One can show that

for every $\boldsymbol{\sigma}$,

$$1/2w_\epsilon(\boldsymbol{\sigma}) \leq w(\boldsymbol{\sigma}) \leq 2w_\epsilon(\boldsymbol{\sigma}).$$

After sampling from the modified distribution, we return the sampled $\boldsymbol{\sigma}$ with probability $w(\boldsymbol{\sigma})/2w_\epsilon(\boldsymbol{\sigma})$. Otherwise, we repeat this procedure until we do return a $\boldsymbol{\sigma}$. As the probability of success is at least $1/4$, we expect to need at most 4 samples from the distribution proportional to $w_\epsilon$.

# 5 A discrete Alexandrov-Fenchel inequality and Lorentzian Polynomials

The key to our proof of Theorem 1 is a discrete analog of the Alexandrov-Fenchel inequality [Fen36, Ale37] proved by Lee, Oh, and Seo [LOS25] using the theory of Lorentzian Polynomials [BH20]. Lorentzian polynomials, under the name *strongly log-concave polynomials* [Gur09] have been used to analyze mixing rates of Markov chains [CGM21, ALOV24]. However, their use in our paper through Lee, Oh and Seo's discrete Alexandrov-Fenchel inequality appears novel.

We will say that a symmetric matrix is *Lorentzian* if its entries are non-negative and it has at most one positive eigenvalue[1]. Given a finite set of coordinates, $C$, a Lorentzian matrix $\boldsymbol{A} \in \mathbb{R}^{C\times C}$, and $k$ vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^C$, Lee, Oh and Seo examine the following polynomial whose variables are the entries of those vectors:

$$\mathcal{V}_{\boldsymbol{A}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \stackrel{\text{def}}{=} \sum_{c:[k]\to C} \prod_{i=1}^{k} \boldsymbol{x}_i(c(i)) \prod_{\{i,j\}\in\binom{[k]}{2}} \boldsymbol{A}(c(i), c(j)). \tag{3}$$

The following theorem may be derived from Theorem 3.1, Proposition 4.1, and the discussion following it through equation (4.1) of their paper [LOS25]:

**Theorem 5.** *Let $\boldsymbol{A} \in \mathbb{R}^{C\times C}$ be a Lorentzian matrix. Then, for all vectors $\boldsymbol{x}_1 \in \mathbb{R}^C$ and $\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k \in \mathbb{R}^C_{\geq 0}$, we have*

$$\mathcal{V}_{\boldsymbol{A}}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_k)^2 \geq \mathcal{V}_{\boldsymbol{A}}(\boldsymbol{x}_1, \boldsymbol{x}_1, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_k)\mathcal{V}_{\boldsymbol{A}}(\boldsymbol{x}_2, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_k). \tag{4}$$

To prove this, Lee, Oh, and Seo show that $\mathcal{V}_{\boldsymbol{A}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is the polarization of a Lorentzian polynomial and then apply results of Brändén and Huh. They point out that their Proposition 4.1 is in fact Proposition 4.5 of [BH20]. The first inequality of this type of which we are aware appears in the work of Van Lint [vL81, Lemma 2.5].

Two types of vectors will arise in our proofs: the vectors $\boldsymbol{v}_1, \ldots \boldsymbol{v}_n$ that we want to balance and the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ that appear in the definition of $\mathcal{V}_{\boldsymbol{A}}$. To avoid confusion, we will call the first set the *input vectors* and the second set the *selection vectors*.

We define the *Lorentz matrix* of vectors $\{\boldsymbol{v}_c\}_{c\in C}$ to be the matrix $\boldsymbol{A}$ with $\boldsymbol{A}(c, \widehat{c}) = 1 - \langle \boldsymbol{v}_c, \boldsymbol{v}_{\widehat{c}} \rangle$. The following lemma is the first step to using Theorem 5 to analyze the distribution (1).

**Lemma 6.** *Let $\boldsymbol{A}$ be the Lorentz matrix of vectors $\{\boldsymbol{v}_c\}_{c\in C}$. If all $\boldsymbol{v}_c$ have norm at most one, then $\boldsymbol{A}$ is a Lorentzian matrix.*

[1] If a non-negative symmetric matrix has a positive entry, then it has at least one positive eigenvalue. So, the only Lorentzian matrices with no positive eigenvalues are the all-zero matrices.

*Proof.* Let $\boldsymbol{J} \in \mathbb{R}^{C \times C}$ be the all-1 matrix and let $\boldsymbol{V} \in \mathbb{R}^{d \times C}$ be the matrix whose columns are the vectors $\boldsymbol{v}_c$. Then

$$\boldsymbol{A} = \boldsymbol{J} - \boldsymbol{V}^T \boldsymbol{V}.$$

As all the inner products of the vectors are at most 1, the matrix $\boldsymbol{A}$ is non-negative. As $\boldsymbol{V}^T \boldsymbol{V}$ is positive semidefinite, $\boldsymbol{z}^T \boldsymbol{A} \boldsymbol{z} \leq 0$ for all $\boldsymbol{z}$ orthogonal to $\mathbf{1}$. This implies that $\boldsymbol{A}$ has at least $|C| - 1$ non-positive eigenvalues, and so it has at most one positive eigenvalue. □

## 6 Subgaussian concentration

To relate the distribution of $\boldsymbol{s}$ to Theorem 5, we consider the Lorentz matrix of the input vectors and their opposites. To index these, let $\mathcal{C} = \{1^+, 1^-, 2^+, 2^-, \ldots, n^+, n^-\}$, and set $\boldsymbol{v}_{i^+} = \boldsymbol{v}_i$ and $\boldsymbol{v}_{i^-} = -\boldsymbol{v}_i$. Let $k = n$ and for $1 \leq i \leq n$ define the selection vector $\boldsymbol{x}_i$ so that

$$\boldsymbol{x}_i(c) = \begin{cases} 1 & \text{if } c = i^+ \\ 1 & \text{if } c = i^- \\ 0 & \text{otherwise.} \end{cases}$$

With these selection vectors, all nonzero terms in (3) must have $c(i) \in \{i^+, i^-\}$ for all $i$. Thus, the $c : [n] \to \mathcal{C}$ that can lead to nonzero terms are in bijection with $\boldsymbol{\sigma} \in \{\pm 1\}^n$, and

$$\mathcal{V}_{\boldsymbol{A}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^n} w(\sigma) \stackrel{\text{def}}{=} \mathcal{Z},$$

where $\mathcal{Z}$ is called the *partition function* of the distribution.

We will need to consider generalizations of the distribution given by $w$ that have been "tilted" by an "external field" $\boldsymbol{h} \in \mathbb{R}^n$. We define

$$w_{\boldsymbol{h}}(\boldsymbol{\sigma}) = e^{\langle \boldsymbol{h}, \boldsymbol{\sigma} \rangle} w(\boldsymbol{\sigma}), \tag{5}$$

and $\mathcal{Z}(\boldsymbol{h}) = \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^n} w_{\boldsymbol{h}}(\sigma)$. Thus, we sample $\boldsymbol{\sigma}$ with probability

$$\mu_{\boldsymbol{h}}(\boldsymbol{\sigma}) = \frac{w_{\boldsymbol{h}}(\boldsymbol{\sigma})}{\mathcal{Z}(\boldsymbol{h})}. \tag{6}$$

We write $\mathbb{E}_{\boldsymbol{h}}$ and $\mathbf{Var}_{\boldsymbol{h}}$ to indicate expectation and variance when $\boldsymbol{\sigma}$ is distributed according to this distribution. The assumption that there are no triple of parallel unit vectors ensures that $\mathcal{Z}(\boldsymbol{h})$ does not equal zero and thus these distributions are well-defined.

We can express $\mu_{\boldsymbol{h}}(\boldsymbol{\sigma})$ in terms of $\mathcal{V}_{\boldsymbol{A}}$ by changing the selection vectors. Upon changing $\boldsymbol{x}_i$ so that

$$\boldsymbol{x}_i(c) = \begin{cases} e^{\boldsymbol{h}(i)} & \text{if } c = i^+ \\ e^{-\boldsymbol{h}(i)} & \text{if } c = i^- \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

we obtain

$$\mathcal{V}_{\boldsymbol{A}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \mathcal{Z}(\boldsymbol{h}).$$

We now upper bound the covariance matrix of the signed sum $\boldsymbol{s}$ for every external field.

**Lemma 7.** *Assume that the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ all have Euclidean at most 1 and that no three are parallel unit vectors. Let $h \in \mathbb{R}^n$, $\boldsymbol{\sigma}$ be distributed according to $w_{\boldsymbol{h}}(\boldsymbol{\sigma})/\mathcal{Z}(\boldsymbol{h})$, and $\boldsymbol{s} = \sum_i \boldsymbol{\sigma}(i)\boldsymbol{v}_i$. Then for every $\boldsymbol{\theta} \in \mathbb{R}^d$,*

$$\mathbf{Var}\left(\langle \boldsymbol{\theta}, \boldsymbol{s}\rangle\right) \leq \|\boldsymbol{\theta}\|^2 .$$

We begin by introducing the notation we use in the proof. We will consider the functions $\mathcal{V}_{\boldsymbol{B}}$ where $\boldsymbol{B}$ is the Lorentz matrix of the input vectors, their negatives, the all-zero vector $\mathbf{0}$, and a vector $\epsilon\boldsymbol{\theta}$. We index vector $\mathbf{0}$ by coordinate 0 and vector $\epsilon\boldsymbol{\theta}$ by coordinate $-1$. We will set $k = n+2$, and consider two more selection vectors $\boldsymbol{x}_0$ and $\boldsymbol{x}_{-1}$ that we define to be the elementary unit vectors in directions 0 and $-1$. Concretely, we are instantiating (3) by setting $\widehat{\mathcal{C}} = \{-1, 0\} \cup \mathcal{C}$, $K = \{-1, 0, 1, \ldots, n\}$ and taking

$$\mathcal{V}_{\boldsymbol{B}}(\boldsymbol{y}_{-1}, \boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) \stackrel{\text{def}}{=} \sum_{c: K \to \widehat{\mathcal{C}}} \prod_{i \in K} \boldsymbol{x}_i(c(i)) \prod_{\{i,j\} \in \binom{K}{2}} \boldsymbol{B}(c(i), c(j)).$$

For selection vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, each of which will be one of $\boldsymbol{x}_0$ or $\boldsymbol{x}_{-1}$, we define

$$\mathcal{Z}(\boldsymbol{a}, \boldsymbol{b}) \stackrel{\text{def}}{=} \mathcal{V}_{\boldsymbol{B}}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n),$$

where we keep $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ as they were defined in (7).

*Proof of Lemma 7.* Let $\epsilon > 0$ satisfy $\epsilon \|\boldsymbol{\theta}\| < 1$. We will eventually let $\epsilon$ go to zero.

Theorem 5 tells us that

$$\mathcal{Z}(\boldsymbol{x}_0, \boldsymbol{x}_{-1})^2 \geq \mathcal{Z}(\boldsymbol{x}_0, \boldsymbol{x}_0)\mathcal{Z}(\boldsymbol{x}_{-1}, \boldsymbol{x}_{-1}). \tag{8}$$

We now expand the terms that appear in this inequality. As $\boldsymbol{x}_0$ pins a coordinate to 0 and that coordinate indexes the all-zero vector,

$$\mathcal{Z}(\boldsymbol{x}_0, \boldsymbol{x}_0) = \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^n} w_{\boldsymbol{h}}(\boldsymbol{\sigma})(1 - \langle \mathbf{0}, \mathbf{0}\rangle) \prod_{i=1}^n (1 - \langle \mathbf{0}, \boldsymbol{\sigma}(i)\boldsymbol{v}_i\rangle) \prod_{i=1}^n (1 - \langle \mathbf{0}, \boldsymbol{\sigma}(i)\boldsymbol{v}_i\rangle) = \mathcal{Z}(\boldsymbol{h}).$$

Similarly

$$\begin{aligned}
\mathcal{Z}(\boldsymbol{x}_0, \boldsymbol{x}_{-1}) &= \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^n} w_{\boldsymbol{h}}(\boldsymbol{\sigma})(1 - \langle \mathbf{0}, \epsilon\boldsymbol{\theta}\rangle) \prod_{i=1}^n (1 - \langle \mathbf{0}, \boldsymbol{\sigma}(i)\boldsymbol{v}_i\rangle) \prod_{i=1}^n (1 - \langle \epsilon\boldsymbol{\theta}, \boldsymbol{\sigma}(i)\boldsymbol{v}_i\rangle) \\
&= \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^n} w_{\boldsymbol{h}}(\boldsymbol{\sigma}) \prod_{i=1}^n (1 - \epsilon \langle \boldsymbol{\theta}, \boldsymbol{\sigma}(i)\boldsymbol{v}_i\rangle), \qquad \text{and} \\
\mathcal{Z}(\boldsymbol{x}_{-1}, \boldsymbol{x}_{-1}) &= \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^n} w_{\boldsymbol{h}}(\boldsymbol{\sigma})(1 - \langle \epsilon\boldsymbol{\theta}, \epsilon\boldsymbol{\theta}\rangle) \prod_{i=1}^n (1 - \langle \epsilon\boldsymbol{\theta}, \boldsymbol{\sigma}(i)\boldsymbol{v}_i\rangle) \prod_{i=1}^n (1 - \langle \epsilon\boldsymbol{\theta}, \boldsymbol{\sigma}(i)\boldsymbol{v}_i\rangle) \\
&= (1 - \epsilon^2 \langle \boldsymbol{\theta}, \boldsymbol{\theta}\rangle) \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^n} w_{\boldsymbol{h}}(\boldsymbol{\sigma}) \left( \prod_{i=1}^n (1 - \epsilon \langle \boldsymbol{\theta}, \boldsymbol{\sigma}(i)\boldsymbol{v}_i\rangle) \right)^2 .
\end{aligned}$$

Define

$$\boldsymbol{\Phi}(\boldsymbol{\sigma}) = \prod_{i=1}^n (1 - \epsilon \langle \boldsymbol{\theta}, \boldsymbol{\sigma}(i)\boldsymbol{v}_i\rangle),$$

so when we choose $\boldsymbol{\sigma}$ with probability $w_{\boldsymbol{h}}(\boldsymbol{\sigma})/\mathcal{Z}(\boldsymbol{h})$, we can write

$$\begin{aligned}\mathcal{Z}(\boldsymbol{x}_0, \boldsymbol{x}_{-1}) &= \mathcal{Z}(\boldsymbol{h})\mathbb{E}_{\boldsymbol{h}}\boldsymbol{\Phi}(\boldsymbol{\sigma}), \qquad \text{and} \\ \mathcal{Z}(\boldsymbol{x}_{-1}, \boldsymbol{x}_{-1}) &= (1 - \epsilon^2 \langle \boldsymbol{\theta}, \boldsymbol{\theta} \rangle)\mathcal{Z}(\boldsymbol{h})\mathbb{E}_{\boldsymbol{h}}\boldsymbol{\Phi}(\boldsymbol{\sigma})^2.\end{aligned}$$

Substituting these identities into (8) and canceling $\mathcal{Z}(\boldsymbol{h})^2$ from both sides yields

$$\begin{aligned}\left(\mathbb{E}_{\boldsymbol{h}}\boldsymbol{\Phi}(\boldsymbol{\sigma})\right)^2 &\geq (1 - \epsilon^2 \langle \boldsymbol{\theta}, \boldsymbol{\theta} \rangle)\mathbb{E}_{\boldsymbol{h}}\boldsymbol{\Phi}(\boldsymbol{\sigma})^2 \qquad \Longleftrightarrow \\ \epsilon^2 \|\boldsymbol{\theta}\|^2 \mathbb{E}_{\boldsymbol{h}}\boldsymbol{\Phi}(\boldsymbol{\sigma})^2 &\geq \mathbb{E}_{\boldsymbol{h}}\boldsymbol{\Phi}(\boldsymbol{\sigma})^2 - \left(\mathbb{E}_{\boldsymbol{h}}\boldsymbol{\Phi}(\boldsymbol{\sigma})\right)^2 \\ &= \mathbf{Var}_{\boldsymbol{h}}\left(\boldsymbol{\Phi}(\boldsymbol{\sigma})\right),\end{aligned}$$

which implies

$$\|\boldsymbol{\theta}\|^2 \mathbb{E}_{\boldsymbol{h}}\boldsymbol{\Phi}(\boldsymbol{\sigma})^2 \geq \frac{\mathbf{Var}_{\boldsymbol{h}}\left(\boldsymbol{\Phi}(\boldsymbol{\sigma})\right)}{\epsilon^2}.$$

The first-order approximation of $\boldsymbol{\Phi}(\boldsymbol{\sigma})$ as $\epsilon$ goes to zero is

$$\boldsymbol{\Phi}(\boldsymbol{\sigma}) \approx 1 - \epsilon \sum_i \langle \boldsymbol{\theta}, \boldsymbol{\sigma}(i)\boldsymbol{v}_i \rangle = 1 - \epsilon \langle \boldsymbol{\theta}, \boldsymbol{s} \rangle .$$

So, $\mathbb{E}_{\boldsymbol{h}}\boldsymbol{\Phi}(\boldsymbol{\sigma})^2$ approaches 1 as $\epsilon$ approaches 0 and

$$\frac{\mathbf{Var}_{\boldsymbol{h}}\left(\boldsymbol{\Phi}(\boldsymbol{\sigma})\right)}{\epsilon^2} \to \frac{\mathbf{Var}_{\boldsymbol{h}}\left(\epsilon \langle \boldsymbol{\theta}, \boldsymbol{\sigma} \rangle\right)}{\epsilon^2} = \mathbf{Var}_{\boldsymbol{h}}\left(\langle \boldsymbol{\theta}, \boldsymbol{s} \rangle\right),$$

because variance is unchanged by addition of constants.

We conclude that

$$\|\boldsymbol{\theta}\|^2 \geq \mathbf{Var}_{\boldsymbol{h}}\left(\langle \boldsymbol{\theta}, \boldsymbol{s} \rangle\right). \qquad \square$$

### 6.1 1-Subgaussian Concentration

In our proof that $\boldsymbol{s}$ is 1-subgaussian, we will need to integrate the result of Lemma 7 over a path of external fields. Our proof will make use of the logarithm of the partition function, which we define to be the function

$$F(\boldsymbol{h}) \stackrel{\text{def}}{=} \ln \mathcal{Z}(\boldsymbol{h}).$$

The following facts about $F(\boldsymbol{h})$ are well known. See, for example, [CLV23, Lemma 12].

**Claim 8.** *The gradient and Hessian of $F$ are given by*

$$\nabla_{\boldsymbol{h}} F(\boldsymbol{h}) = \mathbb{E}_{\boldsymbol{h}}\boldsymbol{\sigma} \qquad \textit{and} \qquad \nabla^2_{\boldsymbol{h}} F(\boldsymbol{h}) = \mathbf{Cov}_{\boldsymbol{h}}\left(\boldsymbol{\sigma}\right),$$

*where*

$$\mathbf{Cov}_{\boldsymbol{h}}\left(\boldsymbol{\sigma}\right) = \mathbb{E}_{\boldsymbol{h}}\boldsymbol{\sigma}\boldsymbol{\sigma}^T - \left(\mathbb{E}_{\boldsymbol{h}}\boldsymbol{\sigma}\right)\left(\mathbb{E}_{\boldsymbol{h}}\boldsymbol{\sigma}\right)^T$$

*is the covariance matrix of $\boldsymbol{\sigma}$.*

*Proof.* To compute the gradient, consider the $i$th coordinate.

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{h}(i)} F(\boldsymbol{h}) &= \frac{\partial}{\partial \boldsymbol{h}(i)} \ln Z(\boldsymbol{h}) \\
&= \frac{\frac{\partial}{\partial \boldsymbol{h}(i)} Z(\boldsymbol{h})}{Z(\boldsymbol{h})} \\
&= \frac{\sum_{\boldsymbol{\sigma}} \frac{\partial}{\partial \boldsymbol{h}(i)} w_{\boldsymbol{h}}(\boldsymbol{\sigma})}{Z(\boldsymbol{h})} \\
&= \frac{\sum_{\boldsymbol{\sigma}} \frac{\partial}{\partial \boldsymbol{h}(i)} \exp(\langle \boldsymbol{h}, \boldsymbol{\sigma} \rangle) w(\boldsymbol{\sigma})}{Z(\boldsymbol{h})} \\
&= \frac{\sum_{\boldsymbol{\sigma}} \boldsymbol{\sigma}(i) \exp(\langle \boldsymbol{h}, \boldsymbol{\sigma} \rangle) w(\boldsymbol{\sigma})}{Z(\boldsymbol{h})} \\
&= \mathbb{E}_{\boldsymbol{h}} \boldsymbol{\sigma}(i).
\end{aligned}$$

We can similarly show that

$$\frac{\partial^2}{\partial \boldsymbol{h}(i) \partial \boldsymbol{h}(j)} F(\boldsymbol{h}) = \mathbb{E}_{\boldsymbol{h}} \boldsymbol{\sigma}(i) \boldsymbol{\sigma}(j) - \mathbb{E}_{\boldsymbol{h}} \boldsymbol{\sigma}(i) \mathbb{E}_{\boldsymbol{h}} \boldsymbol{\sigma}(j).$$

□

We now prove a strengthening of Theorem 1 that allows for an external field $\boldsymbol{h}$. Theorem 1 is the special case with $\boldsymbol{h} = \mathbf{0}$. For that $\boldsymbol{h}$, $\mathbb{E}_{\mathbf{0}} \boldsymbol{\sigma} = \mathbf{0}$, which simplifies the proof slightly.

**Theorem 9.** *Assume that the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ all have Euclidean at most 1 and that no three are parallel unit vectors. Let $\boldsymbol{h} \in \mathbb{R}^n$ and let $\boldsymbol{\sigma}$ be chosen with probability proportional to $w_{\boldsymbol{h}}(\boldsymbol{\sigma})$. Let $\boldsymbol{s} = \sum_i \boldsymbol{\sigma}(i) \boldsymbol{v}_i$. Then, $\boldsymbol{s}$ is a 1-subgaussian random vector. That is, for all vectors $\boldsymbol{\theta}$,*

$$\mathbb{E}_{\boldsymbol{h}} \exp(\langle \boldsymbol{\theta}, \boldsymbol{s} - \mathbb{E}_{\boldsymbol{h}} \boldsymbol{s} \rangle) \leq \exp(\|\boldsymbol{\theta}\|^2 / 2).$$

*Proof.* Let $\boldsymbol{\theta}$ be a vector in $\mathbb{R}^d$. We will show

$$\ln \left( \mathbb{E}_{\boldsymbol{h}} \exp(\langle \boldsymbol{\theta}, \boldsymbol{s} - \mathbb{E}_{\boldsymbol{h}} \boldsymbol{s} \rangle) \right) \leq \|\boldsymbol{\theta}\|^2 / 2.$$

To turn this into an expression in $\boldsymbol{\sigma}$, let $\boldsymbol{V}$ be the matrix whose columns are $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ so that $\boldsymbol{s} = \boldsymbol{V} \boldsymbol{\sigma}$. Define

$$\boldsymbol{\phi} = \boldsymbol{V}^T \boldsymbol{\theta}$$

so that

$$\langle \boldsymbol{\theta}, \boldsymbol{s} \rangle = \langle \boldsymbol{\phi}, \boldsymbol{\sigma} \rangle .$$

The connection to the partition functions is

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{h}} \exp(\langle \boldsymbol{\theta}, \boldsymbol{s} \rangle) &= \mathbb{E}_{\boldsymbol{h}} \exp(\langle \boldsymbol{\phi}, \boldsymbol{\sigma} \rangle) \\
&= \frac{\mathcal{Z}(\boldsymbol{\phi} + \boldsymbol{h})}{\mathcal{Z}(\boldsymbol{h})},
\end{aligned}$$

and so

$$\ln \left( \mathbb{E}_{\boldsymbol{h}} \exp(\langle \boldsymbol{\theta}, \boldsymbol{s} - \mathbb{E}_{\boldsymbol{h}} \boldsymbol{s} \rangle) \right) = F(\boldsymbol{\phi} + \boldsymbol{h}) - F(\boldsymbol{h}) - \langle \boldsymbol{\phi}, \mathbb{E}_{\boldsymbol{h}} \boldsymbol{\sigma} \rangle .$$

This reduces our problem to that of showing

$$F(\boldsymbol{\phi}+\boldsymbol{h}) - F(\boldsymbol{h}) \leq \|\boldsymbol{\theta}\|^2/2 + \langle \boldsymbol{\theta}, \mathbb{E}_{\boldsymbol{h}}\boldsymbol{s}\rangle\,.$$

We will do this by integrating on a path from $\boldsymbol{h}$ to $\boldsymbol{\phi}+\boldsymbol{h}$. For a real variable $t \in [0,1]$, define

$$f(t) \stackrel{\text{def}}{=} F(t\boldsymbol{\phi}+\boldsymbol{h}).$$

As $f(t)$ is twice continuously differentiable,

$$f(1) - f(0) = \int_{t=0}^{1} f'(t)dt = f'(0) + \int_{t=0}^{1}(1-t)f''(t)dt.$$

By the chain rule, $f'(t) = \langle \boldsymbol{\phi}, \nabla F(t\boldsymbol{\phi}+\boldsymbol{h})\rangle$. So Claim 8 implies $f'(0) = \langle \boldsymbol{\phi}, \mathbb{E}_{\boldsymbol{h}}\boldsymbol{\sigma}\rangle = \langle \boldsymbol{\theta}, \mathbb{E}_{\boldsymbol{h}}\boldsymbol{s}\rangle$. The chain rule also gives

$$\begin{aligned}
f''(t) &= \boldsymbol{\phi}^T \nabla^2 F(t\boldsymbol{\phi}+\boldsymbol{h})\boldsymbol{\phi} \\
&= \boldsymbol{\phi}^T \mathbf{Cov}_{t\boldsymbol{\phi}+\boldsymbol{h}}(\boldsymbol{\sigma})\,\boldsymbol{\phi} \qquad \text{(by Claim 8)} \\
&= \boldsymbol{\theta}^T \boldsymbol{V}\mathbf{Cov}_{t\boldsymbol{\phi}+\boldsymbol{h}}(\boldsymbol{\sigma})\,\boldsymbol{V}^T\boldsymbol{\theta} \\
&= \boldsymbol{\theta}^T \mathbf{Cov}_{t\boldsymbol{\phi}+\boldsymbol{h}}(\boldsymbol{V}\boldsymbol{\sigma})\,\boldsymbol{\theta} \\
&= \boldsymbol{\theta}^T \mathbf{Cov}_{t\boldsymbol{\phi}+\boldsymbol{h}}(\boldsymbol{s})\,\boldsymbol{\theta} \\
&\leq \|\boldsymbol{\theta}\|^2\,,
\end{aligned}$$

by Lemma 7. Plugging this upper bound back in to the expression for $f(1) - f(0)$ proves the theorem:

$$f(1) - f(0) = f'(0) + \int_{t=0}^{1}(1-t)f''(t)dt \leq f'(0) + \int_{t=0}^{1}(1-t)\|\boldsymbol{\theta}\|^2 dt = \langle \boldsymbol{\theta}, \mathbb{E}_{\boldsymbol{h}}\boldsymbol{s}\rangle + \|\boldsymbol{\theta}\|^2/2. \quad \square$$

## 7 Glauber dynamics

Next we analyse the standard algorithm to sample from the Gibbs distribution $\mu_h$ induced by (5), namely Glauber dynamics. The key result is a Poincaré inequality, which is also derived from Theorem 5.

For $i \in [n]$, let $P_i$ be the matrix that, when acting on probability distributions represented by row vectors, resamples the $i$th coordinate. Namely, $P_i(\boldsymbol{\sigma})$ keeps $\boldsymbol{\sigma}_{-i}$, the partial configuration other than $i$, and replaces the $i$th coordinate by drawing from the distribution $\mu_h$ conditioned on $\boldsymbol{\sigma}_{-i}$. When applied to column vectors $f : \{\pm 1\}^n \to \mathbb{R}$, $P_i$ replaces $f(\boldsymbol{\sigma})$ with a weighted average of $f(\boldsymbol{\sigma})$ and $f(\tau)$ where $\tau$ is the $\{\pm 1\}^n$ vector that differs from $\boldsymbol{\sigma}$ in coordinate $i$: $(P_i f)(\boldsymbol{\sigma}) = (w_{\boldsymbol{h}}(\boldsymbol{\sigma})f(\boldsymbol{\sigma}) + w_{\boldsymbol{h}}(\tau)f(\tau))/(w_{\boldsymbol{h}}(\boldsymbol{\sigma}) + w_{\boldsymbol{h}}(\tau))$. Let

$$P \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} P_i.$$

Also, define

$$D_i \stackrel{\text{def}}{=} I - P_i.$$

Let $\alpha_i = 1 - \|\boldsymbol{v}_i\|^2$ and $\alpha_0 = \min_i \alpha_i$. We prove the following approximate tensorization of variance for distribution $\mu_h$, which is equivalent to the Poincaré inequality for its Glauber dynamics. We refer readers to [CLV21] for a detailed introduction.

**Lemma 10.** *For every $f : \{\pm 1\}^n \to \mathbb{R}$,*

$$\mathbf{Var}_{\boldsymbol{h}}\left(f(\boldsymbol{\sigma})\right) \leq \frac{1}{\alpha_0} \sum_{i=1}^{n} \mathbb{E}_h\left[((D_i f)(\boldsymbol{\sigma}))^2\right]. \tag{9}$$

Our proof rests on a consequence of Theorem 5 which we state, prove, and then massage to prove Lemma 10. For $\boldsymbol{z}$ and $\boldsymbol{\sigma}$ in $\mathbb{R}^n$, we define the function

$$\boldsymbol{\Phi}(\boldsymbol{z}, \boldsymbol{\sigma}) \stackrel{\text{def}}{=} \prod_{i=1}^{n} (1 - \boldsymbol{z}(i)\boldsymbol{\sigma}(i)).$$

**Lemma 11.** *Assume that $\|\boldsymbol{v}_i\|^2 \leq 1 - \alpha_0$ for each $i$. Let $S$ be a finite set of vectors in $\mathbb{R}^n$ and for each $\boldsymbol{z} \in S$ let $c_{\boldsymbol{z}}$ be a real number. If all vectors in $S$ have norm at most $\sqrt{\alpha_0}$, then*

$$\mathbf{Var}_{\boldsymbol{h}}\left(\sum_{\boldsymbol{z}\in S} c_{\boldsymbol{z}} \boldsymbol{\Phi}(\boldsymbol{z}, \boldsymbol{\sigma})\right) \leq \frac{1}{\alpha_0} \mathbb{E}_{\boldsymbol{h}} \left\| \sum_{\boldsymbol{z}\in S} c_{\boldsymbol{z}} \boldsymbol{z} \boldsymbol{\Phi}(\boldsymbol{z}, \boldsymbol{\sigma}) \right\|^2, \tag{10}$$

*where the expectation and variance are taken over $\boldsymbol{\sigma}$ distributed according to $w_{\boldsymbol{h}}$.*

We prove Lemma 11 by adding coordinates to the vectors $\boldsymbol{v}_i$ and $\boldsymbol{z}$ in a way we describe below, and then applying the following lemma to reason about them. The difference between Lemma 11 and Lemma 12 is that Lemma 11 considers functions defined by inner products with $\boldsymbol{\sigma}$, while Lemma 12 considers inner products with different vectors.

For vectors $\boldsymbol{u}$ and $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ of the same dimension, define the function

$$\boldsymbol{\Phi}^{w}_{\boldsymbol{u}}(\boldsymbol{\sigma}) \stackrel{\text{def}}{=} \prod_{i=1}^{n} (1 - \boldsymbol{\sigma}(i) \langle \boldsymbol{u}, \boldsymbol{w}_i \rangle).$$

**Lemma 12.** *Let $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ be vectors of norm at most 1 such that for all $j \neq k$*

$$\langle \boldsymbol{w}_j, \boldsymbol{w}_k \rangle = \langle \boldsymbol{v}_j, \boldsymbol{v}_k \rangle.$$

*Let $U$ be a finite set of vectors. For each $\boldsymbol{u} \in U$, let $c_{\boldsymbol{u}}$ be a real number. If all the vectors in $U$ have norm at most 1, then*

$$\mathbf{Var}_{\boldsymbol{h}}\left(\sum_{\boldsymbol{u}\in U} c_{\boldsymbol{u}} \boldsymbol{\Phi}^{w}_{\boldsymbol{u}}(\boldsymbol{\sigma})\right) \leq \mathbb{E}_{\boldsymbol{h}} \left\| \sum_{\boldsymbol{u}\in U} c_{\boldsymbol{u}} \boldsymbol{u} \boldsymbol{\Phi}^{w}_{\boldsymbol{u}}(\boldsymbol{\sigma}) \right\|^2, \tag{11}$$

*where the expectation and variance are taken over $\boldsymbol{\sigma}$ distributed according to $w_{\boldsymbol{h}}$.*

*Proof.* Let $\boldsymbol{B}$ be the Lorentz matrix of the vectors $\{\pm \boldsymbol{w}_i\}_{i\in[n]}$, $\mathbf{0}$, and $\{\boldsymbol{u}\}_{\boldsymbol{u}\in U}$. The vectors $\pm \boldsymbol{w}_i$ are indexed by coordinates $i^+$ and $i^-$ and $\mathbf{0}$ is indexed by coordinate 0. For each $\boldsymbol{u} \in U$, we index

the vector $\boldsymbol{u}$ by a coordinate we call $\boldsymbol{u}$[2]. We also use the selection vectors $\boldsymbol{x}_i$ defined in (7), and let $\boldsymbol{x}_0$ be the elementary unit vector in direction 0. And, we introduce just one more selection vector $\boldsymbol{x}_{-1}$ defined by

$$\boldsymbol{x}_{-1}(\boldsymbol{u}) = c_{\boldsymbol{u}}, \quad \text{for } u \in U.$$

For selection vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, define

$$\mathcal{Y}(\boldsymbol{a}, \boldsymbol{b}) \stackrel{\text{def}}{=} \mathcal{V}_{\boldsymbol{B}}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n).$$

As in the proof of Lemma 7, we apply Theorem 5 to show

$$\mathcal{Y}(\boldsymbol{x}_0, \boldsymbol{x}_{-1})^2 \geq \mathcal{Y}(\boldsymbol{x}_0, \boldsymbol{x}_0)\mathcal{Y}(\boldsymbol{x}_{-1}, \boldsymbol{x}_{-1}). \tag{12}$$

The assumption $\langle \boldsymbol{w}_j, \boldsymbol{w}_k \rangle = \langle \boldsymbol{v}_j, \boldsymbol{v}_k \rangle$ guarantees that

$$\begin{aligned}
\mathcal{Y}(\boldsymbol{x}_0, \boldsymbol{x}_0) &= \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^n} \exp(\langle \boldsymbol{h}, \boldsymbol{\sigma} \rangle) \prod_{\{i,j\} \in \binom{[n]}{2}} (1 - \boldsymbol{\sigma}(i)\boldsymbol{\sigma}(j) \langle \boldsymbol{w}_i, \boldsymbol{w}_j \rangle) \\
&= \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^n} \exp(\langle \boldsymbol{h}, \boldsymbol{\sigma} \rangle) \prod_{\{i,j\} \in \binom{[n]}{2}} (1 - \boldsymbol{\sigma}(i)\boldsymbol{\sigma}(j) \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle) \\
&= \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^n} w_{\boldsymbol{h}}(\boldsymbol{\sigma}) \\
&= \mathcal{Z}(\boldsymbol{h}).
\end{aligned}$$

We may similarly compute

$$\begin{aligned}
\mathcal{Y}(\boldsymbol{x}_0, \boldsymbol{x}_{-1}) &= \sum_{\boldsymbol{u} \in U} c_{\boldsymbol{u}} \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^n} w_{\boldsymbol{h}}(\boldsymbol{\sigma}) \prod_{i=1}^{n} (1 - \langle \boldsymbol{u}, \boldsymbol{\sigma}(i) \boldsymbol{w}_i \rangle) \\
&= \mathcal{Z}(\boldsymbol{h}) \mathbb{E} \sum_{\boldsymbol{u} \in U} c_{\boldsymbol{u}} \boldsymbol{\Phi}^{\boldsymbol{w}}_{\boldsymbol{u}}(\boldsymbol{\sigma}) \\
\mathcal{Y}(\boldsymbol{x}_{-1}, \boldsymbol{x}_{-1}) &= \mathcal{Z}(\boldsymbol{h}) \mathbb{E} \sum_{\boldsymbol{u}, \widehat{\boldsymbol{u}} \in U} c_{\boldsymbol{u}} c_{\widehat{\boldsymbol{u}}} (1 - \langle \boldsymbol{u}, \widehat{\boldsymbol{u}} \rangle) \boldsymbol{\Phi}^{\boldsymbol{w}}_{\boldsymbol{u}}(\boldsymbol{\sigma}) \boldsymbol{\Phi}^{\boldsymbol{w}}_{\widehat{\boldsymbol{u}}}(\boldsymbol{\sigma}) \\
&= \mathcal{Z}(\boldsymbol{h}) \mathbb{E} \left( \sum_{\boldsymbol{u} \in U} c_{\boldsymbol{u}} \boldsymbol{\Phi}^{\boldsymbol{w}}_{\boldsymbol{u}}(\boldsymbol{\sigma}) \right)^2 - \mathcal{Z}(\boldsymbol{h}) \mathbb{E} \sum_{\boldsymbol{u}, \widehat{\boldsymbol{u}} \in U} c_{\boldsymbol{u}} c_{\widehat{\boldsymbol{u}}} \langle \boldsymbol{u}, \widehat{\boldsymbol{u}} \rangle \boldsymbol{\Phi}^{\boldsymbol{w}}_{\boldsymbol{u}}(\boldsymbol{\sigma}) \boldsymbol{\Phi}^{\boldsymbol{w}}_{\widehat{\boldsymbol{u}}}(\boldsymbol{\sigma}) \\
&= \mathcal{Z}(\boldsymbol{h}) \mathbb{E} \left( \sum_{\boldsymbol{u} \in U} c_{\boldsymbol{u}} \boldsymbol{\Phi}^{\boldsymbol{w}}_{\boldsymbol{u}}(\boldsymbol{\sigma}) \right)^2 - \mathcal{Z}(\boldsymbol{h}) \mathbb{E} \left\| \sum_{\boldsymbol{u} \in U} c_{\boldsymbol{u}} \boldsymbol{u} \boldsymbol{\Phi}^{\boldsymbol{w}}_{\boldsymbol{u}}(\boldsymbol{\sigma}) \right\|^2.
\end{aligned}$$

Plugging these terms into (12), canceling $\mathcal{Z}(\boldsymbol{h})^2$, and rearranging terms gives

$$\begin{aligned}
\mathbb{E} \left\| \sum_{\boldsymbol{u} \in U} c_{\boldsymbol{u}} \boldsymbol{u} \boldsymbol{\Phi}_{\boldsymbol{u}}(\boldsymbol{\sigma}) \right\|^2 &\geq \mathbb{E} \left( \sum_{\boldsymbol{u} \in U} c_{\boldsymbol{u}} \boldsymbol{\Phi}^{\boldsymbol{w}}_{\boldsymbol{u}}(\boldsymbol{\sigma}) \right)^2 - \left( \mathbb{E} \sum_{\boldsymbol{u} \in U} c_{\boldsymbol{u}} \boldsymbol{\Phi}^{\boldsymbol{w}}_{\boldsymbol{u}}(\boldsymbol{\sigma}) \right)^2 \\
&= \mathbf{Var}_h \left( \sum_{\boldsymbol{u} \in U} c_{\boldsymbol{u}} \boldsymbol{\Phi}^{\boldsymbol{w}}_{\boldsymbol{u}}(\boldsymbol{\sigma}) \right).
\end{aligned}$$

□

[2]This overloading of the same symbol for a vector and a coordinate is convenient, and we hope it does not create confusion.

*Proof of Lemma 11.* We construct $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ from $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ by setting

$$\boldsymbol{w}_i = \boldsymbol{v}_i \oplus \sqrt{\alpha_0}\boldsymbol{e}_i,$$

where $\boldsymbol{e}_i$ is the elementary unit vector in $\mathbb{R}^n$ in direction $i$, and $\oplus$ indicates concatenation of vectors. For $i \neq j$, $\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = 0$ and so $\langle \boldsymbol{w}_i, \boldsymbol{w}_j \rangle = \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle$. By our assumption that $\|\boldsymbol{v}_i\|^2 \leq 1 - \alpha_0$, we can conclude that $\|\boldsymbol{w}_i\|^2 = \|\boldsymbol{v}_i\|^2 + \alpha_0 \leq 1$. For each $\boldsymbol{z} \in S$, define

$$\widehat{\boldsymbol{z}} \stackrel{\text{def}}{=} \mathbf{0}_d \oplus \boldsymbol{z}/\sqrt{\alpha_0},$$

and

$$U = \{\widehat{\boldsymbol{z}} : \boldsymbol{z} \in S\}.$$

Note that each $\widehat{\boldsymbol{z}}$ consists of $d$ zeros followed by a rescaling of $\boldsymbol{z}$. For each $\boldsymbol{z}$ we naturally set $c_{\widehat{\boldsymbol{z}}} = c_{\boldsymbol{z}}$. As $\|\boldsymbol{z}\| \leq \sqrt{\alpha_0}$, $\|\widehat{\boldsymbol{z}}\| \leq 1$. We have established the conditions that allow us to apply Lemma 12.

To see that the left-hand sides of (11) and (10) are equal, note that our constructed vectors satisfy

$$\langle \widehat{\boldsymbol{z}}, \boldsymbol{w}_i \rangle = \boldsymbol{z}(i).$$

So,

$$\boldsymbol{\Phi}^{\boldsymbol{w}}_{\widehat{\boldsymbol{z}}}(\boldsymbol{\sigma}) = \prod_{i=1}^{n}(1 - \boldsymbol{\sigma}(i)\langle \widehat{\boldsymbol{z}}, \boldsymbol{w}_i \rangle) = \prod_{i=1}^{n}(1 - \boldsymbol{\sigma}(i)\boldsymbol{z}(i)) = \boldsymbol{\Phi}(\boldsymbol{z}, \boldsymbol{\sigma}).$$

The right-hand sides are equal because

$$\begin{aligned}
\left\| \sum_{\widehat{\boldsymbol{z}} \in U} c_{\widehat{\boldsymbol{z}}} \widehat{\boldsymbol{z}} \boldsymbol{\Phi}^{\boldsymbol{w}}_{\widehat{\boldsymbol{z}}}(\boldsymbol{\sigma}) \right\|^2 &= \frac{1}{\alpha_0} \left\| \sum_{\boldsymbol{z} \in S} c_{\boldsymbol{z}} (\mathbf{0}_d \oplus \boldsymbol{z}) \boldsymbol{\Phi}^{\boldsymbol{w}}_{\widehat{\boldsymbol{z}}}(\boldsymbol{\sigma}) \right\|^2 \\
&= \frac{1}{\alpha_0} \left\| \mathbf{0}_d \oplus \sum_{\boldsymbol{z} \in S} c_{\boldsymbol{z}} \boldsymbol{z} \boldsymbol{\Phi}(\boldsymbol{z}, \boldsymbol{\sigma}) \right\|^2 \\
&= \frac{1}{\alpha_0} \left\| \sum_{\boldsymbol{z} \in S} c_{\boldsymbol{z}} \boldsymbol{z} \boldsymbol{\Phi}(\boldsymbol{z}, \boldsymbol{\sigma}) \right\|^2 .
\end{aligned}$$

□

To apply Lemma 11 to prove Lemma 10, we need to express every function $f(\boldsymbol{\sigma})$ as a sum of the form $\sum_{\boldsymbol{z} \in S} c_{\boldsymbol{z}} \boldsymbol{\Phi}(\boldsymbol{z}, \boldsymbol{\sigma})$, where each $\boldsymbol{z}$ has norm at most $\sqrt{\alpha_0}$. The difficulty is that we need for the right-hand sides of the inequalities established in these lemmas to line up as well. If we were to ignore the right-hand sides and the norm restriction, we could take $S = \{\pm 1\}^n$ and for each $\boldsymbol{z} \in S$ set

$$c_{\boldsymbol{z}} = 2^{-n} f(-\boldsymbol{z}).$$

To see what this would provide, note that for $\boldsymbol{z}, \boldsymbol{\sigma} \in \{\pm 1\}^n$

$$\boldsymbol{\Phi}(\boldsymbol{z}, \boldsymbol{\sigma}) = \begin{cases} 2^n & \text{if } \boldsymbol{\sigma} = -\boldsymbol{z} \\ 0 & \textit{otherwise.} \end{cases} \tag{13}$$

So we would have

$$\sum_{\boldsymbol{z} \in \{\pm 1\}^n} c_{\boldsymbol{z}} \boldsymbol{\Phi}(\boldsymbol{z}, \boldsymbol{\sigma}) = \sum_{\boldsymbol{z} \in \{\pm 1\}^n} 2^n \mathbf{1}_{\boldsymbol{\sigma} = -\boldsymbol{z}} 2^{-n} f(-\boldsymbol{z}) = f(\boldsymbol{\sigma}).$$

We fix the defects in this construction in two phases. First, we employ a sum over $S = \{0, \pm 1\}^n$ and give a concrete expression for the coefficients $c_{\boldsymbol{z}}$. We then show that this can be replaced by a sum over $S = \{0, \pm\delta\}^n$ for any $\delta > 0$.

**Lemma 13.** *Let $S = \{0, \pm 1\}^n$. For every function $f : \{\pm 1\}^n \to \mathbb{R}$, there exist real coefficients $c_{\boldsymbol{z}}$ for $\boldsymbol{z} \in S$ so that*

$$\sum_{\boldsymbol{z} \in \{0,\pm 1\}^n} c_{\boldsymbol{z}} \boldsymbol{\Phi}(\boldsymbol{z}, \boldsymbol{\sigma}) = f(\boldsymbol{\sigma}), \tag{14}$$

*and for $i \in [n]$*

$$\sum_{\boldsymbol{z} \in \{0,\pm 1\}^n} c_{\boldsymbol{z}} \boldsymbol{z}(i) \boldsymbol{\Phi}(\boldsymbol{z}, \boldsymbol{\sigma}) = -\boldsymbol{\sigma}(i)(D_i f)(\boldsymbol{\sigma}). \tag{15}$$

*Proof.* Let $f_i = P_i f$, so $D_i f = f - f_i$. Also define

$$g = f - \sum_{i=1}^{n} f_i, \qquad \text{so } f = g + \sum_{i=1}^{n} f_i.$$

Let $S_0 = \{\pm 1\}^n$. Equation (13) tells us that the functions $\{\Phi(\boldsymbol{z}, \cdot) : \boldsymbol{z} \in S_0\}$ are a basis of all functions on $\{\pm 1\}^n$ . Similarly, let $S_i = \{\boldsymbol{z} : \boldsymbol{z}(i) = 0 \text{ and } \boldsymbol{z}(j) = \pm 1 \text{ for } j \neq i\}$. The functions $\{\Phi(\boldsymbol{z}, \cdot) : \boldsymbol{z} \in S_i\}$ form a basis for all functions on $\{\pm 1\}^n$ that do not depend on their $i$th coordinate. As $f_i$ does not depend on its $i$th coordinate, there exist coefficients $c_{i,\boldsymbol{z}}$ such that for all $\boldsymbol{\sigma} \in \{\pm 1\}^n$,

$$g(\boldsymbol{\sigma}) = \sum_{\boldsymbol{z} \in S_0} c_{0,\boldsymbol{z}} \Phi(\boldsymbol{z}, \boldsymbol{\sigma}) \qquad \text{and} \qquad f_i(\boldsymbol{\sigma}) = \sum_{\boldsymbol{z} \in S_i} c_{i,\boldsymbol{z}} \Phi(\boldsymbol{z}, \boldsymbol{\sigma}).$$

This implies

$$f(\boldsymbol{\sigma}) = \sum_{i=0}^{n} \sum_{\boldsymbol{z} \in S_i} c_{i,\boldsymbol{z}} \Phi(\boldsymbol{z}, \boldsymbol{\sigma}).$$

Furthermore, by the identity $\boldsymbol{z}(j)\Phi(\boldsymbol{z}, \boldsymbol{\sigma}) = -\boldsymbol{\sigma}(j)\Phi(\boldsymbol{z}, \boldsymbol{\sigma})$ which holds when $\boldsymbol{z}(j) \in \{\pm 1\}$ and $\boldsymbol{z}(j)\Phi(\boldsymbol{z}, \boldsymbol{\sigma}) = 0$ when $\boldsymbol{z}(j) = 0$, it holds

$$\begin{aligned}
\sum_{i=0}^{n} \sum_{\boldsymbol{z} \in S_i} c_{i,\boldsymbol{z}} \boldsymbol{z}(j) \Phi(\boldsymbol{z}, \boldsymbol{\sigma}) &= -\sum_{i=0}^{n} \sum_{\boldsymbol{z} \in S_i} c_{i,\boldsymbol{z}} \boldsymbol{\sigma}(j) \Phi(\boldsymbol{z}, \sigma) \mathbf{1}_{i \neq j} \\
&= -\boldsymbol{\sigma}(j) \left( g(\boldsymbol{\sigma}) + \sum_{i \notin \{0,j\}} f_i(\boldsymbol{\sigma}) \right) \\
&= -\boldsymbol{\sigma}(j)(D_j f)(\boldsymbol{\sigma}).
\end{aligned}$$

For $\boldsymbol{z} \in S_i$ for some $i$, set $c_{\boldsymbol{z}} = c_{i,\boldsymbol{z}}$. This definition is valid because the sets $S_i$ are disjoint. Set $c_{\boldsymbol{z}}$ to zero for all other $\boldsymbol{z}$. □

Before we can apply Lemma 11, we need to replace the set of evaluations points $\{0, \pm 1\}^n$ with $\{0, \pm\delta\}^n$ for $\delta$ small enough that $\delta\sqrt{n} \leq \alpha_0$. This is easy, and we can do it for any $\delta > 0$. Rather than deriving the formula for the coefficients, we just prove one exists.

To start, note that $\boldsymbol{\Phi}(\boldsymbol{z}, \boldsymbol{\sigma})$ and $\boldsymbol{z}(i)\boldsymbol{\Phi}(\boldsymbol{z}, \boldsymbol{\sigma})$ are polynomials in $\boldsymbol{z}$ in which every variable has degree at most 2. Let $\mathcal{P}_{\boldsymbol{z}}^2$ be the set of such polynomials. We use the following consequence of Lemma 13.

**Corollary 14.** *There exists a linear function $L$ mapping $\mathcal{P}_{\boldsymbol{z}}^2$ to the reals such that*

$$L\left[\boldsymbol{\Phi}(\boldsymbol{z},\boldsymbol{\sigma})\right] = f(\boldsymbol{\sigma}), \tag{16}$$

*and for $i \in [n]$*

$$L\left[\boldsymbol{z}(i)\boldsymbol{\Phi}(\boldsymbol{z},\boldsymbol{\sigma})\right] = -\boldsymbol{\sigma}(i)(D_i f)(\boldsymbol{\sigma}). \tag{17}$$

*Proof.* The evaluation of a polynomial is a linear function of its coefficients, as is the sum of evaluations. □

**Lemma 15.** *For any $\delta > 0$, let $S = \{0, \pm\delta\}^n$. For every function $f : \{\pm 1\}^n \to \mathbb{R}$, there exist real coefficients $c_{\boldsymbol{z}}$ for $\boldsymbol{z} \in S$ so that*

$$\sum_{\boldsymbol{z}\in S} c_{\boldsymbol{z}} \boldsymbol{\Phi}(\boldsymbol{z},\boldsymbol{\sigma}) = f(\boldsymbol{\sigma}), \tag{18}$$

*and for $i \in [n]$*

$$\sum_{\boldsymbol{z}\in S} c_{\boldsymbol{z}} \boldsymbol{z}(i)\boldsymbol{\Phi}(\boldsymbol{z},\boldsymbol{\sigma}) = -\boldsymbol{\sigma}(i)(D_i f)(\boldsymbol{\sigma}). \tag{19}$$

*Proof.* Lagrange interpolation provides for every $\boldsymbol{y} \in S$ a polynomial $p_{\boldsymbol{y}}(\boldsymbol{z}) \in \mathcal{P}_{\boldsymbol{z}}^2$ so that

$$p_{\boldsymbol{y}}(\boldsymbol{z}) = \begin{cases} 1 & \text{if } \boldsymbol{y} = \boldsymbol{z} \\ 0 & \text{for every other } \boldsymbol{z} \in S. \end{cases}$$

Moreover, for every $q(\boldsymbol{z}) \in \mathcal{P}_{\boldsymbol{z}}^2$,

$$q(\boldsymbol{z}) = \sum_{\boldsymbol{y}\in S} q(\boldsymbol{y}) p_{\boldsymbol{y}}(\boldsymbol{z}).$$

So, for any linear function $L : \mathcal{P}_{\boldsymbol{z}}^2 \to \mathbb{R}$ and any polynomial $q(\boldsymbol{z}) \in \mathcal{P}_{\boldsymbol{z}}^2$,

$$L[q(\boldsymbol{z})] = L\left[\sum_{\boldsymbol{y}\in S} q(\boldsymbol{y}) p_{\boldsymbol{y}}(\boldsymbol{z})\right] = \sum_{\boldsymbol{y}\in S} q(\boldsymbol{y}) L[p_{\boldsymbol{y}}(\boldsymbol{z})].$$

That is, we may take $c_{\boldsymbol{y}} = L[p_{\boldsymbol{y}}(\boldsymbol{z})]$. Corollary 14 provides the desired linear function $L$. □

*Proof of Lemma 10.* Let $f$ be any function from $\{\pm 1\}^n$ to $\mathbb{R}$. Set $\delta = \alpha_0/\sqrt{n}$ and $S = \{0, \pm\delta\}^n$. By Lemma 15, there exist real coefficients $c_{\boldsymbol{z}}$ for $\boldsymbol{z} \in S$ so that

$$\sum_{\boldsymbol{z}\in S} c_{\boldsymbol{z}} \boldsymbol{\Phi}(\boldsymbol{z},\boldsymbol{\sigma}) = f(\boldsymbol{\sigma}) \quad \text{and for } i \in [n] \quad \sum_{\boldsymbol{z}\in S} c_{\boldsymbol{z}} \boldsymbol{z}(i)\boldsymbol{\Phi}(\boldsymbol{z},\boldsymbol{\sigma}) = -\boldsymbol{\sigma}(i)(D_i f)(\boldsymbol{\sigma}).$$

To apply Lemma 11, note that

$$\mathbb{E}_{\boldsymbol{h}} \left\| \sum_{\boldsymbol{z}\in S} c_{\boldsymbol{z}} \boldsymbol{z} \boldsymbol{\Phi}(\boldsymbol{z},\boldsymbol{\sigma}) \right\|^2 = \mathbb{E}_{\boldsymbol{h}} \sum_{i\in[n]} \left( \sum_{\boldsymbol{z}\in S} c_{\boldsymbol{z}} \boldsymbol{z}(i) \boldsymbol{\Phi}(\boldsymbol{z},\boldsymbol{\sigma}) \right)^2 = \mathbb{E}_{\boldsymbol{h}} \sum_{i\in[n]} (-\boldsymbol{\sigma}(i)(D_i f)(\boldsymbol{\sigma}))^2 = \mathbb{E}_{\boldsymbol{h}} \sum_{i\in[n]} ((D_i f)(\boldsymbol{\sigma}))^2,$$

as $\sigma(i) \in \pm 1$. So, Lemma 11 gives

$$\mathbf{Var}_{\boldsymbol{h}}\left(f(\boldsymbol{\sigma})\right) \le \frac{1}{\alpha_0} \mathbb{E}_{\boldsymbol{h}} \sum_{i\in[n]} ((D_i f)(\boldsymbol{\sigma}))^2 = \frac{1}{\alpha_0} \sum_{i\in[n]} \mathbb{E}_{\boldsymbol{h}} ((D_i f)(\boldsymbol{\sigma}))^2.$$

□

Now we are ready to prove Theorem 2.

*Proof of Theorem 2.* We consider the tilted distribution (5). It is well known that the spectral gap of the Glauber dynamics $P$ has a variational characterisation [LP17, Lemma 13.7]:

$$\operatorname{gap}(P) = \inf_{\mathbf{Var}_{\boldsymbol{h}}(f)>0} \frac{\langle f, (I-P)f\rangle_h}{\mathbf{Var}_{\boldsymbol{h}}(f)},$$

where $\langle f, g\rangle_h = \mathbb{E}_{\boldsymbol{h}}[fg]$. Since $I - P = \frac{1}{n}\sum_{i=1}^n D_i$ and $D_i = I - P_i$ is an orthogonal projection,

$$\langle f, (I-P)f\rangle_h = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_h\left[(D_i f)^2\right].$$

So, Lemma 10 implies $\operatorname{gap}(P) \geq \alpha_0/n$.

It remains to justify the stated mixing bound. Since $P$ is irreducible, reversible, and has only non-negative eigenvalues, there is a well-known relationship between the spectral gap and the mixing time [LP17, Theorem 12.4]

$$t_{\mathrm{mix},h}(\varepsilon) \leq \left\lceil \frac{1}{\operatorname{gap}(P)} \log\left(\frac{1}{\varepsilon \mu_{\boldsymbol{h},\min}}\right)\right\rceil,$$

where $\mu_{\boldsymbol{h},\min} \stackrel{\text{def}}{=} \min_{\boldsymbol{\sigma}} \mu_h(\boldsymbol{\sigma})$. We bound $\mu_{\boldsymbol{h},\min}$ next. By Cauchy-Schwarz, $\left|\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle\right| \leq 1 - \alpha_0$. Together with $e^{-\|\boldsymbol{h}\|_1} \leq e^{\langle \boldsymbol{h},\boldsymbol{\sigma}\rangle} \leq e^{\|\boldsymbol{h}\|_1}$, this implies

$$\mu_{\boldsymbol{h},\min} \geq 2^{-n} e^{-2\|\boldsymbol{h}\|_1} \left(\frac{\alpha_0}{2-\alpha_0}\right)^{n^2}. \qquad \square$$

We now prove Theorem 3 and Corollary 4.

*Proof of Theorem 3.* Let $\rho \stackrel{\text{def}}{=} \lambda_{\max}(\boldsymbol{T})$. Since $\rho I_n - \boldsymbol{T} \succeq 0$, we may choose vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ such that

$$\forall i \neq j, \quad \langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle = -\boldsymbol{T}_{ij} \qquad \text{and} \qquad \|\boldsymbol{v}_i\|^2 = \rho < 1.$$

With this choice of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, the tilted distribution (5) coincides with the general Ising model (2). Applying Theorem 2 with $\alpha_0 = 1 - \rho$ yields the desired spectral gap bound. $\square$

*Proof of Corollary 4.* Since $|\tanh x - x| \leq |x|^3$, the difference between $\boldsymbol{T}$ and $\boldsymbol{J}$ satisfies

$$\mathbb{E}\,\|\boldsymbol{T} - \boldsymbol{J}\|_{\mathrm{op}}^2 \leq \mathbb{E}\,\|\boldsymbol{T} - \boldsymbol{J}\|_{\mathrm{F}}^2 \leq \sum_{i\neq j} \mathbb{E}\boldsymbol{J}_{ij}^6 = O_\beta(n^{-1}). \tag{20}$$

With high probability $\lambda_{\max}(\boldsymbol{J}) \leq 2\beta + o_n(1)$ (see [BY88] or [AGZ10, Theorem 2.1.22]). Together with (20), this implies $\lambda_{\max}(\boldsymbol{T}) = 2\beta + o_n(1)$ with high probability. The desired spectral gap bound then follows from Theorem 3. $\square$

## 8 General rapid mixing results

It is natural to wonder to what extent the machinery in the last section generalises to other spin systems. We first generalise it to a complete graph with uniform interactions, and then see how it can be applied to more general spin systems.

Let $q \geq 2$ and $n \geq 2$ be integers, and let $\boldsymbol{A} \in \mathbb{R}_{\geq 0}^{q \times q}$ be a symmetric matrix with at most one positive eigenvalue, and no all-zero rows. For $i \in [n]$, let $\boldsymbol{\lambda}_i \in \mathbb{R}_{\geq 0}^{q}$ be a vector of external fields. Define the partition function

$$\mathcal{Z}_{\boldsymbol{A}}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n) \stackrel{\text{def}}{=} \sum_{\boldsymbol{\sigma} \in [q]^n} \prod_{i=1}^{n} \boldsymbol{\lambda}_i(\boldsymbol{\sigma}_i) \prod_{1 \leq i < j \leq n} \boldsymbol{A}_{\boldsymbol{\sigma}_i \boldsymbol{\sigma}_j}. \tag{21}$$

We assume $\mathcal{Z}_{\boldsymbol{A}}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n) > 0$. Define the Gibbs distribution

$$\mu_{\boldsymbol{A},\boldsymbol{\lambda}}(\boldsymbol{\sigma}) \stackrel{\text{def}}{=} \frac{1}{\mathcal{Z}_{\boldsymbol{A}}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n)} \prod_{i=1}^{n} \boldsymbol{\lambda}_i(\boldsymbol{\sigma}_i) \prod_{1 \leq i < j \leq n} \boldsymbol{A}_{\boldsymbol{\sigma}_i \boldsymbol{\sigma}_j}.$$

Let $P_i$ be the averaging operator on site $i$, i.e., for every $f : [q]^n \to \mathbb{R}$ and $\boldsymbol{\sigma}$ in the support of $\mu_{\boldsymbol{A},\boldsymbol{\lambda}}$,

$$(P_i f)(\boldsymbol{\sigma}) = \mathbb{E}_{\boldsymbol{\tau} \sim \mu_{\boldsymbol{A},\boldsymbol{\lambda}}} \left[ f(\boldsymbol{\tau}) \mid \boldsymbol{\tau}_{-i} = \boldsymbol{\sigma}_{-i} \right]. \tag{22}$$

The Glauber dynamics $P$ for distribution $\mu_{\boldsymbol{A},\boldsymbol{\lambda}}$ is defined by

$$P \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} P_i.$$

Also, define

$$D_i \stackrel{\text{def}}{=} I - P_i.$$

The Perron-Frobenius theorem gives a decomposition

$$\boldsymbol{A} = \boldsymbol{\rho}\boldsymbol{\rho}^T - \boldsymbol{R}\boldsymbol{R}^T, \tag{23}$$

where $\boldsymbol{\rho} \in \mathbb{R}_{>0}^{q}$ and $\boldsymbol{R} \in \mathbb{R}^{q \times r}$ for some $r \geq 0$. For $a \in [q]$, suppose the $a$-th row of $\boldsymbol{R}$ is $\boldsymbol{r}_a^T$, and let

$$\boldsymbol{v}_a \stackrel{\text{def}}{=} \frac{\boldsymbol{r}_a}{\boldsymbol{\rho}_a}, \quad c_a \stackrel{\text{def}}{=} 1 - \|\boldsymbol{v}_a\|^2 = \frac{\boldsymbol{A}_{aa}}{\boldsymbol{\rho}_a^2}, \quad \text{and} \quad c_{\boldsymbol{A}} \stackrel{\text{def}}{=} \min_{a \in [q]} c_a. \tag{24}$$

The entries of $\boldsymbol{A}$ can be written as

$$\boldsymbol{A}_{ab} = \boldsymbol{\rho}_a \boldsymbol{\rho}_b (1 - \langle \boldsymbol{v}_a, \boldsymbol{v}_b \rangle) \qquad \text{for } a, b \in [q].$$

We obtain the following spectral gap bound.

**Theorem 16.** *If $c_{\boldsymbol{A}} > 0$, the Glauber dynamics $P$ satisfies $\operatorname{gap}(P) \geq \frac{c_{\boldsymbol{A}}}{n}$.*

In particular, if every diagonal entry of $\boldsymbol{A}$ is positive, then $c_{\boldsymbol{A}} > 0$ and the spectral gap is $\Omega_{\boldsymbol{A}}(1/n)$.

The proof of Theorem 16 follows the same strategy as that of Theorem 2. We first introduce the analogue of Lemma 12.

**Lemma 17.** *Let* $\{\boldsymbol{w}_i^a\}_{i\in[n],a\in[q]} \subseteq \mathbb{R}^d$ *be vectors of norm at most 1 such that*

$$\left\langle \boldsymbol{w}_j^a, \boldsymbol{w}_k^b \right\rangle = \langle \boldsymbol{v}_a, \boldsymbol{v}_b \rangle \qquad \text{for all } j \neq k \text{ and } a, b \in [q].$$

*For* $\boldsymbol{u} \in \mathbb{R}^d$*, define*

$$\boldsymbol{\Phi}_{\boldsymbol{u}}^{\boldsymbol{w}}(\boldsymbol{\sigma}) \stackrel{\text{def}}{=} \prod_{i=1}^{n} (1 - \langle \boldsymbol{u}, \boldsymbol{w}_i^{\boldsymbol{\sigma}_i} \rangle).$$

*Let* $U$ *be a finite set of vectors in* $\mathbb{R}^d$*. For each* $\boldsymbol{u} \in U$*, let* $c_{\boldsymbol{u}}$ *be a real number. If all the vectors in* $U$ *have norm at most 1, then*

$$\mathbf{Var}_{\boldsymbol{A},\boldsymbol{\lambda}} \left( \sum_{\boldsymbol{u}\in U} c_{\boldsymbol{u}} \boldsymbol{\Phi}_{\boldsymbol{u}}^{\boldsymbol{w}}(\boldsymbol{\sigma}) \right) \leq \mathbb{E}_{\boldsymbol{A},\boldsymbol{\lambda}} \left\| \sum_{\boldsymbol{u}\in U} c_{\boldsymbol{u}} \boldsymbol{u} \boldsymbol{\Phi}_{\boldsymbol{u}}^{\boldsymbol{w}}(\boldsymbol{\sigma}) \right\|^2, \tag{25}$$

*where the expectation and variance are taken over* $\boldsymbol{\sigma}$ *distributed according to* $\mu_{\boldsymbol{A},\boldsymbol{\lambda}}$*.*

*Proof.* Let $\boldsymbol{B}$ be the Lorentz matrix of the vectors $\{\boldsymbol{w}_i^a\}_{i\in[n],a\in[q]}$, $\mathbf{0}$, and $\{\boldsymbol{u}\}_{\boldsymbol{u}\in U}$. The vectors $\boldsymbol{w}_i^a$ are indexed by coordinates $i^a$ and $\mathbf{0}$ is indexed by coordinate 0. For each $\boldsymbol{u} \in U$, we index the vector $\boldsymbol{u}$ by a coordinate we call $\boldsymbol{u}$. For each $i \in [n]$, we define the selection vector $\boldsymbol{x}_i$ by

$$\boldsymbol{x}_i(i^a) = \boldsymbol{\lambda}_i(a)\boldsymbol{\rho}_a^{n-1} \qquad \text{for all } a \in [q],$$

and set it to zero on all other coordinates. Let $\boldsymbol{x}_0$ be the elementary unit vector in direction 0. And, we introduce just one more selection vector $\boldsymbol{x}_{-1}$ defined by

$$\boldsymbol{x}_{-1}(\boldsymbol{u}) = c_{\boldsymbol{u}}, \quad \text{for } \boldsymbol{u} \in U,$$

and set it to zero on all other coordinates. For selection vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, define

$$\mathcal{Y}(\boldsymbol{a}, \boldsymbol{b}) \stackrel{\text{def}}{=} \mathcal{V}_{\boldsymbol{B}}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n).$$

We apply Theorem 5 to show

$$\mathcal{Y}(\boldsymbol{x}_0, \boldsymbol{x}_{-1})^2 \geq \mathcal{Y}(\boldsymbol{x}_0, \boldsymbol{x}_0)\mathcal{Y}(\boldsymbol{x}_{-1}, \boldsymbol{x}_{-1}). \tag{26}$$

The assumption $\left\langle \boldsymbol{w}_j^a, \boldsymbol{w}_k^b \right\rangle = \langle \boldsymbol{v}_a, \boldsymbol{v}_b \rangle$ guarantees that

$$\begin{aligned}
\mathcal{Y}(\boldsymbol{x}_0, \boldsymbol{x}_0) &= \mathcal{Z}_{\boldsymbol{A}}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n) \\
\mathcal{Y}(\boldsymbol{x}_0, \boldsymbol{x}_{-1}) &= \mathcal{Z}_{\boldsymbol{A}}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n)\mathbb{E}_{\boldsymbol{A},\boldsymbol{\lambda}} \left[ \sum_{\boldsymbol{u}\in U} c_{\boldsymbol{u}} \boldsymbol{\Phi}_{\boldsymbol{u}}^{\boldsymbol{w}}(\boldsymbol{\sigma}) \right] \\
\mathcal{Y}(\boldsymbol{x}_{-1}, \boldsymbol{x}_{-1}) &= \mathcal{Z}_{\boldsymbol{A}}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n)\mathbb{E}_{\boldsymbol{A},\boldsymbol{\lambda}} \left[ \left( \sum_{\boldsymbol{u}\in U} c_{\boldsymbol{u}} \boldsymbol{\Phi}_{\boldsymbol{u}}^{\boldsymbol{w}}(\boldsymbol{\sigma}) \right)^2 - \left\| \sum_{\boldsymbol{u}\in U} c_{\boldsymbol{u}} \boldsymbol{u} \boldsymbol{\Phi}_{\boldsymbol{u}}^{\boldsymbol{w}}(\boldsymbol{\sigma}) \right\|^2 \right].
\end{aligned}$$

Plugging these terms into (26) and rearranging the terms gives Lemma 17. □

To apply Lemma 17, we use the following interpolation lemma. For $\boldsymbol{z} \in \mathbb{R}^{nq}$, define

$$\boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma}) \stackrel{\text{def}}{=} \prod_{i=1}^{n}(1 - \boldsymbol{z}_{i,\boldsymbol{\sigma}_i}).$$

**Lemma 18.** *For any $\delta > 0$, let $S = \{0, \pm\delta\}^{nq}$. For every $f : [q]^n \to \mathbb{R}$, there are real coefficients $(b_{\boldsymbol{z}})_{\boldsymbol{z}\in S}$ such that, for every $\boldsymbol{\sigma}$ in the support of $\mu_{\boldsymbol{A},\boldsymbol{\lambda}}$,*

$$\sum_{\boldsymbol{z}\in S} b_{\boldsymbol{z}} \boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma}) = f(\boldsymbol{\sigma}), \tag{27}$$

$$\sum_{\boldsymbol{z}\in S} b_{\boldsymbol{z}} \boldsymbol{z}_{i,a} \boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma}) = (D_i f)(\boldsymbol{\sigma}) \qquad \text{for all } i \in [n] \text{ and } a \in [q]. \tag{28}$$

*Proof.* Recall that $\mathcal{P}^2_{\boldsymbol{z}}$ is the set of polynomials $g(\boldsymbol{z})$ in which every variable has degree at most 2. For this proof, extend $P_i$ to an operator on all $f : [q]^n \to \mathbb{R}$ by the same conditional-expectation formula (22) whenever the conditioning event has positive probability, and set it to zero otherwise. Thus $P_i f$ depends only on $\boldsymbol{\sigma}_{-i}$. Viewing $\boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma})$ as a function of $\boldsymbol{z}$, for any fixed $\boldsymbol{\sigma} \in [q]^n$, $i \in [n]$, and $a \in [q]$, both $\boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma})$ and $(1-\boldsymbol{z}_{i,a})\boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma})$ are in $\mathcal{P}^2_{\boldsymbol{z}}$. We show the existence of a linear functional $L$ mapping $\mathcal{P}^2_{\boldsymbol{z}}$ to the reals such that, as a polynomial in $\boldsymbol{z}$,

$$L[\boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma})] = f(\boldsymbol{\sigma}) \qquad \text{and} \qquad L[(1 - \boldsymbol{z}_{i,a})\boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma})] = (P_i f)(\boldsymbol{\sigma}).$$

Suppose there are real coefficients $\boldsymbol{\alpha}_{\boldsymbol{\sigma}}$ and $\boldsymbol{\beta}_{\boldsymbol{\sigma},i,a}$ such that

$$\sum_{\boldsymbol{\sigma}} \boldsymbol{\alpha}_{\boldsymbol{\sigma}} \boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma}) + \sum_{\boldsymbol{\sigma},i,a} \boldsymbol{\beta}_{\boldsymbol{\sigma},i,a}(1 - \boldsymbol{z}_{i,a})\boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma}) = 0. \tag{29}$$

Set $\boldsymbol{y}_{i,a} = 1 - \boldsymbol{z}_{i,a}$. By comparing the monomials in $\boldsymbol{y}$ on both sides, we must have

$$\boldsymbol{\alpha}_{\boldsymbol{\sigma}} = 0 \qquad \text{and} \qquad \boldsymbol{\beta}_{\boldsymbol{\sigma},i,a} + \boldsymbol{\beta}_{\boldsymbol{\tau},i,b} = 0, \tag{30}$$

whenever $\boldsymbol{\tau}_{-i} = \boldsymbol{\sigma}_{-i}$, $\boldsymbol{\sigma}_i = b$, and $\boldsymbol{\tau}_i = a$. Indeed, the first family consists of distinct monomials of total degree $n$, whereas the second family has total degree $n+1$; repeated monomials in the second family arise only by exchanging $a$ and $\boldsymbol{\sigma}_i$. When $a = \boldsymbol{\sigma}_i$, the corresponding monomial occurs only once, so $\boldsymbol{\beta}_{\boldsymbol{\sigma},i,a} = 0$. Since $P_i f$ depends only on $\boldsymbol{\sigma}_{-i}$,

$$(P_i f)(\boldsymbol{\sigma}) = (P_i f)(\boldsymbol{\tau}).$$

Plugging into (30) gives

$$\sum_{\boldsymbol{\sigma}} \boldsymbol{\alpha}_{\boldsymbol{\sigma}} f(\boldsymbol{\sigma}) + \sum_{\boldsymbol{\sigma},i,a} \boldsymbol{\beta}_{\boldsymbol{\sigma},i,a}(P_i f)(\boldsymbol{\sigma}) = 0.$$

Thus the prescribed values respect all linear relations and define a linear functional on the span of these polynomials, which extends to $\mathcal{P}^2_{\boldsymbol{z}}$.

Lagrange interpolation provides for every $\boldsymbol{x} \in S$ a polynomial $p_{\boldsymbol{x}}(\boldsymbol{z}) \in \mathcal{P}^2_{\boldsymbol{z}}$ so that

$$p_{\boldsymbol{x}}(\boldsymbol{z}) = \begin{cases} 1 & \text{if } \boldsymbol{x} = \boldsymbol{z} \\ 0 & \text{for every other } \boldsymbol{z} \in S. \end{cases}$$

Moreover, for every $g(\boldsymbol{z}) \in \mathcal{P}_{\boldsymbol{z}}^2$,

$$g(\boldsymbol{z}) = \sum_{\boldsymbol{x} \in S} g(\boldsymbol{x}) p_{\boldsymbol{x}}(\boldsymbol{z}).$$

So, for any linear functional $L : \mathcal{P}_{\boldsymbol{z}}^2 \to \mathbb{R}$ and any polynomial $g(\boldsymbol{z}) \in \mathcal{P}_{\boldsymbol{z}}^2$,

$$L[g(\boldsymbol{z})] = \sum_{\boldsymbol{x} \in S} g(\boldsymbol{x}) L[p_{\boldsymbol{x}}(\boldsymbol{z})].$$

Take $b_{\boldsymbol{x}} = L[p_{\boldsymbol{x}}(\boldsymbol{z})]$ for the functional constructed above. Applying the interpolation identity to $\boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma})$ gives (27). For $\boldsymbol{\sigma}$ in the support of $\mu_{\boldsymbol{A},\boldsymbol{\lambda}}$, we also have

$$L[\boldsymbol{z}_{i,a} \boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma})] = L[\boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma})] - L[(1 - \boldsymbol{z}_{i,a}) \boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma})] = f(\boldsymbol{\sigma}) - (P_i f)(\boldsymbol{\sigma}) = (D_i f)(\boldsymbol{\sigma}),$$

which gives (28). □

*Proof of Theorem 16.* For simplicity, let $c = c_{\boldsymbol{A}} > 0$. Fix $0 < \epsilon < c$ and, in $\mathbb{R}^r \oplus \mathbb{R}^n \oplus \mathbb{R}^{nq}$, let

$$\boldsymbol{w}_i^a = \boldsymbol{v}_a \oplus \sqrt{c - \epsilon}\, \boldsymbol{e}_i \oplus \sqrt{\epsilon}\, \boldsymbol{e}_{i,a}, \tag{31}$$

where $\boldsymbol{e}_i$ and $\boldsymbol{e}_{i,a}$ are standard basis vectors in $\mathbb{R}^n$ and $\mathbb{R}^{nq}$, respectively. A simple calculation yields $\|\boldsymbol{w}_i^a\|^2 = \|\boldsymbol{v}_a\|^2 + c \le 1$. For all $j \neq k$ and $a, b \in [q]$,

$$\left\langle \boldsymbol{w}_j^a, \boldsymbol{w}_k^b \right\rangle = \langle \boldsymbol{v}_a, \boldsymbol{v}_b \rangle + (c - \epsilon) \langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle + \epsilon \langle \boldsymbol{e}_{j,a}, \boldsymbol{e}_{k,b} \rangle = \langle \boldsymbol{v}_a, \boldsymbol{v}_b \rangle .$$

Furthermore, for each $\boldsymbol{z} \in \mathbb{R}^{nq}$, we let

$$\boldsymbol{u}(\boldsymbol{z}) = \boldsymbol{0}_r \oplus \sum_i \frac{\boldsymbol{z}_{i,1}}{\sqrt{c - \epsilon}} \boldsymbol{e}_i \oplus \sum_{i,a} \frac{\boldsymbol{z}_{i,a} - \boldsymbol{z}_{i,1}}{\sqrt{\epsilon}} \boldsymbol{e}_{i,a}.$$

The inner product $\langle \boldsymbol{u}(\boldsymbol{z}), \boldsymbol{w}_i^a \rangle = \boldsymbol{z}_{i,1} + \boldsymbol{z}_{i,a} - \boldsymbol{z}_{i,1} = \boldsymbol{z}_{i,a}$. So,

$$\boldsymbol{\Phi}_{\boldsymbol{u}(\boldsymbol{z})}^{\boldsymbol{w}}(\boldsymbol{\sigma}) = \prod_{i=1}^n (1 - \langle \boldsymbol{u}(\boldsymbol{z}), \boldsymbol{w}_i^{\boldsymbol{\sigma}_i} \rangle) = \prod_{i=1}^n (1 - \boldsymbol{z}_{i,\boldsymbol{\sigma}_i}) = \boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma}).$$

Since $\boldsymbol{u}(\boldsymbol{z})$ is linear in $\boldsymbol{z}$, choose $\delta > 0$ sufficiently small that $\|\boldsymbol{u}(\boldsymbol{z})\| \le 1$ for every $\boldsymbol{z} \in S \stackrel{\text{def}}{=} \{0, \pm\delta\}^{nq}$. Applying Lemma 17 to the vectors $\{\boldsymbol{u}(\boldsymbol{z}) : \boldsymbol{z} \in S\}$ gives, for any real coefficients $b_{\boldsymbol{z}}$,

$$\mathbf{Var}_{\boldsymbol{A},\boldsymbol{\lambda}} \left( \sum_{\boldsymbol{z} \in S} b_{\boldsymbol{z}} \boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma}) \right) \le \mathbb{E}_{\boldsymbol{A},\boldsymbol{\lambda}} \left\| \sum_{\boldsymbol{z} \in S} b_{\boldsymbol{z}} \boldsymbol{u}(\boldsymbol{z}) \boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma}) \right\|^2. \tag{32}$$

For any $f : [q]^n \to \mathbb{R}$, take the coefficients supplied by Lemma 18 for this grid. Then

$$\mathbf{Var}_{\boldsymbol{A},\boldsymbol{\lambda}} \left( \sum_{\boldsymbol{z} \in S} b_{\boldsymbol{z}} \boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma}) \right) = \mathbf{Var}_{\boldsymbol{A},\boldsymbol{\lambda}} \left( f(\boldsymbol{\sigma}) \right),$$

and

$$\mathbb{E}_{\boldsymbol{A},\boldsymbol{\lambda}} \left\| \sum_{\boldsymbol{z} \in S} b_{\boldsymbol{z}} \boldsymbol{u}(\boldsymbol{z}) \boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{\sigma}) \right\|^2 = \frac{1}{c - \epsilon} \sum_{i=1}^n \mathbb{E}_{\boldsymbol{A},\boldsymbol{\lambda}} ((D_i f)(\boldsymbol{\sigma}))^2.$$

Combining with (32) and taking $\epsilon \downarrow 0$ gives

$$\mathbf{Var}_{\boldsymbol{A},\boldsymbol{\lambda}}\left(f(\boldsymbol{\sigma})\right) \le \frac{1}{c}\sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{A},\boldsymbol{\lambda}}((D_i f)(\boldsymbol{\sigma}))^2.$$

This establishes the desired spectral gap lower bound for $P$. □

Now let $G = (V, E)$ be a graph with $|V| = n$ and adjacency matrix $\mathrm{Adj}(G)$, and $\boldsymbol{A}$ be an interaction matrix for $q$ spins. For external fields $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_i)_{i\in V}$ with $\boldsymbol{\lambda}_i \in \mathbb{R}^q_{\ge 0}$, define

$$\mathcal{Z}_{G,\boldsymbol{A}}(\boldsymbol{\lambda}) \stackrel{\text{def}}{=} \sum_{\boldsymbol{\sigma}\in[q]^V} \prod_{i\in V} \boldsymbol{\lambda}_i(\boldsymbol{\sigma}_i) \prod_{ij\in E} \boldsymbol{A}_{\boldsymbol{\sigma}_i\boldsymbol{\sigma}_j}.$$

Let $\mu_{G,\boldsymbol{A},\boldsymbol{\lambda}}$ be the corresponding Gibbs distribution and let $P$ be the Glauber dynamics for $\mu_{G,\boldsymbol{A},\boldsymbol{\lambda}}$.

The Perron–Frobenius theorem implies that $\boldsymbol{A} = \boldsymbol{\rho}\boldsymbol{\rho}^T + \boldsymbol{A}'$ for some $\boldsymbol{\rho} \in \mathbb{R}^q_{>0}$ and matrix $\boldsymbol{A}'$. After shifting the weight of $\boldsymbol{\rho}$ to the external fields, and decomposing the remainder into positive and negative parts, we may assume that $\boldsymbol{A} = \mathbf{1}\mathbf{1}^T + \boldsymbol{A}^+ - \boldsymbol{A}^-$ for some matrices $\boldsymbol{A}^+, \boldsymbol{A}^- \succeq 0$. Let $\lambda_{\max} \stackrel{\text{def}}{=} \lambda_{\max}(\mathrm{Adj}(G))$ and $\lambda_{\min} \stackrel{\text{def}}{=} \lambda_{\min}(\mathrm{Adj}(G))$ be the maximum and minimum eigenvalues of the adjacency matrix of $G$. Set $\boldsymbol{B} \stackrel{\text{def}}{=} \lambda_{\max}\boldsymbol{A}^+ - \lambda_{\min}\boldsymbol{A}^-$, and $c_{G,\boldsymbol{A}} \stackrel{\text{def}}{=} 1 - \max_{a\in[q]} \boldsymbol{B}_{aa}$.

**Corollary 19.** *With the notation above, if $c_{G,\boldsymbol{A}} > 0$, then $\mathrm{gap}(P) \ge \frac{c_{G,\boldsymbol{A}}}{n}$.*

*Proof.* To reduce this model to the complete-graph setting of (21), we keep the same vertex set and enlarge the spin set from $[q]$ to $V \times [q]$. By assigning proper external fields $\widehat{\boldsymbol{\lambda}}$, we restrict vertex $i$ to spins of the form $(i, a)$, where $a \in [q]$. To preserve the weight of every configuration, the new interaction matrix $\widehat{\boldsymbol{A}}$ must satisfy, for distinct $i, j \in V$,

$$\widehat{\boldsymbol{A}}_{(i,a),(j,b)} = \begin{cases} \boldsymbol{A}_{ab}, & ij \in E, \\ 1, & i \ne j \text{ and } ij \notin E. \end{cases} \tag{33}$$

The remaining entries, with $i = j$, do not affect these configuration weights and can be chosen freely. We choose them so that $\widehat{\boldsymbol{A}}$ satisfies the hypotheses of Theorem 16; in particular, it must be non-negative entrywise and have at most one positive eigenvalue.

We claim $\widehat{\boldsymbol{A}} \stackrel{\text{def}}{=} \mathbf{1}\mathbf{1}^T - \boldsymbol{\Gamma}$ satisfies all constraints, where the block matrix $\boldsymbol{\Gamma}$, indexed by $V \times [q]$, is defined by

$$\boldsymbol{\Gamma} \stackrel{\text{def}}{=} \boldsymbol{I}_n \otimes \boldsymbol{B} - \mathrm{Adj}(G) \otimes (\boldsymbol{A}^+ - \boldsymbol{A}^-).$$

A straightforward calculation yields (33). By the definition of $\boldsymbol{B}$,

$$\boldsymbol{\Gamma} = \left(\lambda_{\max}\boldsymbol{I}_n - \mathrm{Adj}(G)\right) \otimes \boldsymbol{A}^+ + \left(\mathrm{Adj}(G) - \lambda_{\min}\boldsymbol{I}_n\right) \otimes \boldsymbol{A}^-.$$

Both factors involving $\mathrm{Adj}(G)$ are positive semidefinite. Since $\boldsymbol{A}^+, \boldsymbol{A}^- \succeq 0$, both tensor-product terms are positive semidefinite, and hence $\boldsymbol{\Gamma} \succeq 0$ and $\widehat{\boldsymbol{A}}$ has at most one positive eigenvalue. Finally, the non-negativity of $\widehat{\boldsymbol{A}}$ follows from

$$\widehat{\boldsymbol{A}}_{(i,a),(i,b)} = 1 - \boldsymbol{B}_{ab} \ge 1 - \sqrt{\boldsymbol{B}_{aa}\boldsymbol{B}_{bb}} > 0,$$

where the first inequality follows from $\boldsymbol{B} \succeq 0$.

It remains to compute the quantity $c_{\widehat{\boldsymbol{A}}}$ in (24). Since $\boldsymbol{\Gamma} \succeq 0$, write $\boldsymbol{\Gamma} = \boldsymbol{R}\boldsymbol{R}^T$. Then $\widehat{\boldsymbol{A}} = \mathbf{1}\mathbf{1}^T - \boldsymbol{R}\boldsymbol{R}^T$ is a representation of the form (23) with $\boldsymbol{\rho} = \mathbf{1}$. For this representation, (24) gives

$$c_{\widehat{\boldsymbol{A}}} = \min_{(i,a)\in V\times[q]} \widehat{\boldsymbol{A}}_{(i,a),(i,a)} = 1 - \max_{a\in[q]} \boldsymbol{B}_{aa} = c_{G,\boldsymbol{A}} > 0.$$

Therefore, by Theorem 16, the Glauber dynamics $\widehat{P}$ for the distribution $\mu_{\widehat{\boldsymbol{A}},\widehat{\boldsymbol{\lambda}}}$ has the spectral gap bound $\mathrm{gap}(\widehat{P}) \geq \frac{c_{G,\boldsymbol{A}}}{n}$. Since the Glauber dynamics $\widehat{P}$ can be identified with the Glauber dynamics $P$ for $\mu_{G,\boldsymbol{A},\boldsymbol{\lambda}}$, we have

$$\mathrm{gap}(P) = \mathrm{gap}(\widehat{P}) \geq \frac{c_{G,\boldsymbol{A}}}{n}. \qquad \square$$

What Corollary 19 is saying is that $\lambda_{\max}$ controls the allowed interaction strength of the ferromagnetic part of $\boldsymbol{A}$, whereas $\lambda_{\min}$ controls the anti-ferromagnetic part.

When $G$ is a complete graph, $\lambda_{\min} = -1$. If, in addition, $\boldsymbol{A}$ is anti-ferromagnetic, namely $\boldsymbol{A}^+ = 0$, then the condition of Corollary 19 recovers that of Theorem 16. More generally, on line graphs, where $\lambda_{\min} \geq -2$, Corollary 19 implies a $\Omega(1/n)$ spectral gap bound whenever $c_{\boldsymbol{A}} > 1/2$. Previously, it was known [DHJM21, CLV24] that Glauber dynamics is rapid mixing for any anti-ferromagnetic 2-spin system on line graphs with bounded degree. In comparison, the strength of Corollary 19 is that it applies to any $q \geq 2$, does not require bounded degree, and allows arbitrary external fields.

For the anti-ferromagnetic $q$-state Potts model, let $0 < b < 1$ be the penalty for equal spins. After rescaling, the interaction matrix $\boldsymbol{A}$ can be written as

$$\boldsymbol{A} = \mathbf{1}\mathbf{1}^T - \frac{q(1-b)}{q-1+b}\left(\boldsymbol{I}_q - \frac{1}{q}\mathbf{1}\mathbf{1}^T\right).$$

Writing $\ell \stackrel{\mathrm{def}}{=} -\lambda_{\min}(\mathrm{Adj}(G))$, Corollary 19 therefore gives

$$\mathrm{gap}(P) \geq \frac{1}{n}\left(1 - \ell\frac{(q-1)(1-b)}{q-1+b}\right)$$

whenever the expression in parentheses is positive. This extends the parameter range and improves the spectral-gap lower bound in the recent result of Chen and Wang [CW26, Corollary 6.7].

## Acknowledgments

We thank Yitan Wang for many helpful discussions at the start of this research effort.

## AI Methodology

In 2021, we conjectured the signed sum $\sum_i \boldsymbol{\sigma}_i \boldsymbol{v}_i$ induced by the Ising distribution (1) is 1-subgaussian and that the Glauber dynamics on the distribution is rapidly mixing. Black-box optimization techniques were used to provide evidence for these conjectures. We posed these conjectures to ChatGPT-5.6 Sol Ultra, using a prompt like that used to resolve the Cycle Double-Cover Conjecture [Ope26a]. The first interaction affirmed the first conjecture using the Alexandrov-Fenchel

inequality as the key tool, and established an $n^{O(1/\alpha_0)}$ mixing time bound using the local-to-global argument in [AL20]. We then asked ChatGPT-5.6 Sol Ultra to apply the key idea directly to improve the mixing bound, which led to Theorem 2. Upon digesting the proof, the authors recognized that Theorem 2 provides mixing criteria in general settings described in Section 8. The authors have substantially refined the proofs, both by thinking and using ChatGPT-5.6 and ChatGPT-6.

## References


[Ade26] Ishaq Aden-Ali. Optimal online discrepancy minimization in linear time. *arXiv*, abs/2607.04388, 2026. 2

[AGZ10] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Number 118. Cambridge university press, 2010. 16

[AJK+22] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence: optimal mixing of down-up random walks. In *STOC*, pages 1418–1430. ACM, 2022. 2, 3

[AKV24] Nima Anari, Frederic Koehler, and Thuy-Duong Vuong. Trickle-down in localization schemes and applications. *arXiv*, abs/2407.16104, 2024. 3

[AL20] Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *STOC*, pages 1198–1211. ACM, 2020. 23

[Ale37] Aleksandr D. Aleksandrov. On the theory of mixed volumes of convex bodies II (russian). *Mat. Sbornik*, 2:1205–1238, 1937. English translation available in selected works. Part I. Selected scientific papers. Gordon and Breach, Amsterdam, 1996. 5

[ALOV24] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: High-dimensional walks and an FPRAS for counting bases of a matroid. *Ann. of Math. (2)*, 199(1):259–299, 2024. 5

[Ban98] Wojciech Banaszczyk. Balancing vectors and gaussian measures of n-dimensional convex bodies. *Random Structures & Algorithms*, 12(4):351–360, 1998. 1

[BDGL19] Nikhil Bansal, Daniel Dadush, Shashwat Garg, and Shachar Lovett. The Gram–Schmidt walk: A cure for the Banaszczyk blues. *Theory of Computing*, 15(21):1–27, 2019. 1, 2

[BH20] Petter Brändén and June Huh. Lorentzian polynomials. *Ann. of Math. (2)*, 192(3):821–891, 2020. 5

[BLO26] Mathews Boban, Anqi Li, and Shayan Oveis Gharan. Rank-1-perturbed trickledown theorems: Mixing time of Glauber dynamics for the Sherrington–Kirkpatrick model up to $\beta \leq 1/2 + \varepsilon$. *arXiv*, abs/2609.13138, 2026. 3

[BY88] Zhi-Dong Bai and Yong-Qua Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a wigner matrix. *The Annals of Probability*, pages 1729–1741, 1988. 16

[CGM21] Mary Cryan, Heng Guo, and Giorgos Mousa. Modified log-Sobolev inequalities for strongly log-concave distributions. *Ann. Probab.*, 49(1):506–525, 2021. 5

[CGVY26] Xiaoyu Chen, Heng Guo, Eric Vigoda, and Xiongxin Yang. Fast FPRAS for the permanent. *arXiv*, abs/2609.20717v1, 2026. 2

[CLV21] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: entropy factorization via high-dimensional expansion. In *STOC*, pages 1537–1550. ACM, 2021. 11

[CLV23] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Rapid mixing of Glauber dynamics up to uniqueness via contraction. *SIAM J. Comput.*, 52(1):196–237, 2023. 8

[CLV24] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Spectral independence via stability and applications to Holant-type problems. *TheoretiCS*, 3(16):1–49, 2024. 22

[CW26] Zongchen Chen and Sihan Wang. High-dimensional ultra-log-concave distributions, 2026. arXiv:2609.23994. 3, 22

[Dev86] Luc Devroye. *General Principles in Random Variate Generation*. Springer New York, New York, NY, 1986. 4

[DGLN19] Daniel Dadush, Shashwat Garg, Shachar Lovett, and Aleksandar Nikolov. Towards a constructive version of Banaszczyk's vector balancing theorem. *Theory of Computing*, 15(1):1–58, 2019. 1

[DHJM21] Martin E. Dyer, Marc Heinrich, Mark Jerrum, and Haiko Müller. Polynomial-time approximation algorithms for the antiferromagnetic Ising model on line graphs. *Comb. Probab. Comput.*, 30(6):905–921, 2021. 22

[DHMR26] Hang Du, Brice Huang, Sidhanth Mohanty, and Amit Rajaraman. Sampling from the Sherrington–Kirkpatrick model up to criticality. *arXiv*, abs/2610.17717, 2026. 3

[DLSS26a] Ewan Davies, Holden Lee, Juspreet Singh Sandhu, and Jonathan Shi. Potential Hessian ascent III: Sampling the Sherrington–Kirkpatrick model at $\beta < 1/2$. *arXiv*, abs/2605.03718, 2026. 3

[DLSS26b] Ewan Davies, Holden Lee, Juspreet Singh Sandhu, and Jonathan Shi. Weak Poincaré inequalities via approximate stochastic localization: Application to sampling the Sherrington–Kirkpatrick model. *arXiv*, abs/2607.08160, 2026. 3

[EKZ22] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. A spectral condition for spectral gap: Fast mixing in high-temperature Ising models. *Probab. Theory Related Fields*, 182(3–4):1035–1051, 2022. 2, 3

[Fen36] Werner Fenchel. Inégalités quadratiques entre les volumes mixtes des corps convexes. *C. R. Acad. Sci. Paris*, 203:647–650, 1936. 5

[Gur09] Leonid Gurvits. On multivariate Newton-like inequalities. In *Advances in combinatorial mathematics*, pages 61–78. Springer, Berlin, 2009. 5

[HLY26] Zhe Hou, Jingcheng Liu, and Yixiao Yu. Simulated annealing and weak Poincaré inequalities with inverse-polynomial accuracy for the Sherrington–Kirkpatrick model. *arXiv*, abs/2610.09882, 2026. 3

[HSSZ24] Christopher Harshaw, Fredrik Sävje, Daniel A Spielman, and Peng Zhang. Balancing covariates in randomized experiments with the gram–schmidt walk design. *Journal of the American Statistical Association*, 119(548):2934–2946, 2024. 1

[LOS25] Joonkyung Lee, Jaeseong Oh, and Jaehyeon Seo. Counting homomorphisms in antiferromagnetic graphs via Lorentzian polynomials. *arXiv*, abs/2506.13659, 2025. 5

[LP17] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, second edition, 2017. With contributions by Elizabeth L. Wilmer, With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson. 16

[LSS26] Holden Lee, Juspreet Singh Sandhu, and Jonathan Shi. Potential Hessian ascent IV: Sampling the Sherrington–Kirkpatrick model at $\beta < 1$. *arXiv*, abs/2609.30590, 2026. 3

[Ope26a] OpenAI. Prompt used for "a proof of the cycle double cover conjecture", July 2026. Full prompt given to GPT-5.6 Sol Ultra. 22

[Ope26b] OpenAI. A spectral gap throughout the high-temperature Sherrington–Kirkpatrick phase. OpenAI Math Release preprint, 2026. September 24, 2026. 3

[vL81] Jacobus Hendricus van Lint. Notes on Egoritsjev's proof of the van der Waerden conjecture. *Linear Algebra and Its Applications*, 39:1–8, 1981. 5

[Wan26] Sihan Wang. Optimal mixing of Glauber dynamics for the Sherrington–Kirkpatrick model at $\beta < 1/2$. *arXiv*, abs/2608.22159, 2026. 3